Multi-Domain Multi-Objective Optimisation of Urban District Environmental Performance

SHEN CHEN

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Social Sciences
School of Architecture

December 2021

Acknowledgements

I wish to sincerely acknowledge my gratitude to the people who have provided me numerous supports during my PhD study.

Firstly, I am very grateful to my supervisor, Professor Jian Kang, for his constant assistance from the very beginning of this research. He generously transferred his knowledge not limiting to acoustic field to me and trained me to be an independent researcher. Although the relinquishment of his supervision on my PhD thesis in the late stage, his concerns about my research never stop.

Secondly, I am very grateful to my current supervisor, Professor Darren Robinson. Since his supervision on my thesis writing in the late stage, a great number of valuable advices are given. Several long-troubled problems are solved under his suggestion. Thanks for his perseverance that leads me working to the very end of this research work. His sharp mind and fast response also impress me as an inspiration for personal growth.

Thirdly, special thanks to Professor Jian Chen, for his selflessly knowledge sharing regarding to optimisation and algorithms which is vital for an architectural background researcher.
I acknowledge the financial support from China Scholarship Council.

Most importantly thanks to parents, for their persistent support and encouragement all though my PhD study. Without their understanding and support, it would be impossible for me to complete it.

Special thanks to all the people who ever gave me emotional support. You are the reasons that make me never stop.

Abstract

Energy and environmental building performance simulation techniques have advanced considerably throughout the past half a century. Sophisticated and easy to use simulation engines now exist that can simultaneously simulate heat flow, fluid flow, plant systems, daylighting and radiation exchange and the impacts of building occupants on these systems and the corresponding feedbacks. Meanwhile, mass rural-urban migration has meant that the global population is now predominantly urban – we have become *homo urbanus* – with the vast majority, more than three quarters, of global resource use being concentrated into urban settlements. With growing concerns over climate change, there is thus a need to identify ways of reducing the negative environmental impacts of urban settlements, whilst ensuring that the quality of urban life is maintained or enhanced. This principle is enshrined in the UN Sustainable Development Goal 11: Sustainable Cities and Communities. In consequence, the last quarter of a century has seen a considerable increase in research activity to develop computational techniques that can efficiently and accurately simulate energy and environmental performance at higher spatial scales, and accounting for increasingly sophisticated building-occupant, building-building and building-system interactions. It is in this urban complexity that this thesis is situated.

Specifically, this thesis seeks to develop and apply a computational framework with which energy (using sunlight availability as a corollary), thermal and acoustic comfort can be simulated and optimised in the urban context, through a series of urban district use cases located in China. In the first instance, candidate building and urban morphology parameters are identified through cluster analysis. These parameters are then used to support parametric modelling of the above sunlight, thermal comfort and acoustic comfort performance domains, using dedicated simulation techniques. Hierarchical cluster analysis is also applied to these results to synthesise preliminary design guidelines. These simulation results are also used to train meta-models, developed using an adapted combination of generic regression neural network (GRNN) and grey wolf optimiser (GWO) algorithms. The trained and validated meta-models are then used to define the objective functions employed by a non-dominated sorting genetic algorithm with elitist strategy (NSGAII) optimiser to identify feasible Pareto solution sets of our three performance domains, employing the earlier defined morphology parameters. Finally, a specification for an interactive web-based urban design tool is outlined, with which the trained optimiser could be employed to suggest well performing morphologies as inspiration to urban designers.

Table of Content

List of Tables

List of Figures

Declaration

I, the author, confirm that the Thesis is my own work.

I am aware of the University's Guidance on the Use of Unfair Means

Chapter 1 Introduction

1.1 Background of this research

Architecture design is a complex and holistic project and procedure which would impact the overall performance of the design (Shi and Yang, 2013). The holistic complexity involves multiple performance aspects, including not only energy cost, daylight, solar radiation, thermal comfort, environmental impact, but also global and initial cost, space allocation, logistics, structural assessment, etc (Ekici et al., 2019). The mentioned sub-categories of architectural performance, the criteria regard each field may conflict as in all multi-objective problems. So the ultimate goal of architecture design is to spot the balance solution satisfying most of all coherent and conflict objectives.

Design procedure is iterative by repeating the invention and revision cycle though a great many approaches developed along with the development of architecture design, for example sketching, physical modelling, digital modelling and performance simulations (Cobb et al., 2003; Ekici et al., 2019). However, the final design result will be highly affected by the decision made in early design stage (Ekici et al., 2019).

Hence balance and selection in early stage design regarding to multiple objectives are vital for a good holistic design project. This not only applies for sole building design, it appears more influential in building group or neighbourhood planning.

Under the context of growing attention on environmental performance in artificial building area, the overall performance of the environment rather than of single aspect of the performance becomes vital in assessing the quality of a space, not only for indoor but also for the outdoor scenario, although the design procedures.

However, sufficient works are done for the indoor environment, especially focusing on single domain simulations for task-orientated spaces. In the last decade, integrated

and global simulations are gaining popularities in the research field of indoor building environment but still not fully applied in design procedures of practice. As a comparison, the outdoor performance is less researched in terms of single domain simulation and is left barely done regarding multiple domain simulations, especially in explorative manner of non-task-orientated spaces.

To improve design procedure, simulation and optimisation are introduced to the design stages. But due to the high cost of simulation and optimisation in time, efforts and basement knowledge, only simple single-objective simulations are widely used in the early design stage of practice. Even single-objective optimisations based on corresponding simulations applied in early-stage practice are rare. Multi-domain multi-objective simulation and optimisation are still not sufficiently researched in architecture design and planning field in the academy, not to mention its usage in practice to guide early design stage decision making.

Therefore, this research selected multi-domain multi-objective simulation and optimisation applied in the early design stage for decision making supporting on the context of the residential ward outdoor environment as a breakout point. The residential outdoor area is a good example. It is worth studying for being heavily used by occupants but not specifically task-oriented and being possible of implanting recreational areas. Enhancing the residential outdoor area environment involves multiple performance requirements considered from various domains. A good balance of multiple environmental requirements namely of multi-domain objectives is needed all through residential ward planning. Locating this global optimisation proactively would produce the best overall performance and highest efficiency in framing and guiding the planning, which is vital in the residential ward planning and following architectural design.

1.2 Aims and Objectives

Based on the research background, this research focuses on three perspectives: early-stage decision making support for planning, simulation integration, and holistic design with global optimisation.

The aims of this research could be presented as:

1. To help on decision making support in early stage though providing a database of knowledge or guidance. These works are on the basis of the proactive simulations and global optimisation and interpretation of the feasible Pareto solution set in this research.

2. To implement the integration of multi-domain multi-objective performance simulation against the same design variable system, with help of various simulation tools. The design variable system applied in this research in the morphology parameter system of single building and neighbourhood.

3. To implement the holistic design of the outdoor environment of a residential ward by multi-domain multi-objective optimisation (MD-MOO). In other words, a series of balanced design schemes solving the conflictive requirements from multiple environmental objectives could be systematically searches and achieved.

To achieve the above aims, the following objectives are of interest in this research.

1. To explore the possible parameters of describing building distribution in a residential ward and to build the parametric system of morphology of single building and neighbourhoods for the following integrated simulation.

2. To explore the performance evaluative metric and its possible representative indices for three performance domain: traffic noise propagation, sunlight availability

and outdoor thermal comfort. Therefore, the selection of possible representative statistical indices and corresponding evaluative standards could be confirmed.

3. To explore the qualitative relationships between domain performance phenomenon and building distribution pattern and to extract qualitative design rules for building arrangement.

4. To explore quantitative relationships between performance representative indices and morphology parameters of single building and neighbourhoods and to extract numerical guidance for building distribution arrangements.

5. To train and achieve the meta-models for three domain performance, to enable the fast prediction on the overall domain performance level with the integration of all performance metrics in this domain.

6. To achieve multi-domain multi-objective optimisation based on the fast predictive meta-models and to collect the feasible optimised solutions for the three-domain performance design.

7. To acquire the suggestive design patterns and rules of balanced cross-domain performance based on the optimised solutions, to support early stage design in the future.

8. To build the quantitative database of the suggestive design schemes based on optimised solutions from the fast predictive meta-models and to present designer-friendly user interface for an easy application.

1.3 Framework of the Whole Research

The nature of this research is a multi-domain multi-objective optimisation on the basis of simulation analysis of three environmental performance fields. This research is composed of modules of building distribution parametric studies, single-domain performance simulation, simulation data analysis (qualitative and quantitative), single-domain multi-objective optimisation with corresponding meta-models, multi-domain multi-objective optimisation and interpretation of optimised Pareto solution set. The overall framework is indicated in figure 1.1.
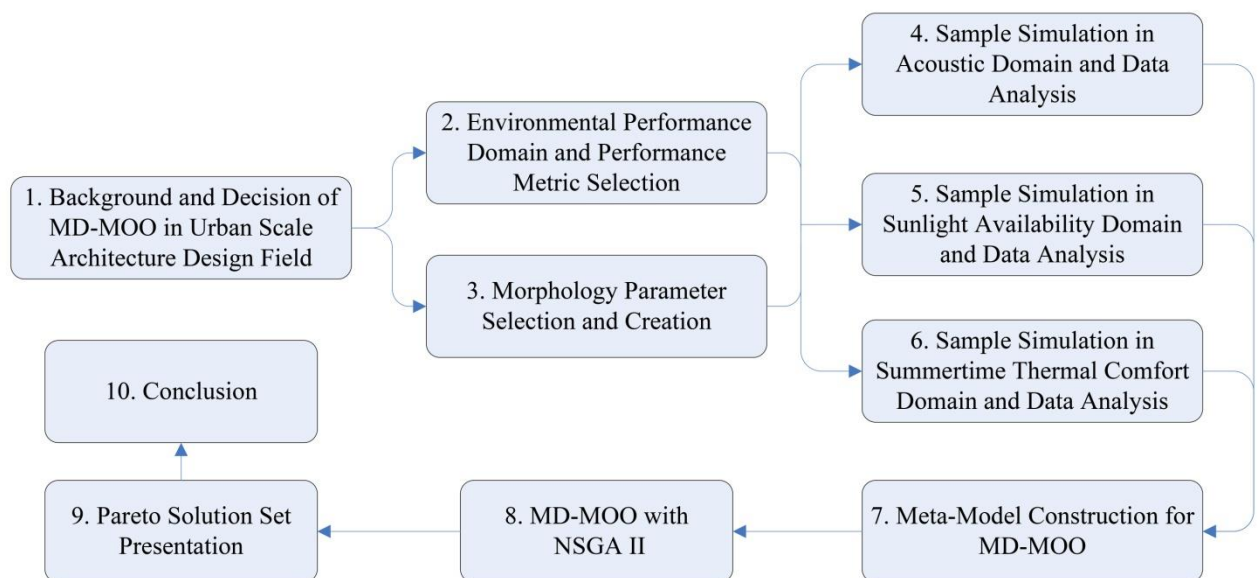
Figure 1.1 Overall Content Structure of the Thesis

Chapter 1 sketches the overview of this thesis from research background, aims and objectives.

Chapter 2 introduces from where and how this research starts, including 2.1 integrated environmental simulation and optimisation in early design stage for supporting decision making, 2.2 multi-domain multi-objective optimisation (MD-MOO); 2.3 MD-MOO application in architectural and urban design; 2.4 integrated environmental assessment system for multi-domain optimisation and decision of performance metric;

2.5 interest location and sample selection for multi-domain environmental performance simulation; 2.6 statistical methods applied in this research; 2.7 summary.

Chapter 3 expatiates the exploration and analysis over building distribution parameters, from aspects of morphology of single buildings and neighbourhoods. the structure of this chapter is: 3.1 definition and application background of individual and neighbourhood morphology parameters; 3.2 consolidation of morphology parameters for simulation; 3.3 data collection and transformation for individual building and neighbourhood; 3.4 cluster analysis for morphology parameters to define key parameters; 3.5 level grading for key morphology parameters; 3.6 correlation and interaction between key morphology parameter pairs and 3.7 summary.

Chapter 4 focuses on the acoustic domain performance simulation and the corresponding qualitative and quantitative analysis. Chapter 4 consists of 4.1 background of acoustic simulation; 4.2 comparison of the currently available acoustic simulation packages; 4.3 set up process of acoustic simulation model in CadnaA.; 4.4 sound source definition in CadnaA; 4.5 selection of representative performance indices for result assessment and analysis. 4.6 and 4.7 explain in detail about the qualitative and quantitative analysis of acoustic simulation results.

Chapter 5 focuses on the sunlight availability domain performance simulation and the corresponding qualitative and quantitative analysis, including 5.1 background of sunlight availability simulation; 5.2 comparison of simulation software; 5.3 simulation model setup; 5.4 selection of performance indices for sunlight availability; 5.5 and 5.6 qualitative analysis for APSH and WPSH simulation data; 5.7 comparison of qualitative analysis results from APSH and WPSH; 5.8 and 5.9 quantitative regression of APSH and WPSH against building morphology parameters; and 5.10 summary.

Similarly Chapter 6 focuses on the outdoor thermal confort domain performance

simulation and the corresponding qualitative and quantitative analysis, including 6.1 background of thermal simulation; 6.2 thermal simulation software comparison; 6.3 simulation model setup in Envi-met; 6.4 parametric study of MRT and WS; 6.5 qualitative analysis for MRT; 6.6 quantitative analysis for MRT; 6.7 qualitative analysis for WS; 6.8 quantitative analysis for WS; 6.9 qualitative and quantitative analysis for LRE; 6.10 integration and of regression results and 6.11 summary

Chapter 7 concentrates on the construction and training of the fast predictive meta-models for three domains. The content includes 7.1 background of application and construction of meta-model in MDO; 7.2 methodology of construction meta-model for md-moo with GWO-GRNN; 7.3 confirmation of morphology parameters and performance indices for SD-SOF for three domains; 7.4 sensitive analysis and selection of morphology parameters and performance indices for SD-MOO; 7.5 mathematical expression of single domain component objective function (SD-SOF); 7.6 mathematical expression of single-domain multi-objectives function (SD-MOF) and meta-model; 7.7 meta-model construction procedures through hybrid generic regression neural network (GRNN) with grey wolf optimiser (GWO);   7.8 meta-model prediction and assessment; 7.9 discussion on update of single-domain meta-models; 7.10 discussion on weight selection in meta-model construction progress; 7.11 discussion on sample size expansion by latin hypercube method and 7.12 summary.

Chapter 8 expatiates the progress of implication of global optimisation over the three domain performance objectives. The structure of this chapter is: 8.1 MDO strategy comparison and selection; 8.2 background and application of NSGAII in MDO; 8.3 methodology of NSGAII optimisation algorithm; 8.4 definition and mathematical expression of MD-MOO function; 8.5 workflow and setup of optimisation with NSGAII algorithm; 8.6 comparison and validation on searching space and iteration times and 8.7 summary.

Chapter 9 is the interpretation of the feasible Pareto solution set achieved from the previous MD-MOO of the three domain performance against the morphology parameters of building and neighbourhood. The structure of this chapter includes 9.1 data transformation for feasible Pareto solutions analysis; 9.2 analysis of global/integrated multi-domain performance improvement; 9.3 analysis of single-domain performance improvement; 9.4 suggested residential ward distribution based on feasible Pareto solutions; 9.5 validation of suggested design schemes; 9.6 interactive tool demonstrating scheme selection. 9.7 summary.

Chapter 10 is the conclusion chapter, summarises the achievement and suggestion for future studies.

Chapter 2 Research Background and Decision of Research Interest

This chapter is about the background of holistic design, early design stage decision making, proactive simulation and multi-domain multi-objective optimisation, as well as the research interested developed based on the above topics. The evaluative standards over the interested performance field are compared. Hence research questions are decided accordingly. The sampling method and sample information are presented, as well as the statistical tools used in this research.

The structure of this chapter is 2.1 integrated environmental simulation and optimisation in early design stage for supporting decision making; 2.2 multi-domain multi-objective optimisation (MD-MOO); 2.3 MD-MOO application in architectural and urban design; 2.4 integrated environmental assessment system for multi-domain optimisation and decision of performance metric; 2.5 interest location and sample selection for multi-domain environmental performance simulation; 2.6 statistical methods applied in this research; 2.7 summary. The flowchart of this chapter is shown in figure 2.1. It is the expansion of box 1 and.2 in overall content structure.
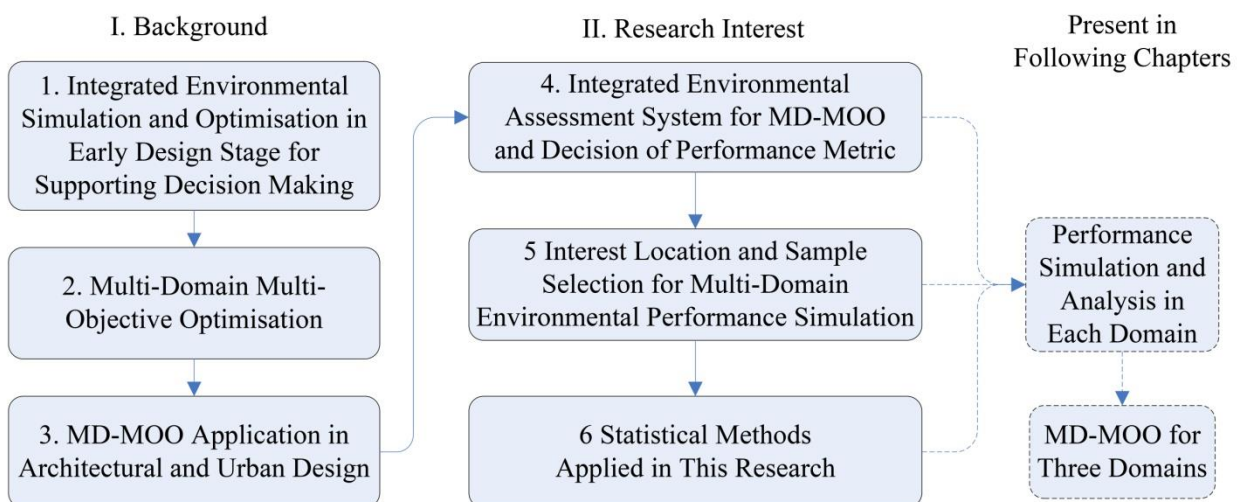
Figure 2.1 Content Structure of Chapter 2
Research Background and Decision of Research Interest

Acronyms for Chapter 2

ACO	ant colony optimisation
ADF	average daylight factor
AHP	analytical hierarchy process
ANN	artificial neural network
ANSI	American National Standards Institute
ANSI	American National Standards Institute
API	application programming interface
APSH	the annual probable sunlight hours
ASHRAE	The American Society Of Heating, Refrigerating And Air Conditioning Engineers
BA	bat algorithm
BIM	building information modelling
BPS	building performance simulation
Bre Br	Building Research Establishment Building Regulation
BS	British Standardisation
CAD	computer aided design
CECS	China Association For Engineering Construction Standardisation
CFD	computational fluid dynamics
CIE	International Commission On Illumination
CMA-ES	covariance matrix adaptation evolution strategy
CS	cuckoo search
DB	Chinese Regional Standards
DE	differential evolution
DoCE	design of computational experiment
EA	evolutionary algorithm
EC	evolutionary computation
Eia	The U.S. Energy Information Administration
GA	genetic algorithm
GB	Chinese National Standardisation (Guobiao)
GUI	graphical user interface
GWO	grey wolf optimiser
HDE	hybrid differential evolution
ISA	interior search algorithm
JG	Chinese Professional Standard Systems
KH	krill herd
L(A)eq	the A weighted equivalent sound level
LHS	Latin hypercube sampling
LRE	longwave radiation from environment
MBA	mine blast algorithm
MCDM	multi-criteria decision making
MD-MOO	multi-domain multi-objective optimisation
MDO	multi-domain optimisation
MFO	moth-flame optimiser

MLR	multiple linear regression
MODE	multi-objective differential evolution
MOEA	multi-objective evolutionary algorithm
MOF	multi-objective optimisation function
MOGA	multi-objective genetic algorithm
MOO	multi-objective optimisation
MOPSO	multi-objective particle swarm optimiser
MOSOS	multi-objective symbiotic organism search
MRT	mean radiant temperature
NSGA	non- dominated sorting genetic algorithm
NSGAII	non-dominated sorting genetic algorithm with elitist strategy
OAT	one-at-a-time approach of simulation
PBD	the performance-based design
PCA	performative computational architecture
PDD	predictive percentage dissatisfied
PMV	predictive mean vote
PSO	particle swarm optimisation
RH	relative humidity
SD-MOO	single domain multi-objective optimisation
SHW	accumulative sunlight hour in winter
SI	swarm intelligence
SOF	single objective optimisation functions
SOO	single objective optimisation
SPEA-2	strength pareto evolutionary algorithm 2
SPL	sound pressure level
SVF	sky view factor
SVM	support vector machine
T_{db}	outdoor dry bulb temperature
UDI	useful daylight illuminance
UHI	the urban heat island intensity
VIF	variance inflation factor
VSC	vertical sky component
WPSH	the winter probable sunlight hours
WS	wind speed
ZPRED	standardised predicted value
ZRESD	standardized residuals

2.1 Integrated Environmental Simulation and Optimisation in Early Design Stage for Supporting Decision Making

This research is aiming to facilitate the experience in early design stage of urban planning in residential background, by involving integrated performance simulation and optimisation. The framework allows designers and planners to consider and balance multi-domain performance objectives, to search the available design space for global optimal solutions and to guide design accordingly.

2.1.1   Holistic Design in Architectural Domain

Conventionally, architecture design and urban planning are empirical works, conducted based on intuition, know-how and judgement, in which decision making highly relies on existing design habits and experiences. The reveal the fact that the traditional design process requires supports from numerical assessment of design and systematic exploration of the design schemes in feasible design space, in order to verify the design's the capability of satisfying multi-disciplinary performance requirements (Ekici et al., 2019). Even conventional design procedures with optimisation studies are based on trial and errors, to test a limit numbers of the subjectively conceived potential alternatives using simulative and numerical tools (Kämpf et al., 2010) .

The performance-based design (PBD) becomes of growing importance as a design concept and approach to fulfil multiple design objectives (Kolarevic, 2004). Followed by Sariyildiz's presentation in 2012, the performative computational architecture (PCA) is a support tool for design process to achieve most desirable combination of design-related parameters to meet performance-related goals in the conceptual design stage (Sariyildiz, 2012). It is the application of holistic design in the architecture domain. It includes the three phases in a holistic design, which are form generation,

performance evaluation and optimisation (Sariyildiz, 2012).

The form-finding refers to the form generation in the iterative design process. The performance evaluation is the performance objective assessment of the design scheme generated by form-finding. The optimisation uses multiple searching approaches to select optimal design solutions in a systematic method. The three stages of PCA are iterative until required design achieved.

Form-finding is initially applied in structural design for example shell design. It is nowadays refers to the exploration progress in architecture design to fulfil predefined performance objectives through computational optimisation approaches, which aims to inform designer with sufficient decision making basis (Ekici et al., 2019). Initial forming finding in early design stage is crucial as it affect final performance greatly, as it is the basement and input data for all subsequent design process all through building's life-cycle (Ekici et al., 2019). With the development of numerical prediction and assessment theories and tools, performance simulation and evaluation is operable in all stage of design process now. But the numeric assessment and guidance in conceptual design is still not sufficient and widely used in current practice (Ekici et al., 2019). The priori knowledge of a design in early stage is highly dependent case-by-case, and especially in short when innovation is involved (Ward, 1969).

Performance evaluation is usually operated in form of physical measurement or computational simulation. Nowadays computational simulation is widely accepted in educational and practice contexts, as one iterative step of the whole design process. The optimisation stage is based on the data collection from the evaluation stage. For architectural context, optimisation are only applied in a few researches rather than widely use in real practice.

Metaheuristic optimisation methods are the most widely adopted approaches in PCA.

They share advantages of 1. Fitting the design process of PCA; 2. Capable of handling both continuous and discrete parameters; 3. Capable of handing large design space; 4. Good at avoiding local optimal; 5. Capable of presenting near optimal solution sets; 6. Reasonable calculation time (Ciftcioglu, Sariyildiz and Bittermann, 2007; Evins, 2013; Machairas, Tsangrassoulis and Axarli, 2014).

Because in architecture design and planning numerous design variables and performance indicator parameters are involved all through the design process, interactive or conflictive parameters need to be equilibrated in design, namely holistic design is essential. As stated in previous studies, there is a clear need of a system providing simultaneous assessment on multiple objectives in conceptual design stage for designer. However, a balance conflictive performance requirement in holistic design is not only difficult all through design process, but especially facing barrier due to uncertainty of identified design parameters and objective in early design stage. For instance, many performance simulations at early stage are not possible without room metrics specified. The two available approaches of holistic design for building simulations are operating multiple disciplinary simulations crossing various software with interoperability or in one integrated simulation platform; or collecting and analysing separated simulation results to aid multi-criteria decision making (MCDM). The weakness of applying holistic design in context of building design is the possibility of unable to achieve single optimal solution or even optimal solution group, due to the high-dimensional nature of the design question.

2.1.2   Decision Making Supporting in Early Design Stage

With the fast improvement in calculation speed and capacity, various performance simulations become the most persuasive supporting tool for aiding decision making in the design process. However, so far the applications of simulation more often concentrate on scheme evaluations and comparison in the late design stage. The whole design process is still following a loop pattern of trial-and-error manner.

Simulation application in early design stage is still limited in spite of the positive and efficient impact on final result from a well organised early design. Multi-criteria decision making (MCDM) is widely used for complex system involving multiple domains and conflicting criteria. It is an analysis method to compare and rank available solutions for a complex system as building related design. The most widespread approach is assigning weighting for various solutions to allow comparison. Analytical Hierarchy Process (AHP) is the most accepted approach applied in building construction field, mainly on building structure field to reduce uncertainty and assess reliability of the solutions. This research provides the available Pareto solutions out of possible design space of the multi-criteria problem, which could help on future MCDM process.

In early design stage, support on decision-making and guidance on design arrangement is vital due to its impact efficiency on final result of performance or cost (Attia et al., 2012; Hygh et al., 2012). Adverse decision may result in reduced and limited design space for further improvement and optimisation (Østergård, Jensen and Maagaard, 2016). To guide the design is considered highest priority of simulation software to major architect (Østergård, Jensen and Maagaard, 2016). According to interviews of 230 architects, characteristics of timely feedback on performance of design and capability of ranking multiple design variations, referred as intelligence and usability by architects, are most preferred than characteristics of accuracy and interoperability when selecting building performance simulation (BPS) tool (Attia et al., 2012). However, simulations applied in design procedures are more for assessment and evaluation of schemes, for instance code compliance, quality control, etc., rather than proactive studies (Attia et al., 2012; Kanters and Horvat, 2012). Among the building simulation tools listed by US Department of Energy, Less than 8% of over 400 has early design enhancement potential (Batueva and Mahdavi, 2014) and only 1% out of 392 are pre-design informative.

The major challenges of application of simulation in early design stage are stated by

Østergård as lack of information, input uncertainty, vast design space, increasing level of model resolution, time-consuming modelling and rapid change of design. According to the literature reviews of previous works and design progress in industry, the challenges of simulation application in early design stage as a support tool lie in the following aspects.

First, high level of uncertainty and variability exist in the draft scheme, input variables for simulation and even performance objectives for assessment. For some cases, at early stage the aim of the design in terms of performance is not decided, hence so as the appropriate assessment objectives and corresponding inputs. Other conditions with clear design performance objectives, the concept designs may provide less details to valid the requirements of a simulation. For instance, scale of room and building, weather data, user's metabolic condition and preference.

Second, the current available building performance simulation could only evaluate one solution in available design space at a time, namely conducted by one-at-a-time approach (OAT). The one-at-a-time approach (OAT) of simulation leads to few alterations on design variables once a run, and it barely provides direct guide for designers on scheme enhancement (Østergård, Jensen and Maagaard, 2016). Furthermore, there is vast design space at early design stage. Hence, conventional building simulation is hard to effectively search or test the design space and guide directions of design enhancement, which makes it difficult to be used in early design stage as an efficient explorative tool for solutions.

Third, currently simulation software are not fully integrated which allows considering of multi-disciplinary objects in one go, in other words, a lack of capacity in simulation multiple objectives in multiple domain simultaneously. This not only exists in indoor context simulation, but also is a significant gap in outdoor simulation context. For indoor context, the current integrated works concentrate on building energy consumption and cost combined with thermal and lighting performance, while

acoustic performance is usually considered and evaluated in late design stage. As a comparison, researches of integrated simulation in outdoor context are much less in total amount, not mentioning its application in industry. The majority of works in outdoor contexts are single disciplinary simulation regarding to thermal comfort, artificial lighting or noise propagation.

Except for the above mentioned challenges, others have been mentioned by Østergård, et al. as first, lack of continuity of modelling with increasing resolution and interoperability among multiple simulation software from early stage to later stage; second, lack of adoption of existing knowledge in guiding further simulations; and third, long modelling and simulation time relative to early design stage (Østergård, Jensen and Maagaard, 2016). Simultaneously, it is also mentioned the difficulties of performing simulation and optimisation though all stages of building design are stated as limited use of knowledge, lack of simulation guidance, contradictive and strict requirements, interoperability between software, discrepancy between simulation and real-life measurements

It is noted in the review work, that six approaches aiding early design stage simulation or all process simulations could facilitate the challenges aforementioned (Østergård, Jensen and Maagaard, 2016). They are knowledge-based approach, proactive simulation, statistical method, computer aided design (CAD)-building performance simulation (BPS) interoperability, holistic design and global optimisation, etc.

The knowledge based approach refers to through application of existing knowledge including design rules, experiences and previously done simulations, the preset variables and default value combinations of new simulation could be decided. Knowledge bases approach could improve simulation consistency and validity. As shown in multiple researches, statistical approach is more widely applied on uncertainty analysis, sensitivity analysis and multivariate analysis. Proactive simulation and global optimisation would be explained in the following sections.

2.1.3 Proactive Simulation Approaches

To response to the issues of iterative, time-consuming, lag behind enquires and evaluative nature of simulation, some concept are arise. Pre-design informative building performance simulation (BPS) is mentioned referring to operating simulation prior to the early design stage and forming databases accordingly of predefined schemes to support of decision making in early design (Attia *et al.*, 2012). Another similar concept of proactive simulation refers to a series of structured organised simulations that aiming to explore available design space to guide later designs rather than to evaluate existing design scheme (Østergård, Jensen and Maagaard, 2016).

Exhaustive simulation before design and filtering based on criteria on results is one of the concepts for proactive simulation. It requirement calculation support, i.e. cloud computing. The strength of exhaustive simulation is a sufficient cover and exploit of design space and parameter combination, on the other hand the corresponding weakness is supper labour and time consuming when modelling and calculation. Large scale and comprehensive parametrical modelling and cloud computing is the key techniques to allow the easy access of this approach.

According to the comparison of two concepts, the framework of this research is a combination of the two. The prepositioned simulations of existing projects and design rules database are pre-design informative BPS, while the following multi-objective optimisation and the suggested optimisation looping based on automatic parametric modelling and Latin Hypercube Sampling could be categorised as proactive approach.

As shown in previous researches, to support early design stage there mentioned a tool which is capable of parametric, room-level simulations to generate input for overall building design process before any design decisions, in terms of integrated aspects as energy consumption, air quality, daylight and thermal comfort (Petersen, 2011). An advice tool aiding intelligent facade design at conception design stage for further

continuity in preliminary and detailed design stage, in terms of energy and visual comfort based on imbedded EnergyPlus simulations is also mentioned (Ochoa and Capeluto, 2009). Similar as in this research, the tool presented by Ochoa et al. extracted the simulation inputs in EnergyPlus into architectural languages, for instance, occupancy level, facade openness or building depth, etc. The substitution of large amount of input variables by few design properties not only enables easy understanding rules of design to be generated for architects to guide further simulations, but also enable creation of design alternatives with the tool though complying with the rules. This coincidence demonstrates the need of pre-digestion of complex inputs in simulation to few design quality descriptions from architect view point.

A combination of large amount of structured proactive simulations and statistical analysis on results is named as statistical approach of systematically exploring design space and providing design support in previous work (Østergård, Jensen and Maagaard, 2016). It refers to an expansion of a baseline simulation model by systematically sampling in the design space formed by all simulation inputs parameters, under guidance of the probability density functions of the parameters. Multiple sampling rules are available, including the Latin hypercube as discussed in this research. The results of expanded simulations are analysed to form general rules or support experiment design in simulation, by uncertainty analysis, sensitivity analysis, multivariate analysis and meta-model creation, etc. This approach help reduce uncertainty and variability of simulation in early design stage. It is more efficient than one-at-a-time simulation with only few adjustments in input variables. The structured exploration may enlarge the solution space through widening the possible span of input variables and statistical analysis presents the possible ranges of input variables.

Sensitivity analysis is applied in multiple studies to decide the most influential simulation input parameters on building performance. Sensitivity analysis applied in

building energy context is broadly reviewed by Tian. Multiple methods could be applied to discover relationships of input variables, i.e. multivariate regression used in this research.

2.1.4   Integrated Simulation Tools

As classified by Petersen (Petersen, 2011) and Citherlet (Citherlet, 2001), the combination of CAD platform with BPS unit include several methods of integration, run-time interoperability, file exchange, and standalone. The integrated method refers to BPS function imbedded in CAD environment, for example solar analysis in various modelling software. The run-time interoperable combination is the most expected method by designers from industry, which refers to access BPS through plug-ins or (application programming interface) API from modelling tools allowing run-time or concurrent simulation following the model editing. Examples of this kind are Grasshopper and Dynamo plug-ins, SketchUp and Revit with API of Sefaira and OpenStudio, which are capable of multiple light, solar, thermal simulations both for indoor and outdoor environment. The file change method is as used in building information modelling (BIM), to connect model and simulation through readable and writable file exchange. The last standalone method is most widely used in research context that the building simulationist remodels the select building or room in independent accurate simulation software, i.e. EnergyPlus and Envi-met for thermal performance and energy consumption; Radiance for lighting condition; CadnaA for acoustic performance.

A variety of software vendors provides building performance simulation (BPS) though add-ons or dynamically coupled engines over modelling platform, for instance, Autodesk's Green Building Studio for Revit, EcoDesigner Star for ArchiCAD. The application programming interface (API) of a third party vendor is also linked to modelling tools to enable BPS, for instance Sefaira, IESVE, OpenStudio linked to SketchUp. The simulation engines adopted in various plug-ins and APIs include

EnergyPlus, Daysim, and Radiance. This coupled approach is most widely applied in industry for late stage simulations; however for early stage applications it has several limitations: the interoperability of file format crossing software, heavy computations of detailed simulation engines and large amount of inputs due to multi-objective context. The availability of multi-objective simulation relies on the incorporation of multiple BPS over CAD platform by software vendors, while some research communities also make contributions. For example, multiple engines are imbedded in Ladybug and Honeybee as plug-ins for parametrical modelling tool Rhinoceros with Grasshopper. This is the currently most promising package suitable for multi-discipline simulations in early design stage, because of the simplified algorithm, fast calculation and availability in real-time simulation for some objectives. However, its limitation lies on further validation required and not suitable for detailed analysis.

To provide easy-accessed interactive tool for real-time optimisation result for practitioners and researchers is essential. In this research, multiple linear regression and meta-model construction respond to the limitation of uncertainty and vast calculation. The interactive tool developed with imbedded optimised Pareto solution database could be application medium for users to skip the formulation of problem and selection of optimisation method. The global optimisation system established in this research could be used as template to format other multi-objective optimisation and decision making problems in building design and planning.

2.1.5   Summary

This section describes the background and previous works of holistic design of integrated environmental domains as the support at early design stage. The concepts of holistic design, decision making support at early design stage, proactive simulation at early design stage and possible integrated simulation tools are explained. The necessity and challenges of implementation holistic design from the early design stage is expatiated.

This research is attempting to approaching the goal by response to the aforementioned challenges. By involving proactive simulations of real projects at multiple assessment level, the existing design knowledge and local design habits are extracted as guidelines for further simulation, which is close the mentioned knowledge-base approach. Statistical method are used to analysis proactive simulation to aid exploring possible simulation input variables and reduce uncertainty at early stage simulation. The simulations on multiple domains are conducted over various standalone software to avoid interoperability among multiple simulation software. The multi-objective optimisation based on multi-domain simulations focuses on the systematic search on all feasible solutions in design space. A fast responding meta-model is constructed for the optimisation to replace the high accuracy simulation procedures in early stage to avoid over-time-consuming modelling and calculation. Rules and patterns extracted from feasible optimal solutions could directly guide further designs and simulations, through narrow down design spaces and offering extensive design suggestions for designers before any model had been built.

2.2 Multi-Domain Multi-Objective Optimisation (MD-MOO)

Real optimisation problems are usually nonlinear and with multiple objectives. When dealing with multi-objective optimisation (MOO), the optimisation calculation for each objective may have coherence and confliction. Hence, multi-objective optimisation algorithm with capability of global optimal solution gains firm attention in nowadays.

2.2.1   Definition of Multi-Domain Multi-Objective Optimisation (MD-MOO)

Multi-domain optimisation is a methodology of designing complex system and sub-system through excavating and applying the cooperative relationships among multiple domains to achieve optimal solutions of the whole system or optimised

satisfying solution for practice. The core of multi-domain deign is the comprehensive consideration of cross-domain cooperative and conflictive relationships and global performance optimisation of the holistic system.

A MD-MOO is implemented through the approximate optimisation strategy based on design of computational experiment (DoCE) and meta-model. The basic idea of the approximate optimisation strategy is to sample enough data with assistance of DoCE, and to construct meta-model using the sampled data, then to result in fast convergence of achieving global optimum by adopting and analysing the meta-model.

As in this research, the three involved research fields are domains of architectural design discipline. Hence, here we name MDO as multi-domain optimisation instead of the original term. The multi-domain optimisation (MDO) of three environmental performance domains would be executed through multi-domain multi-objective optimisation method (MD-MOO) as the global optimisation system. Sub-system optimisation, namely single domain multi-objective optimisation (SD-MOO) in one field of the acoustic, sunlight availability or thermal comfort could also be execute before the global MD-MOO, but not applied to this research. Implement of MD-MOO is using approach of solving multi-objective optimisation function (MOF), which could be expressed as a calculation of searching minimum of a MOF under a series of constraint functions and within the design space. The MOF would be constituted by multiple single objective optimisation functions (SOF), which would presents cooperative or conflictive trends during solving minimum, and constraint array of all independent variables (x) and dependent variables (y). One SOF, for example the function of sunlight hour index, is the component function of a MOF or single domain performance metric, for example sunlight availability metric. In SOF and MOF, the independent variable (x) is the performance parameters, in other words the individual and neighbourhood morphology parameters; the dependent variable (y) is the component parts of performance metric, name it as performance index.

A MD-MOO is consisted in four main parts from view of knowledge content: optimisation strategy, sampling method, meta-model construction, and optimisation algorithm.

Optimisation strategy refers to the optimisation design of the sub-systems. Its aim is to simplify the MD-MOO system from structure viewpoint. It includes the division of sub-systems from a comprehensive practical problem, coordination of the sub-system progress, reduction of communication and calculation load and comprehension of optimisation results of individual sub-system for holistic optimisation calculation. Decoupling method is aiming to divide the whole system into sub-systems in order of convenient individual analysis within each domain field. The main division strategies include hierarchical, non-hierarchical and mix layer system. In hierarchical system, data deliver from the parent layer to the offspring layer uni-directionally, namely the offspring system only receiving data from parent system without data exchanging between peer systems. On the contrary, in non-hierarchical system sub-systems are paralleled and capable of data exchange with each other. In this research, hierarchical system is adopted. The acoustic, sunlight and thermal performance optimisation systems are considered independent sub-systems. While the integrated environmental performance system is the parent system required holistic optimisation.

Meta-model is a fast approximate model by estimate a mathematical relationship of input and output parameters, as a substitution of calculation heavy models. It is aiming to reduce calculation scale during MOO of the whole system. It is essential because the complexity of each of the prediction models from multiple research fields involved in MD-MOO may fail the global search calculation and iteration, due to excessive calculation load. Building performance simulation tools are mostly of high simulation accuracy and time or effort consuming, especially for thermal simulations (Østergård, Jensen and Maagaard, 2016). This characteristic of fast calculation of meta-model, namely quick feed back tends to be welcomed in early design stage. Meta-model could be established based on experiment or observational data, building

performance database, or valid simulated data from multiple detailed simulation software. Meta-model is constructed by regression and fitting method. Various techniques could be used to construct meta-model, for instance Artificial Neural Network (ANN), support vector machine (SVM), Kriging, multivariate linear regression, etc. Once achieving a meta-model, its prediction ability will be assessed to the original simulation or prediction model. Previous works of meta-model construction in early design stage had involved performance indicators about daylight factor, heating/cooling load, thermal comfort, indoor air quality and net cost, etc. The shortcoming of meta-model is it is established for specific problem or relationship, which means not interoperable and interchangeable to other similar questions. Although mentioned meta-model training for each performance indicator could be laborious (Østergård, Jensen and Maagaard, 2016), combining meta-model with multiple objective optimisation could overcome this limit by involving few global objectives constructed by several performance indicators.

Optimisation algorithm is the specific mathematical algorithm dealing with searching for optimal solutions of single or multiple objective functions in a specific range in the holistic or sub system. If solving single objective function (SOF) for optimisation, it is referred as single-objective optimisation algorithm; if solving multiple objective functions (MOF) simultaneously for optimisation, it is referred as multi-objective optimisation algorithm. An optimisation algorithm is consisted of procedures of searching in design space, iteration and assessment of convergence.

2.2.2   Workflow of Applying Multi-Domain Multi-Objective Optimisation (MD-MOO)

The general procedures of a MD-MOO are as below and shown in figure 2.2.

1. Based on requirement of each domain, adopt appropriate simulation models to accomplish analysis with consideration of accuracy and validity. Then construct the optimisation model of the multi-domain question with clarification of

objective function, constraint function, design variables, design space.

2. Acquire homogeneously distributed sampling in initial design space, though DoCE method.

3. Calculate the real response value of the initial and expanded sample points from the high accuracy models.

4. Construct meta-model in the design space using the abovementioned sample points and their real response values.

5. Choose approximate meta-model constructive strategy: static or dynamic.

6. Examine the accuracy of meta-model, if satisfied go to 7, if not go to 8.

7. Execute optimisation calculation based on the meta-model, to search for optimums. Output the approximate optimum as optimum of the original objective function.

8. Expand sample size by DoCE to update meta-model until accuracy requirement is satisfied. Go to 3 or 4.

9. Examine the validity and feasibility of the optimal solutions whether satisfying constraint conditions and variable bounds of the multi-domain optimisation (MDO) question, if yes, save the approximate optimum as final optimum, if no, repeat from 1, with adjusting on the construction of meta-model and optimisation algorithm.
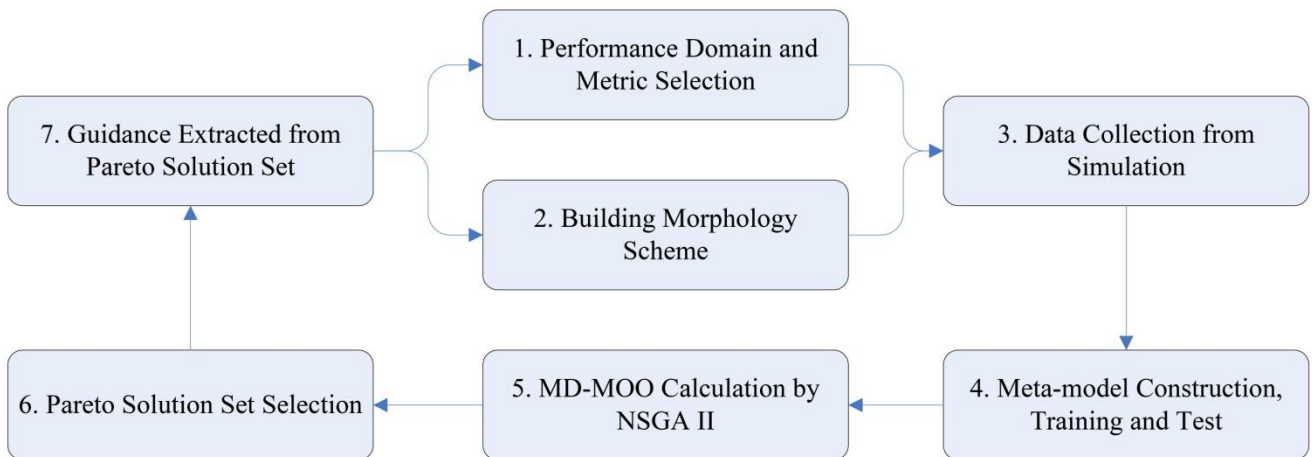
Figure 2.2 Design Workflow with MD-MOO

2.2.3   Previous Development of Optimisation Algorithm

Conventionally, since its development, the real world design problem optimisations applied gradient descent optimisation algorithms. Its basic idea is to calculate the derivative of the formulation of the optimisation objective. Hence the limitation of gradient descent optimisation is suffered from local minima stagnation (Kelley, 1999) and dependence on initial solution. Also it is more time consuming. The second phase of optimisation algorithm development is of population based stochastic algorithms. Its basic idea is to obtain multiple numbers of solutions at a time, to involve the features of local minima avoidance. It is calculating the maximisation or minimisation of the single objective function of the design problem. Under metaheuristic optimisation method, swarm intelligence (SI) and evolutionary computation (EC) are two natural inspired searching strategies.

The mechanism of swarm intelligence (SI) is to mimic the interaction of the individual in a society with the rest individuals and environment. 1. Define the behaviour rules of the individual; 2.   Generate overall behaviour through interactions between individuals in the swarm system. The swarm society could simulate the group behaviour of ants, wasps, termites, bees, fishes, birds, land animal herds like wolf packs (Ekici *et al.*, 2019). Most accepted SI algorithms are particle swarm optimisation (PSO) first proposed with two paradigms (Eberhart and Kennedy, 1995), ant colony optimisation (ACO) first presented with three paradigms (Dorigo, Birattari and Stutzle, 2006).

The mechanism of evolutionary computation (EC) is 1. Encode a D-dimensional individual into chromosomes to form initial population as the first parent population; 2. In each generation, the individuals in parent population will be mated, in a cross over manner to form the offspring or children generation; 3. Certain proportions of the individuals in the offspring will be mutated to generate randomness to escape from local optimal; 4. The offspring is combined to the parent to form a new parent

generation for the iteration. Some algorithms of EC are widely accepted and verified. The Genetic Algorithm (GA) is firstly presented with machine learning for multiple application example (Goldberg, 1989).The Differential Evolution (DE) is first presented for what question (Storn and Price, 1997).

Some widely used algorithm of single objective optimisation (SOO) are Moth-Flame Optimiser (MFO) (Mirjalili, 2015), Bat Algorithm (BA) (Yang, 2010), Particle Swarm Optimisation (PSO) (Kennedy and Eberhart, 1995), Ant Colony Optimisation (ACO) (Dorigo, Birattari and Stutzle, 2006), Genetic Algorithm (GA) (Holland, 1992), Cuckoo Search (CS) (Gandomi, Yang and Alavi, 2013), Mine Blast Algorithm (MBA) (Sadollah *et al.*, 2013), Krill Herd (KH) (Gandomi and Alavi, 2012), Interior Search Algorithm (ISA) (Gandomi, 2014), etc. There are reviews stating advantages of these algorithms with variety of operators as capable of handling uncertainties (Beyer and Sendhoff, 2007), local minima (Knowles, Watson and Corne, 2001), misleading global solutions (Deb and Goldberg, 1993), better constraints (Coello, 2002), etc.

Widely accepted multi-objective optimisation algorithms includes Multi-Objective Particle Swarm Optimiser (MOPSO) (Coello, Pulido and Lechuga, 2004), Non-dominated Sorting Genetic Algorithm (NSGA) (Deb *et al.*, 2000, 2002; Deb and Goel, 2001), Non-Dominated Sorting Genetic Algorithm II (NSGAII) (Deb *et al.*, 2002) and Multi-Objective Symbiotic Organism Search (MOSOS) (Panda and Pani, 2016).

2.3 MD-MOO Application in Architectural and Urban Design

2.3.1   General Trend of Optimisation in Architectural and Urban Design

Since 2000, there is a boost of publications related to building optimisation (Evins, 2013; Machairas, Tsangrassoulis and Axarli, 2014; Nguyen, Reiter and Rigo, 2014; Østergård, Jensen and Maagaard, 2016) It is reviewed that majority of the

publications about building optimisation are related to HVAC and control (Machairas, Tsangrassoulis and Axarli, 2014). For early stage building design, the optimisation applied are mainly of single objective, while the multi-objective optimisation is gradually becoming the trend (Evins, 2013). In most researches about multi-objective optimisation in building design field, two objectives are typically adopted, yet a few uses three objectives (Chantrelle *et al.*, 2011).

Optimisations about influential parameters on building design also gained popularity in recent years. It is mentioned that the aim of building design optimisation is not searching for best solution, but to explore the design space for alternative solutions and preferable design space (Attia *et al.*, 2013), or in other word parameter variations (Østergård, Jensen and Maagaard, 2016). Few studies are focusing on the simulation-based optimisation method (Nguyen, Reiter and Rigo, 2014).

A good many of works are based on single building performance and optimisation, focusing on building envelop forming (Huang and Niu, 2016; Kheiri, 2018), energy efficiency (Attia *et al.*, 2013; Shi *et al.*, 2016; Cui *et al.*, 2017; Eltaweel and Su, 2017; Kheiri, 2018), certain physical domain performance, sustainable design in various research branches (Evins, 2013; Tian *et al.*, 2018), spatial plan efficiency (Dutta and Sarthak, 2011).

A few of work adopted multi-objective optimisation in urban form fields in the recent decades. One work reviews simulation-based generation and optimisation applied in energy-driven urban design (Shi, Fonseca and Schlueter, 2017), which is the most close research field to this research. Multi-genomic algorithms is applied in form definition of the non-standard from building consisted with triangles. This form allows the optimisation to circumvent the limit of design variables, and introduces new decision-making variables (Ngo and Labayrade, 2014). A collaborative interactic genetic algorithm is applied in a floor plan creation, with consideration os creative design space, design space exploration, design representation, design evaluation,

design collaboration and visaulisation for interactivity (Banerjee, Quiroz and Louis, 2008). In the interactive tool EcoGen developed by Marsault, etc., the multi-objective interactive genetic algoritm is applied for efficient and diersified optimisation solutions, adaptive cross-over mutation and seclection (ACROMUSE) multi-objective extension is for sovling constraints. It is a multi-objective genetic local search (MOGLS)-base method, stated simpler to apply than NSGAII-based mehod (Marsault, 2017). The latest version of EcoGen2.1 as a tool of generative eco-design, is capable of generate one building form or group of buildings according to various critera in solar energy, thermal consumption and luminous comfort. The difficulities lie on the realtime and interactive response to gernerate proposals in the early design stage (Marsault and Torres, 2019). Evolutionary algorithm is encouraged in this work to use in an expliratory process in architrecure design solution finding. It shares the similarity view as in this research that conventional design process is typologocal based; in which designers tend to limit the design space due to the design methodology inherent from topology-based knowledge and requirement. Therefore exploratory precess would provide potential new set of topologies still satifying the constraints rather than fixed to existing design pattern (Nguyen *et al.*, 2016). In some cases, genetric algorithm is attempted as tool of form finding in generative design. However, the scale is still limited to abstract blocks in relative small scales (Navarro-Mateu and Cocho-Bermejo, 2020).

Despite the popularity of optimisation method applied in building design field, its adoption in practice is limited due to time-consuming calculation, inability of coping with uncertainties, knowledge requirements for optimisation problem formulation and algorithm selection, and the error-prone bridging between optimisation tool and simulation software (Attia *et al.*, 2013). It is also reviewed that the expected features of building performance optimisation from researchers and practitioners focus at easy Graphical User Interface (GUI), parallel computation, and simulation and optimisation tool bridging for real-time optimisation (Attia *et al.*, 2013).

2.3.2   Application of Optimisation in Daylight and sunlight

Considering from building facade design viewpoint, Genetic Algorithm (GA) is applied to optimise the daylight factor and solar incidence and structural weight of a roof (Ciftcioglu, Sariyildiz and Bittermann, 2007), and applied for optimise the daylight uniformity ratio within a gallery building (Rakha and Nassar, 2011). GA is applied for trade-off multiple illuminance objectives and multi-objective micro-GA is applied for optimising illuminance and glare objectives (Gagne and Andersen, 2012). Daylight and solar radiation is optimised integrated (Turrin, von Buelow and Stouffs, 2011).

Consider building facade design, a self-adaptive Differential Evolution (DE) algorithm is used in maximise solar irradiation (Bizjak, Marko ; Žalik, Borut ; Lukač, 2015). GA is used in an optimisation of building shadow area against building floor area, considering from aspect of environmental impact (Huang, Chang and Shih, 2015). The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and Hybrid Differential Evolution (HDE), the two algorithms are used to maximise solar potential (Kämpf and Robinson, 2010). Multi-objective Genetic Algorithm (MOGA) is used for maximising total radiation, considering building shape coefficient for one community building (Zhang, Zhang and Wang, 2016).

The Evolutionary Algorithm (EA) is used to optimise solar radiation in summer, equinox and winter in an urban context (Vermeulen *et al.*, 2015). It considering the periodic urban fabric may help in solar simulation review. PSO is used to optimise solar gain and residential building area in urban context (Liu, Liu and Duan, 2007). GA is used to optimise solar gain in summer and winter in urban context (Oliveira Panão, Gonçalves and Ferrão, 2008). GA is used to maximise solar energy received by buildings in an urban context, optimise against building amounts (Conceição António, Carlos A ; Monteiro, João Brasileiro ; Afonso, 2014). EA is used to maximise solar energy received by buildings for high-rise buildings (Vermeulen *et al.*,

2015). GA is also used for minimise the solar radiation, against residential building block layout (Kyu and Kim, 2015).

Some outdoor environmental performance like covered outside space, block ventilation, unit ventilation, outdoor solar radiation, unit solar radiation, circulation and unit count are considered from an viewpoint of environmental impact aspect in optimisation (Menges, 2012).

2.3.3   Application of Optimisation in thermal comfort domain

TRNSYS Simulation-based Artificial Neural Network (ANN) with Non-Dominated Sorting Genetic Algorithm with Elitist Strategy (NSGAII) is used for minimising the average absolute thermal comfort and annual energy consumption for single building facade design (Magnier and Haghighat, 2010). NSGAII is used for balance for indoor thermal comfort and annual primary energy consumption for single office building with adaptive facade design (Huang, Chang and Shih, 2015). NSGAII with ANN is used for optimise indoor annual thermal comfort and energy demand taking one residential building as example (Gou *et al.*, 2018). NSGAII is used for balance of percentage of thermal discomfort hour in residential buildings with its annual energy consumption for a prediction tool (Yu *et al.*, 2015). NSGAII is used in single-objective optimisation of discomfort hour in residential buildings against facade overhang design (Sghiouri *et al.*, 2018). The non-dominated sorting genetic algorithm II (NSGAII), multi-objective particle swarm optimisation (MOPSO), the multi-objective genetic algorithm (MOGA) and multi-objective differential evolution (MODE), are used and compared in optimisation of multiple objective thermal problem for single residential building sample, including total percentage of cumulative discomfort time, life-cycle cost and carbon dioxide equivalence (Li *et al.*, 2017). SPEA-2 is used for optimise the summer discomfort time and energy usage and useful daylight illuminance (UDI) for one school building (Zhang *et al.*, 2017). MOGA combined with fitness functions from multi-linear regression (MLR) and

ANN approaches is used for optimise thermal comfort and energy consumption for the facade of a single building (Lin *et al.*, 2018).

Multi-objective optimisation based on simulation of incident irradiation with Radiance against building geometry and solar orientation using evolutionary algorithm over Matlab. For cold regions requiring long heating season, one of the compounded objective functions is to minimise thermal energy consumption could be approximated as maximising incident solar irradiation with consideration of the offset by thermal loss. This is compounded objective could provide a readily interpretable results of the optimal solutions (Kämpf *et al.*, 2010).

2.3.4   Application of Optimisation in Multiple Domain

In some case, there applies multi-objective optimisation or single objective optimisation applied for multi-objective problems in cross-domain researches.

Particle Swarm Optimisation (PSO) is adopted in an optimisation of hourly illuminance outcome and annual hourly energy consumption for building envelops (Futrell, Ozelkan and Brentrup, 2015). NSGAII is applied to optimise the daylight and thermal discomfort times (Chen, Yang and Sun, 2016). Optimisation is adopted for balance of daylight and cooling energy consumption for a parametric building (Chen, Janssen and Schlueter, 2018), or to balance daylight and thermal comfort (Zhang *et al.*, 2017). The Multi-Objective Evolutionary Algorithm (MOEA) is used in minimise irradiation abatement due to thermal loss, considering building shape and volume (Kämpf *et al.*, 2010). MOEA is also used for balance of solar radiation and energy consumption in high-rise office context (Yi, 2014).

Strength Pareto Evolutionary Algorithm 2 (SPEA-2) is used in a multi-domain optimisation, involving daylight, energy usage, thermal requirements and capital cost for a folding building facade (Negendahl and Nielsen, 2015). Daylight and evaluated

logistics are optimised with NSGAII (Su and Yan, 2015).

In one case, a special manner is used to achieve multi-domain multi-objective optimisation (MD-MOO) by breaking the multi-objective problem into several single objective optimisation (SOO) problems and solving them in a sequence. This performance-based parametric design exploration takes an office building as a sample and adopts GA in Galapagos platform for three independent single objective problem optimisations. The first objective is to minimise solar irradiation by calculating the optimal of shading device for roof and south facade of the atrium, involving four parameters, angle, depth, number and distance. The second objective is a combined fitness function of solar irradiation and daylight factor, by calculating optimal involving number, depth and distance of the shading device for east and west facades. The third objective is the target daylight factor on each floor, by calculating optimal angle for east and west shading devices.

The advantage of this research is a full loop of performance-based evolution calculation (EC) optimisation. It accomplished the iteration of initial simulation, optimisation calculation; parametrically remodel the optimal solution, simulation of the optimal solution, new generation forming. The optimal would be determined when constraint conditions are met. Three key factors contribute the basement of the full iteration of performance-based optimisation, which are the fast calculation of single objective optimisation with GA, the capability of real time simulation for solar irradiation and daylight factor and the parametric model capability in Grasshopper for building scale model (Ercan and Elias-ozkan, 2015).

However, it is worth mentioned that this research is actually solving a single domain multi-objective problem, namely in architectural daylight and sunlight domain, the multiple objectives of solar irradiations and daylight factors on various simulation grid surfaces could be globally optimised against four parameters of shading device on three facades and roof. In this research it is separates into three single objective

optimisation questions, which means the influences of the later optimal on previous optimised performance are not considered. Surely, the simplification of single optimisation would make the iteration of performance simulation for optimal easier. Secondly, the optimal calculation process is searching for a single minimisation of the fitness function, with records of all near optimal solutions in the pools. The final design is selected within the pool, and does not have to be the real optimal. This is also the side effect of applying single objective optimisation. For multiple objective optimisation even for some single objective optimisations, a Pareto solution set could be calculated, that all solutions in the set are equally good for the optimisation problem.

In this thesis, three domain problems are researched, and in each domain a compounded single objective function is used. Covering three domains, a multi-domain multi-objective optimisation via NSGAII algorithm is optimised. The advantages are: first, considering multi-objective optimisation could achieve global optimal of the problem; second, the adoption of Pareto solution set could collect all possible equally fit optimal solutions. Because of the complexity of parametric modelling for an urban scale model of residential neighbourhood, a fast decoding of optimal solution into 3D model for iterative simulation is hardly achievable at the moment. Also due to the thermal simulation requires simulating the computational fluid dynamics (CFD) of certain space in a period of interest, the time cost is much longer for this urban scale space. Because of the two reasons mentioned above, the iterative simulation for optimal in each generation are not involved in this optimisation procedures.

Since the global optimal is the target to achieve, the selection of proper optimisation algorithm is vital according to the requirement of the optimisation problem. In this research, the final aim is to provide a considerable amount of optimal building distribution design possibilities and patterns which for designers to choose from. Hence, not only the maximum or minimum value is required from the global objective

function as a sole optimal solution. It is required to collect all possible well performed optimised solutions that fit the integrated performance criteria.

The conventional single-objective optimisation algorithm mentioned above and the method of transferring multi-objective problem into single-objective function to solve are searching for the only one optimal solution. If an optimal solution set is needed, it is conducted through collect the optimum together with the satisfying quasi-optimums, in last iteration or last several iterations.

Non-dominated sorting genetic algorithm since its development is aiming to solve multi-objective optimisation with multiple optimums. Hence the optimising aim of NSGA is matching the requirement of the optimisation problem in this research. NSGAII is tested being more effective in searching design space. Therefore it is selected as a basis of making improvement in the algorithm.

An algorithm is not possibly equally efficient for all real world problem, so that researchers may propose new algorithms or improve existing ones, for example hybrid existing optimisers, to fit for a certain research question (Jangir and Jangir, 2018). This research applies both EC and SI, respectively adopted algorithm of Non-Dominated Sorting Genetic Algorithm with Elitist Strategy (NSGAII) and Grey Wolf Optimiser (GWO).

2.4 Integrated Environmental Assessment System for Multi-Domain Optimisation and Decision of Performance Metric

The integrated environmental assessment system refers to the overall assessment criteria from various environmental aspects on specific performance evaluation parameters. Here three domains of acoustic, sunlight possibility and thermal comfort are involved. The specification of assessment variables and corresponding criteria refer from various authorities and sources. Various source published by British and Chinese architecture design, planning and environment authorities are reviewed, which are consisted of design criteria, guidance, regulation and local design habits, as well as local market expectations and cultural preferences.

2.4.1 Environmental Evaluation and Assessment Criteria from Various Sources

Regarding to various scopes of environmental evaluation and assessment, multiple international, national, regional and local authorities have published documents. Publications systems regarding to standard and suggestions of building engineering based in Britain and China are compared and reviewed. Other internationally recognised documents focusing on single domain are also compared.

The principle parts of the quoted criteria are adopted from publications of National Standardization Bodies in UK and China, which are British Standardisation (BS) and Chinese National Standardisation (GuoBiao, GB). These legislated publications are compulsory with full scope of building engineering, and limit the tolerable bounds of certain evaluative aspects. Although there are regional and local obligatory publications, the mandatory provisions are compliant to which in national publications. Hence, regional and local standards and ordinances are not quoted.

The criteria publications based in UK could be divided into two sections according to

different levels of legislative requirement. The obligatory publications are comprised of building regulation, codes, standards or ordinances. The series of voluntary criteria including recommended practices (partly are approved American National Standards Institute (ANSI) standards), design guides (of standard or specific design aspect), guidelines (general rule and a piece of advice), technical memoranda and measurement and calculation guides. The multi-domain environmental assessing criteria are quoted from mainly obligatory regulations and codes, and partly from voluntary guidelines.

Chinese standardisation system regarding to architecture, includes national standards (GB), professional standard systems (JG), and regional standards (DB), of which are obligatory or voluntary. The national standards could be divided into basic, general and specific according to scope of content. In the obligatory standard, also suggested provisions are adopted. The voluntary standards are noted with T representing they are suggested. In addition, there are voluntary institutional standards published by China Association for Engineering Construction Standardisation (CECS).

In both UK and China the provisions in obligatory publications are compulsory. Comparing the criteria value of identical evaluative parameter, Chinese regulations and standard has higher requirement than UK. Therefore, the criteria with high requirement level are quoted during construction of multi-domain evaluative system.

As for The voluntary publications are more of advices of a good practice. Hence the context difference would lead to discrepancies in suggestions. The advisory publications in UK provide advices of higher level performance than the advisory publications in China. Furthermore, the application of advisory publications in practice in UK is more frequently adopted compared to the condition in China. Hence generally speaking the eventual designing results in UK appear under higher level requirement and limitations than which in China does. Hence, the provisions in advisory publications in UK are also quoted in evaluation system. However due to the

different expectations in occupation of a designed space, the advisory provisions on same environment may vary in UK and China. Hence the adoption of feasible provision and adaptation to local context in China is essential.

Unlike for acoustic environment, Environmental Quality Standard for Noise (GB 3096-2008) has present detailed items of limitation and suggestion, in Chinese standard system, environmental quality standard for outdoor light or thermal comfort is not existed yet in Chinese standard system. The current environmental quality controls are focusing on soil, surface water, underground water and air. Visual and thermal comfort advisory standard is highly suggested to be established for improved constructive environment.

Quotations from UK standard system are compared in following sections accordingly. The quotations referred from Chinese and UK standard system in this research are listed table 2.1 and 2.2 respectively.

2.4.2   Determination of Assessing Parameters and Critical Values for Acoustic Domains

Residential noise control is a vital topic through the world. Environmental noise could have serious effects on human health, which include annoyance, sleep disturbance and cardiovascular disease, considered by the World Health Organisation (Hurtley, 2009; WHO, 2011) Research shows that with increasing of noise exposure, the prevalence of noise-related health effect also increases (Miedema and Oudshoorn, 2001; WHO, 2011). The major source of environmental noise is road traffic noise (EU Directive, 2002). A report estimates that about one in three urban inhabitants in EU countries are highly annoyed by road traffic noise (WHO, 2011).

Most traditional acoustic researches related to dwellings are about noise impact at building facades or indoor noise level. It is suggested that human response to noise is

conditioned, not only by the exposure level at the dwellings but also in their surrounding neighbourhood environment (Klæboe, Engelien and Steinnes, 2006; Botteldooren, Dekoninck and Gillis, 2011; Klæboe, 2011). Additional approaches are required to help with noise abatement, except for technology development at source emission, road surface, etc. some attempts are utilising the variation in soundscape to improve sound environment (Skånberg and Öhrström, 2002). A good attempting is to utilise the distribution of building inside a residential area, to optimise the sound environment in the residential area.

In order to have access to control the effect of environment noise, in European cities, noise maps and noise level distribution at facade of dwellings are frequently produced (EU Directive, 2002). Although some noise mapping has been done in Hong Kong related to dwellings (Lee, Chang and Park, 2008; Lam and Chung, 2012; Lam and Ma, 2012), there is still lack of noise mapping data base in mainland China for designer to review while planning residential wards. Because based on detailed traffic and building data, to generate large amounts of noise maps is very time-consuming (King and Rice, 2009), to build up noise map data base for design purpose in short term is impractical. Hence, this research aims to extract a series of general designing rule related to residential ward sound environment, out of the case simulations and analyses. These rules could help with optimising the building arrangement and distribution at prime design stage, if lack of other noise-related data to review.

In the scenario of residential wards, acoustic performance mainly refers to the result of traffic noise propagation from the roads and streets within and around the ward, regardless of human activity noises. As in compact residential ward in SU-ZHE-WAN region in China, traffics are mostly guided into car parks close to the ward entrance. So traffics within the ward are excluded in this research to focus on the external traffic noise propagation in a ward.

External streets and road are considered as linear sound source wrapping around the

interested residential wards. The assessing measure of the acoustic performance is sound pressure level (SPL). In the simulation of traffic noise propagation in the ward, assessing points or receivers to collect SPL data are arranged 1) on the grid points evenly distributed in the ward; 2) pairedly distributed along the front row (street facing) residential buildings, in-front-of and behind the facades facing traffic and in noise shadow respectively.

As the objective of this research of outdoor environment, the criteria specifying outdoor requirement are directly used, while the criteria in regard to indoor requirement are referred or corresponding indoor parameters are borrowed to outdoor context.

Criteria regarding to acoustic performance in Chinese standards are presented from aspects of indoor and outdoor. General criteria are mentioned in Residential Architecture Design (GB50096-2011), more details presented in Code for Design of Sound Insulation of Civil Buildings (GB50118-2010), Environmental Quality Standard for Noise (GB 3096-2008), Technical Specification for Construction of Healthy Housing (CECS 179:2009).

The daytime (6:00-22:00) requirement could be summarised as for indoor environment, the satisfactory standard is L(A)eq of 45dBA for main functional room like bedroom and living room, with lower bound of insulation ability of external window and wall respectively of 30 and 45dBA. In other word, when adopting the lowest possible insulation level of building itself, to fulfil indoor requirement, the allowed highest level of SPL close to window is 75dBA (45+30dBA). The requirements for outdoor environmental noise are 55dBA (suggested), 55dBA (satisfactory), and 60dBA (lower limit) for region of residential and commercial mixture, and 70dBA for each side of traffic artery (≥100veh/h). The suggested receiver height and interval for indoor environment is 1.2-1.6m, and over 1.5m, while for outdoor height is over 1.2m. Inside Distance between receiver and reflective

surface should be over 1m, while outside distance to wall and window of noise sensitive building equals to 1m.

According the quotations, 55dBA is the satisfactory threshold for outdoor acoustic performance in residential area and 70dBA for locations on the side of traffic artery. But due to the traffic estimations of the roads in this research adopt the highest capacity value, the resulting SPL is higher than practice, with 80-90dBA on the side of artery. Hence, an appropriate loosening is need for adaptation of the practice standard into this research context. Therefore, the 65dBA is adopted as threshold value of higher limit of noise level in residential area in this research.

According to E. Öhrström etc 2006, a quiet side concept is mentioned. The sound levels from road traffic at the most-exposed side of one dwelling should not exceed LAeq 60dB (24h), under condition that a quiet side LAeq 45dB (24h) is available. This refers to if most exposed side sound level not exceeding 60dBA, quiet side concept could be applied to promote in-site sound environment. If quiet side of a dwelling is accessible, it corresponds to a reduction of 5dBA ($L_{Aeq,24h}$) at most-exposed side. The quiet side concept echoes to the building shielding of non-residential buildings to reduce adverse effect on dwellings.

This approach of involving multiple design variables and considering multiple design objectives simultaneously through a fast meta-model based global optimisation, based on proactive simulation database, makes higher integrated performance more achievable from earliest stage.

Similar to the abovementioned statistical approach, in this research the non-exhaustive proactive simulation of selected series of real project samples is followed by a statistical analysis to define possible range of input variables and underlying rules. The non-exhaustive simulations are based a structured sampled real

projects, instead of generated models of various parameter combination. Sensitive analysis in operated though multivariate linear regression.

Although large amount of one-at-a-time simulations in multiple software are avoided, the limitation of the non-exhaustive proactive simulation is that not full possible ranges of all input variables are explored. But the sampling from real projects represents the performance variation band of existing schemes and input combination habit. This limitation is compensated by the exhaustive search of following global optimisation in an expanded range based on the statistical range from real project samples, stated in chapter 8.

The other approach of exhaustive proactive simulation is to sample the design space by certain list of variables though design of computational experiment (DoCE) method, for instance Latin hypercube sampling (LHS), to form sufficient amount of solutions for operate simulation. The shortcoming of this approach is required large amount of one-at-a-time modelling building and simulations in multiple domains high-accuracy simulation software which could be highly time-consuming and effort intensive.

The ambiguous suggestion in the end of the research is advocating global sampling on the design space and automating model generation though parametrical modelling, to prepare for high accuracy simulations and update global optimisation functions.

Table 2.1 Quotations Regarding Integrated Optimisation in Chinese Standard System

Scope	Category	Name	Order	quotation
Urban-Rural Planning	National Specific standard	Code of Urban Residential Areas Planning & Design	GB50180-2018	5.0.2.1 residential building separation regarding to sunlight requirement: sunlight hour for residential building ≥2h/3h between 8:00-16:00 at ground floor window sill/ 0.9m over interior floor on great cold day
Architectural Design	National General standard	Code for Design of Civil Buildings	GB50352-2019	7.1 light environment should compliant to standard for daylighting design for buildings GB50033-2013 7.2 ventilation: descriptive provisions regarding to window attributes, no environmental evaluative parameter applied, ie, wind speed 7.3 thermal environment should compliant to Code for thermal design of civil building GB 50176-2016 7.4 acoustic environment should compliant to Code for design of sound insulation of civil buildings GB50118-2010
	National Specific standard	Regulations of Residential Architecture Design	GB50096-2011	7.1 indoor sunlight, natural lighting, shading; 7.2 natural ventilation; 7.3 sound insolation and attenuation: 7.3.1 Daytime bedroom/ living room:L(A)eq≤45dBA 8.6.4 suggested 26°C for AC bedroom/ living room
Architectural Indoor Environment	National Basic standard	Architectural Climate Zoning Standard	GB50178-93	3.3 climate zoning III of hot summer and cold winter region
	National Genetic standard	Code for Design of Sound Insulation of Civil Buildings	GB50118-2010	3.0.2 planning for noise control for residential area against heavy traffic and rail: adopt noise insensitive buildings as noise barriers at residential area boundary; introduce noise barrier when noise sensitive buildings violate acoustic environment criteria 4 residential buildings： 4.1.1 Daytime bedroom/ living room:L(A)eq≤45dBA, or ≤40dBA (for higher standard) 4.2.5 air-bourn sound insulation of external window≥25dBA, ≥30dBA (for bedroom/living room) 4.2.6 air-bourn sound insulation of external wall≥45dBA Appendix A indoor meansure 5.1receiver height above ground 1.2-1.6m;5.2 distance between receiver and reflective surface≥1m; 5.3 receiver interval≥1.5m; 5.4 distance between receiver and source≥1.5m
		Standard for Daylighting Design for Buildings	GB50033-2013	It focuses on indoor task oriented evaluation not suitable for this research. Critical illuminance of exterior daylight and critical illuminance of interior daylight Target region of research locates in daylight climate zoning Ⅳ

			3.0.4 design illuminance of exterior daylight=13500lux for allowing complete indoor daylighting
	Standard for Assessment Parameters of Sunlight On Building	GB/T 50947-2014	5.0.1 measurement point interval 0.3-0.6m for window, 0.6-1m for building and 1-5m for site assessment. 5.0.6 on vertical surfaces of various of wall openings, the assessment height is 0.9m over indoor ground floor level.
	Code for Thermal Design of Civil Building	GB 50176-2016	8.2.1 suggested angle between incident dominant wind direction and building orientation: ≤30°(strip-shaped); ≥30-60°(point-shaped) 9 sun-shading design to avoid over heating. No criteria on radiation intensity about human perception. Radiation absorbing coefficient of various wall surface, could calculate surface radiative ability and temperature. Temperature are focusing on HVAC efficiency, with suggested temperature of 18°C in winter, 26°C in summer
	Design Code for Heating Ventilation And Air Conditioning of Civil Buildings	GB 50736-2012	3.0.2 thermal comfort parametric of HVAC indoor environment: satisfactory standard for cooling condition, temperature 26-28°C, RH≤70%,wind velocity≤0.3m/s, or 0.5m/s (for short stay). 4.1.8 when calculating summer indoor ventilation, the outdoor temperature adopted should use the average temperature of 14:00 of the hottest month in past 10-30 years. 6.2.1 to encourage natural ventilation in residential buildings, the angle between dominant wind direction in summer should be 60-90°and over 45°. 7.4.13 return air inlet suction speed ≤1.5m/s at location of occupant's continuous stay.
	Evaluation Standard for Indoor Thermal Environment In Civil Buildings	GB/T 50785-2012	Ditto Adopt subjective thermal comfort assessment by predictive mean vote (PMV) and predictive percentage dissatisfied (PDD).
National Specific standard	Design Standard for Energy Efficiency of Public Buildings	GB 50189-2015	3.1.3 The master plan for the building complex should attenuate UHI. The master plan and site plan should be in favour of natural ventilation and winter sunlight hour. Main building orientation should select by local preference and avoid winter dominant wind direction. Commentary of 3.1.3 started from climate condition of location, planning need to combine architectural design with architectural micro climate, technology and energy efficiency. architectural energy efficiency design need consider sunlight, dominant wind, natural ventilation, orientation, etc. specifically, in winter increase sunlight, solar radiation and avoid dominant wind; in summer reduce heat gain to the max and increase natural ventilation Commentary of 3.2.8 based on the observation and simulation of cases in south china, when outdoor Tdb≤28°C, RH≤80%, wind velocity≤1.5m/s, if opening of external window accounts over 8% of room area, most indoor area could achieve thermal comfort.

				7.2.6 evaluated on winter solstice day, accumulative sunlight hours on lighting surface for solar thermal collector≥4h, for photovoltaic module≥3h.
		Assessment Standard for Green Building	GB/T 50378-2019	8.2.8 under typical winter wind velocity and direction, at height 1.5m over ground, wind velocity should ≤5m/s (pedestrian around buildings), ≤2m/s (outdoor recreational area and playground); Under typical summer wind velocity and direction, should not allow wind turbulence and no-wind zone.
	Professional standard	Design Standards for Energy Efficiency of Residential Buildings In Severe Cold And Cold Zone	JGJ 26-2018	Commentary of 3.0.1 indoor thermal environment evaluative parameter include temperature, humidity, wind velocity, surface temperature, etc, but only adopt temperature and aeration factor, due to uncontrollable feature of humidity and wind velocity as central air conditioning rarely used in residential buildings.
		Design Standards for Energy Efficiency of Residential Buildings In Hot Summer And Cold Winter Zone	JGJ 134-2010	Highest roof surface temperature could reach 62, 64°C in Nanjing, Wuhan; west facade surface temperature could reach 51, 55°C in Nanjing, Wuhan.
Outdoor And Indoor Environment	National Generic standard	Environmental Quality Standard for Noise	GB 3096-2008	3.4 daytime means 6:00-22:00 5.1 daytime environmental noise level=55dBA for type1 region of residential and commercial mix. 6 daytime environmental noise level=70dBA for each side of traffic artery (≥100veh/h) 6.2 receiver height ≥1.2m, outside distance to wall and window of noise sensitive building=1m Appendix A traffic estimation of each traffic artery type is based on this section.
	Institutional Standard	Technical Specification for Construction of Healthy Housing	CECS 179:2009	3.3.1 wind environment in residential area 1 with utilisation of local dominant wind direction the suggested building distribution strategies are matrix-style, free-style, low-front-high-rear, and rhythmic high-low alteration. For strip-shaped building of long facade, should adopt arcade or ground floor void properly. 3. under typical local climate condition, the simulated wind environment in residential area require wind velocity ≤1.5m/s (pedestrian around buildings). Avoid local wind turbulence and no-wind zone. 3.4.2 suggested indoor wind velocity ≤0.3m/s (cooling), ≤0.2m/s (heating). 3.5.1 daytime outdoor environmental noise in residential area should ≤50dBA (suggested), ≤55dBA (satisfactory), ≤60dBA (lower limit). 3.6.1 accumulative sunlight hour on window of ground floor room≥2h (large city in zone Ⅲ), ≥3h (small-medium city in zone Ⅲ).

Table 2.2 Quotations Regarding Integrated Optimisation in UK Standard System

Scope	Category	Name	Order	quotation
Daylight and sunlight (indoor and outdoor)	Legislation documents	Right of Light Act 1959		Compulsory. Window with uninterrupted daylight available for 20 years is protected for obstruction by Right of Light Act 1959. Assessed on case-by-case basis.
	British Standard	BS 8206: Lighting for buildings, Part 2: 1992 Code of practice for daylighting	BS 8206-2	Indoor daylight (Voluntary): ADF≥1.5% for living room; ≥2% for kitchen; ≥0.5% for bedroom.
	Institutional Guides	Lighting Guide 10: Daylighting - a Guide for Designers	CIBSE LG10/14	Daylight factor could be calculated manually using the techniques provided
	Institutional Guides	Building Research Estalishment (BRE) BR 209: Site layout planning for daylight and sunlight: a guide to good practice 1991	BR 209	Right of Light is assessed in EIA for new development, by the angle between centre of the tested window and the proposed development with assessment procedure provided in BR209. Sufficient light amount is not objectively quantified in the Right of Light. It is assessed on a case by case basis by the courts.

Outdoor overshadowing(voluntary):Permanent shadow≤25% assessed area with max value 40% recommended on 21 March; if possible transient shadow path tested on hourly basis for open space and buildings on 21 March, 21 June, and 21 December. Assessed on a case by case basis.

Outdoor sunlight (voluntary): ≥2h direct sunlight at least 50% of the amenity area on 21 March on which building cast longest shadows. The two hours sun contour is suggested to be plotted to mark the target area and arrange outdoor space functions according to sunlight requirements.

Sunlight hour at external facade window (compulsory): sunshine hour≥25% APSH annually and≥5%APSH in winter
Assessed by counting unobstructed dots in the view of window concerned when proposed building superimposing on the sunlight availability indicator maps. Only presented three reference maps of London of 1486h available sunlight hour, of Manchester 1392h and Edinburgh/Glasgow 1267h. |

| Sustainable design (indoor and outdoor) | Reference book: Specialists' guide for emerging and meeting UK and international statutory and requirements | Integrated sustainable design of buildings, Routledge, Appleby, P. (2012) | N/A | Wind impact assessment required as part of an EIA for developments with a potential negative impact on the wind environment in the surrounding public domain, such as might apply to tall buildings. Currently no UK wide planning policy or legislation that applies to this and the requirements for a planning application will be dependent upon legal precedent and the specific rules laid down by a local authority.

In the Lawson Comfort Criteria 1977, regarding to suggested wind speed upper limits and exceeding percentages at space used by pedestrians for a full range of activates. If wind speed for outdoor sitting space is over Beaufort Force 3 or 5.4m/s for 1% of the year, the space needs to be amended.

However the only recommendation for comfort level regarding to wind is mentioned for indoor condition: a recommended maximum percentage of down draft is 15% namely a maximum air speed around 0.15m/s at operative temperature of 22°C. |
| Indoor thermal comfort | | Ergonomics of the thermal environment. Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, 2005 | BS EN ISO 7730:2005 | For free-running indoor environment, wind speed at 0.8m/s or more which is likely to cause inconvenience of loose papers lifted from surfaces in offices.

at wind speed 0.15m/s, a upper limit of operative temperature for free-running indoor condition could expect 27.4°C. If wind speed increase, the satisfactory operative temperature could be even higher. |

2.4.3 Determination of Assessing Parameters and Critical Values for Lighting Domains

British daylight and sunlight standard exhibit a trend of providing descriptive suggestions and verifying the daylight and sunlight availability to support the certain functions at the target unit. Target values of performance parameters are barely presented primarily in the standards. Then the spaces are organised accordingly matching the unevenness of performances with spaces' requirements. This is also mentioned in BR 209, fast and hard rules of target value are difficult to specify due to various context and standing point of different groups. One possible reason is that overall integrated performance of a site is unknown at primary design stage until schemes are tested in current design process. This research is aiming of exploring the possibility of providing holistic environmental performance preferences before specific design starts. This design progress provides are novel aspect of view compared to the current design process stated in UK and Chinese standard system.

Light related provisions on outdoor context are not the predominant part in standard system in both UK and China. The only provisions about outdoor daylight, sunlight and overshadowing are compared, and indoor evaluative parameters are discussed for possibility of adaption to outdoor context. The following terms are defined in BR209: Daylight means diffuse light from overcast sky. An unobstructed view of sky can provide daylighting in a building. Sunlight refers to direct sun rays exposed in the habitable rooms. Overshadowing means the shadow cast by proposed building in garden, amenity area and open space of the existing building (Littlefair, 2011).

Outdoor lighting involved in Chinese standard system, principally refers to task-orientated artificial lighting with some involvement of daylighting, with quantitative illuminance requirements according to various tasks on various venues, i.e. from coarse visual tasks in stadium to medium precision visual task at harbour to high precision visual task in semi-open factory. Regarding to indoor daylight, in Standard for Daylighting Design for Buildings (GB50033-2013), there are required indoor daylighting hours for various building types with critical illuminance targets for various tasks, i.e. school, office, etc., but residential building is not included. Indoor daylight requirement in residential context in UK standard system is presented

by critical percentage of average daylight factor (ADF) for main habitable rooms (BS 8206-2). However, this research only evaluates test points on horizontal surfaces in outdoor condition. Because ADF is requires proportions of obstructed illuminance to unshaded outdoor illuminance on horizontal surface from CIE overcast sky, it is impossible to adapt this indicator into this research as an outdoor index on horizontal calculation grid. Hence no outdoor daylighting is evaluated without borrowing any indoor daylighting parameters. Quantitative daylighting requirement considering occupants' healthy and comfort requirements are expected in future compilations of standards.

Regarding to overshadowing, qualitative and descriptive requirements are presented in multiple standards in Chinese system in forms of recommended clauses regarding to encouragement of sunlight and avoidance of over-shading in recreational zone and play ground in residential area (GB50180-2018, GB50096-2011). It is suggested to expand quantitative criteria considered from healthy and comfort aspects of occupants in outdoor environment in future revisions, except for the only clause of recommended upper limit of overshadowing percentage assessed on the great cold day. Overshadowing assessed in BR209 guidance also only focuses on outdoor functional area by means of case-by-case simulation. The assessing parameter is also the allowed the percentage of permanent shadows on 21st March. Transient overshadow path on an hourly basis for open space and buildings are also suggested for new development, on 21 March, 21 June and 21 December if possible. Due to the fact that in early stage of design the outdoor functional areas are not specified, the verification of overshadowing is not practicable no matter in the approach of UK or Chinese standard system. Besides, the nature of assessing permanent shadow duplicates the accumulative sunlight hour analysis. Hence, conventional overshadow testing is not applied in this research.

In UK design guidance BR209, regarding to outdoor sunlight context, the recommendations concentrate on gardens and open spaces, including gardens, parks, playing fields, children's playgrounds, outdoor swimming pools, paddling pools, sitting out area and focal points for views, assessed on 21st March. It is suggested that at least half of the amenity area should receive at least 2h direct sunlight on the assessment day on which building cast longest shadows. The two hours sun contour is

suggested to be plotted to mark the target area and arrange outdoor space functions according to sunlight requirements.

Outdoor quantitative sunlight criteria related to residential context in Chinese standards are only mentioned at ground floor window sill of main habitable rooms in residence with over 2 or 3h accumulated hour on the great cold day; and at surfaces of thermal collector and photovoltaic module on winter solstice day of 4 or 3h accumulated hour to ensure producing efficiency. This assessment is on the basis of short term accumulative sunlight hour on the critical day. Indoor sunlight potential is indirect estimated by sunlight hour on external window sill rather than directly assessed. This is in accord with the fact that recent researches related to sunlight hour are predominantly related to effective requirement of solar energy harvesting. This research calculates accumulative sunlight hour distribution on horizontal and vertical grids, which could also contribute to arrangement on solar energy harvest ends in new development. Sunlight availability of a residence is assessed by insolation hour in UK standard system. Provided in UK BRE BR209, Annual Probable Sunlight Hours (APSH) method is used in UK for assessment. APSH method is used to predict the sunlight available during both summer and winter for the main window of each habitable room that faces within 90 degree of due south. 21 September to 21 March is considered to be winter period, and the remaining months constitute the summer. For new building, suggested APSH for a whole year is no less than 25% of the total available APSH with no less than 5% of total available in winter season (WPSH). APSH and WPSH calculation require cloudiness of the climate to be excluded. APSH refers to the sunshine hour during the year with cloudy hours excluded from the overall daylight hour at the certain location, which could provide effective insolation. The method of calculating APSH provide by BR209 guidance is counting unobstructed dots in the view of window concerned when proposed building superimposing on the sunlight availability indicator map, with only three reference maps of London with 1486h sunshine hours, of Manchester 1392h and Edinburgh/Glasgow 1267h. This assessment strategy differs from which in Chinese system, is a long term accumulative hour, with consideration of climate cloudiness at researched location.

Comparatively, long term accumulative sunlight hour adopted in UK standards is a

more holistic and informative parameter than sunlight hour assessed on a representative day as in Chinese standards. Hence, the UK criteria is adopted as assessing standard while the short term sunlight criteria in Chinese standards are derived to a long term accumulative metric of three month winter time span in this research as a comparison. The two evaluative parameters of long-term accumulative sunlight hour are simulated and compared for their representative and distinguishing ability on sunlight possibility.

To simulate long-term accumulative sunlight hour, namely APSH and WPSH, firstly the total availability of sunlight in certain location represented by sunshine duration or sunlight hour is needed for the researched city. The term sunshine duration is defined by The World Meteorological Organization as the cumulative time during which an area receives direct irradiance from the Sun exceeding 120 watts per square meter. Therefore, sunshine hour is calculated by counting all possible hour with direct irradiance over 120W/m2 from daylight hours presented in the local weather data provided in EPWMaps of Ladybug tools, sourced from EnergyPlus weather date provided by World Meteorological Organization region and Country. This method is in accordance with the one adapted in the TAS engineering software package by Environment Design Solutions Ltd (EDSL). The calculation result of sunshine duration in London, Manchester and Edinburgh which are 1478h, 1449h and 1248h respectively which are comparable to the abovementioned value provided in BR209. As Hefei is adopted as representative city, based on the definition of sunshine duration, through Ladybug calculation the annual sunshine duration in Hefei is 1653h. This estimation could be verified by data (1971-2000) from China Meteorological Data Service Centre, that in the SU-ZHE-WAN area of research interest, where locates in the Middle-Lower Yangtze plains, the annual sunshine duration is 1500-2500h. Because of the clear four seasons in Hefei, it is not acceptable to set winter season from 21 September to 31 March as in BRE 09. In this research, winter is set to 01 December to 28 February as in Chinese four season convention. Hence, the total sunshine duration in winter is 348h at Hefei according to Ladybug calculation. As aforementioned UK criteria of APSH and WPSH is over 25% for the whole year and 5% for winter of the local available sunshine hour, the insolation criteria at Hefei is no less than 413h annually with no less than 83h in winter.

As in table 2.1, in Chinese standard GB50180-2018, at researched location Hefei in architectural climate zoning III, residential building must fulfil requirement of sunlight hour no less than 3h between effective sunlight hour (8:00-16:00) at ground floor window sill of main habitable room / 0.9m over interior floor on the great cold day. The specification of effective insolation time period is a simple alternative to define the time span by solar irradiation over 120W/m2. It is stated in the standard, for large scale cities with high population density and built density, when insolation critical value on the winter solstice of lowest solar altitude, is hard to achieve for proposed project, even for new development, In this case the great cold day is adapted as assessment day of a lower evaluation standard for zoning I, II, III and large scale city in zoning IV. Hence, it could be hypothesised that 3h sunlight at Hefei in winter is almost the lowest daily threshold. Hence, the estimated critical value of accumulative winter sunlight hour based on Chinese standard is 270h (=3h×3months×30days). Therefore a novel evaluative parameter of accumulative sunlight hour in winter (SHW) based on short-term Chinese sunlight standard (on great cold day) and expanded to long-term accumulative value (December-February) is created, with criteria value of no less than 270h accumulative sunlight hour. Conventional sunlight availability in both countries' standards is assessed on external facade or window. To make it possible for examination of impact of building distribution on sunlight distribution and therefore guiding outdoor space arrangement, the sunlight assessment is expanded to horizontal simulation grid over the site, except for ground floor vertical surfaces representing the windows.

As one objective of this research is to maximise sunlight hours year-round and in winter, there are consideration about it may cause overheating in summer and glare in both outdoor and indoor. However, the building mass distribution makes much more significant impact on shadow casting and insolation hour on facade and ground in winter than in summer. Meanwhile, vegetation arrangement and other detailed building design contribute much more on sun shading in summer when solar altitude is high than building mass distribution. In this case, the optimisation of insolation is to maximise the sunlight hour both in winter and year-round without considering overheating. Similarly, in case of potential indoor glare and overheating due to optimised direct sun exposure, there are architectural techniques applicable indoor to mitigate them: through approaches applicable in individual building design procedure

after site layout stage, i.e. horizontal shading, limit opening area, application of low-e coating glazing, purging heat gain by ventilation, building envelop insulation, reducing solar absorption at building envelop by apply high albedo facade and reducing roof heat gain with green roof (Solar gain in buildings, 2017)

In addition, it is stated in BR209 the importance of remaining skylight for existing windows assessing by the angle of visible sky. This metric could be presented by the vertical angle, average daylight factor on facade (ADF) and vertical sky component (VSC). But to make is adaptive to horizontal simulation grid over ground in this research, sky view factor (SVF) has similar definition to VSC but with discrepancy on projection area of visible sky patches to the tested surfaces. Besides, in GB 50189-2015 it is suggested to attenuate urban heat island intensity (UHI) through organisation of master plan and site plan by encouraging site ventilation. Respecting the effect of sky view factor (SVF) in reducing urban heat island intensity, SVF is adopted as an evaluative parameter of spatial openness, visual comfort and as well as UHI avoidance. However no critical value is stated in standards, SVF value is expected the higher the better.

2.4.4   Determination of Assessing Parameters and Critical Values for Thermal Domains

Gehl first studied outdoor activities related to local microclimate from sunlight aspect by of practice observation method of counting people sitting on sunny or shady benches (Gehl, 1971). In the past, a vast number of research projects on outdoor thermal comfort in various climates around the world, some focusing on thermo-physiological perspective by means of modelling and assessment (Höppe, 2002; Gulyás, Unger and Matzarakis, 2006), some focusing on subjective thermal comfort level by means of investigation (Spagnolo and de Dear, 2003; Cheng and Ng, 2006)

The general framework of outdoor thermal comfort assessment could be divided into four levels, which are physical, physiological, psychological and social/behavioural levels (Chen and Ng, 2012). The physical level involves in objective influencing factors, i.e. building form and microclimate indices and adopts research approaches of

measurements and modelling. This is the most conventional research perspective which is highly informative and directly supportive to planning and design procedures. Physiological level involves objective biometeorological indices, i.e. in thermalregulation and energy balance models, using research approaches of modelling and monitoring. While psychological and behavioural level researches rely on survey, interview, observation and prediction approaches, requiring considerable amount of participants. These two assessment levels involve subjective factors, i.e. expectations, past experiences, neutrality, autonomy of occupants; and errands and preferences of occupants, respectively.

The currently most accepted assessment method in researches for solely outdoor thermal comfort uses models at physiological level which defining relationship between local microclimate condition and biometeorological indices. The thermal assessment at physiological level is divided into steady-state models (Fanger, 1972; Givoni, 1976; Mayer and Höppe, 1987; Hamdi, Lachiver and Michaud, 1999; Kenny et al., 2009) and non steady-state models (Gagge, Stolwijk and Nishi, 1971; Gagge, Fobelets and Berglund, 1986; Höppe, 2002; Bruse, 2005). Multiple steady-state models regarding to human thermal regulation and energy balance are constructed, which are originally developed for indoor assessment and become commonly adopted for outdoor assessment, i.e. PMV model (Fanger, 1972). PMV model is a matured assessment system adopting human sensational indicators, but it is suitable for context of large mount of people being surveyed. In non steady-state assessment, or to say human dynamic thermal adaptation assessment, the most used model is Pierce Two-node Model (Gagge, Stolwijk and Nishi, 1971; Gagge, Fobelets and Berglund, 1986) in which considering body as two isolated thermal parts of skin and core.

As mentioned in ANSI/ASHRAE Standard 55 - Thermal Environmental Conditions for Human Occupancy, thermal comfort is affected by ten key factors: mean radiant temperature, dry bulb temperature, humidity, air velocity, vertical air temperature differences, radiant temperature asymmetry, floor temperature, drafts, metabolic rate and clothing level. This is originally discussing indoor environment which is highly influenced by thermal conditioning and enclosure performance. However, the factors are also valid in analysing outdoor scenario.

If consider thermal comfort from subjective aspect, predictive mean vote model emphasise air temperature, radiant temperature, relative humidity, air velocity, active rate and clothing level to be determiners of subjective thermal perception.

To describe a physical thermal environment conventionally, air temperature, mean radiant temperature, water vapour pressure (namely humidity) and wind speed are needed. And these parameters would be impacted by building distribution design. However vegetation design could improve local temperature and humidity, that it could be considered as next stage outdoor environment optimisation after first stage building distribution design.

Mean radiant temperature (MRT) and wind speed (WS) are adopted in this research as indicators of thermal comfort. No biometerorological model is used. Because relative humidity, active rate and clothing level vary greatly in outdoor scenarios and not controllable as for indoor thermal environment, only MRT and WS is used as representatives of impact from interaction of building mass distribution and environment. With the help of observation and control on MRT and WS in-between buildings in residential ward, the extreme condition of heat stress and wind hazard could be avoided in early design stage. While specific thermal comfort for the habitants could be adjusted according to personal preference on activity rate and clothing level. Due to the focus of building mass distribution in this research, water body and vegetation are not included in the simulation. Therefore, relative humidity which is greatly influenced by both of them is not analysed.

Wind speed is the priority dimension considered in integrated outdoor environment evaluation from thermal aspect. Wind speed value at the specified grid points is used, namely section of 3D wind field. Although wind direction is available from simulation results of Envi-met, but dominant wind direction varies in different cities, for the purpose of generalisation of the result in different location, only value of wind speed is analysed.

Qualitative outdoor thermal environment considering from occupant comfort aspect is also less involved in Chinese standard system. However, general descriptive requirements are stated regarding to avoiding serious defects, not only for residential

context. This is identical as in UK standards, i.e. potential negative impact on the wind environment in the surrounding public domain of a building needs assessment (Appleby, 2012). There are also descriptive recommendations of design strategies regarding to outdoor thermal environment especially for residential environment in Chinese standards, i.e. sun-shading, natural ventilation encouragement, facade orientation, etc, but no quantitative recommendation involved. Recommendations also stated in Appleby's book that according to the most widely accepted wind specialist criteria, the Lawson Comfort Criteria 1977, regarding to suggested wind speed upper limits and exceeding percentages at space used by pedestrians for a full range of activates. It is recommended that outdoor sitting space of wind speed over Beaufort Force 3 or 5.4m/s for 1% of the year needs to be amended. Wind speed below 0.3m/s is considered of Beaufort Force 0.

Although outdoor thermal critical value is not sufficient, provisions of indoor context are also referred. Except for a few quantitative requirements on temperature, wind speed, and relative humidity, the majority of work concentrates on attributes and performances of building components and materials on heat absorption, conduction and radiation in Chinese standards. In the Chinese professional standard GB/T 50785-2012 and British standard BS EN ISO 7730: 2005, subjective thermal comfort assessment system namely statistical representations of likely satisfaction of occupants with the thermal environment is adopted, presented by metric of predictive mean vote (PMV) and predictive percentage dissatisfied (PDD). Although it is one of the most promising assessing systems for outdoor thermal comfort, due to the lack of information of corresponding parameters i.e. metabolic rate and clothing coefficient in the researched cases, subjective thermal comfort concept is not utilised.

Nevertheless when discussing adaptation to free-running indoor environment, BS EN ISO 7730: 2005 provides referable wind speed at 0.8m/s or more which is likely to cause inconvenience of loose papers lifted from surfaces in offices. It could be derived that wind speed at 0.8m/s in outdoor context would have impact on reading. It is also mentioned that at wind speed 0.15m/s, a upper limit of operative temperature for free-running indoor condition could expect 27.4°C. If wind speed increase, the satisfactory operative temperature could be even higher. Quoted from Appleby's book, the only recommendation for comfort level regarding to wind is mentioned for indoor

condition: a recommended maximum percentage of down draft is 15% namely a maximum air speed around 0.15m/s at operative temperature of 22℃. Standards of free-running indoor thermal environment are more general with fewer specifications in Chinese standards.

HVAC controlled thermal environment is the predominant research field in indoor thermal standard due to the controllable interaction among thermal parameters. It is mentioned in Design Code for Heating Ventilation And Air Conditioning of Civil Buildings (GB 50736-2012) that satisfactory standard for cooling condition of indoor environment are 26-28℃ in temperature, less than or equal to 70% relative humidity, wind velocity less than or equal to 0.3m/s (for continuous stay) or 0.5m/s (for short stay). Wind speed for heating condition should be below 0.2m/s stated in Technical Specification for Construction of Healthy Housing (CECS 179:2009).This could be referred as ideal condition for evaluating outdoor thermal environment in summer. Speed 1.5m/s is also indicated when presenting upper limit of return air inlet suction speed for location of occupant's continuous stay. A quantitative outdoor thermal parametric directly is also mentioned during discussion of indoor natural ventilation. In GB 50189-2015, it is stated that based on data of cases in south china, when outdoor dry bulb temperature (Tdb) below 28℃, relative humidity (RH) below 80%, wind velocity below 1.5m/s, if opening of external window accounts over 8% of room area to ensure sufficient natural ventilation, most indoor area could achieve thermal comfort.

Regarding to outdoor wind condition, it is mentioned in BR209, multiple design strategies are listed for reducing wind sensitivity of buildings in planning approach, due to the relatively strong wind environment in UK. It is even mentioned that to provide wind shelter in planning, compromises on natural lighting may occur due to blocking the gaps between buildings and enclosed building arrangement. This is opposite to the practice in China, for zone III natural wind is highly encouraged no matter indoor or outdoor in summer, only to avoid winter dominant wind in window orientation and size design. Hence building gaps are advocated both in wind and daylight distribution design.

Therefore to sum up, 28℃, 0.3m/s and 1.5m/s is used as critical values of satisfactory

MRT, lower limit and satisfactory WS for outdoor thermal evaluation. Although there are content about radiation absorbing coefficient of various wall surfaces which could calculate surface radiative ability and temperature in Code for Thermal Design of Civil Building (GB 50176-2016), there is no standard about radiation intensity regarding to thermal comfort to refer to for longwave radiation from environment (LRE) critical value.

It is also mentioned that at 14:00 of the hottest month in past 10-30 years, the average outdoor temperature should be used for calculating summer indoor ventilation. Hence, among the 24h simulated thermal performances of mean radiant temperature (MRT), wind speed (WS) and longwave radiation from environment (LRE), the slides at 14:00 are selected for analysis and discussion. Not only 14:00 could be referred from previous work, but also it is the starting time of active outdoor activities for elder and kid populations of local preferences.

2.4.5   Summary

With comparison of quotations in Chinese and UK standard system regarding to outdoor environmental performance considered from acoustic, lighting and thermal aspects. Predominant indicator and critical values are decided.

2.5 Interest Location and Sample Selection for Multi-Domain Environmental Performance Simulation

The objective of this research is to discover interaction between building distribution and integrated environmental performances in residential ward of south-east China. Hence, sample residential wards are selected randomly from cities of the selected architectural climate zoning. All 44 residential wards are practice projects and had been built up in representative medium-large-sized cities in SU-ZHE-WAN region in China, so that they could ensure the models for simulation reflecting the real city texture of residence. Selected samples allow variation of distribution characteristics (age, density, plot ratio, green coverage rate, etc) and environmental requirements. Therefore the diversity ensures a composition of full scale of conditions of residential wards prepared for next stage simulation and analysis. Furthermore, the building

distribution of selected wards may have not been evaluated and assessed to fit current building regulations because of various ages, which means they may contain different environment qualities inside the wards. The sample selection keeps a variation of planning quality, which unfold the city appearance honestly.

2.5.1   Population and Economic of the Interest Location

Sample residential wards are selected from cities located in SU-ZHE-WAN region in southeast part of China. SU, ZHE, WAN are three adjacent provinces in southeast China, where share a high economic development pace. The research conclusion and output that will be presented in form of guidelines would only be applied in this particular region. However, the clear four seasons variation in this climate zoning makes the guidelines could also be applicable partly to other climate type with identical partial climate characteristics, i.e. hot in summer warm in winter zone of hot humid climate in China.

Sample residential wards are selected from vary scale cities in the region, and all been built to ensure that they reflect the real situation in cities. Samples are mostly selected from are medium-large scale, some from capital cities of the located provinces.

Geographic and Economic conditions of each city which sites are located are revealed in Appendix A together with the location map of all sample wards. The table shows the city's area, population, GDP and its economic classification on geographical location. These can help to verify the cities have less disparity. Similar level of economic among the cities where samples are selected from makes the various indictors of residential wards comparable across cities. GDP is used as an evaluating parameter of economic level. However the more comparable indices could be the consumption level and construction condition of real estate in each city in the past years. The geographic and economic information is based on the latest annual report by the local government of each city.

2.5.2   Climate Zoning for Architecture Design and Corresponding Climate Characteristics of SU-ZHE-WAN Region

In China, vast land is divided into five climate zonings for architecture design. They are extreme cold zoning, cold zoning, hot in summer cold in winter zoning, hot in summer warm in winter zoning and mild zoning (Architectural Climate Zoning Standard, GB50178-93). Similarly, there are light climate zonings are divided according to average yearly daylighting time and critical illuminance of exterior daylight (Standard for Daylighting Design for Buildings, GB50033-2013).

The climate in SU-ZHE-WAN region represents the typical climate in southeast China, consisted of three provinces of SU, ZHE and WAN. This zone possesses subtropical monsoon climate with four distinctive seasons. Spring includes March, April and May; June, July and August for summer; September, October and November for autumn; and December, January and February for winter. Coldest month is January, whilst hottest month is July. All these three province locate in the same climate zoneIII according to the Architectural Climate Zoning Standard GB50178-93 , where typical climate is cold in winter but hot in summer. Similarly, the three provinces are all in the light climate zoning IV of 10 hours effective daylight hour at critical illuminance of exterior daylight of 5500lx. In climate zoneIII, the typical climate in summer or winter is partly shared with other zonings, which means the research results (guidelines) is fully suitable for design in hot in summer cold in winter zoning, and referable and partial applicable for other zoning design. The locations of all selected residential wards are plotted on figure A1 in Appendix A against the China architectural climate zoning map.

Most parts in Climate ZoneIII share the following climate characteristics: sultry summer, moist cold winter and low diurnal temperature variation; high annual precipitation; respectively lack of sunshine; continuous overcast and rain, frequent heavy rain at end of spring and beginning of summer; east coast area affected by tropical storms and summer typhoons, may resulting to a stormy-windy weather. The detailed climate descriptions related to architecture design in this zone are:

1. Average air temperature in July is between 25~30℃, average air temperature in January is between 0~10℃; minimum air temperature usually can reach -10℃, even -20℃, as a result of winter cold wave; The number of day in which its diurnal average temperature all through the year≥25℃, is between 40~110d, the number of day which its diurnal average temperature all through the year≤5℃, is

between 90~0d.

2. Annual average relative humidity is high and has little variation through the year, which is 70%~80%; annual rainy days are around 150ds, sometimes over 200ds; annual precipitation is 1000~1800mm.

3. Annual total solar irradiance is 110-160W/m^2.

The details of climate of all selected sites are listed in Appendix A. The table shows the climate type of the cities and their architectural climate zoning. The climate details include first generic temperature and humidity; second wind speed and dominant direction, which may have obvious influence on thermal simulation; third is sunlight irradiation, which will help in daylighting and thermal simulation.

2.5.3   Weather Data of Samples Adopted in Simulations

In Architectural Climate Zoning Standard GB50178-93, it only enumerates climate data of the representative cities and cities on the edge of the zoning. For the city which is not on the list, adopting the data of the most adjacent city of the same type on the list is suggested.

Because the selected cities share similar climate characteristics and for a purpose of simplify the settings procedures of climate and local coordinates for each of all 44 cases in thermal and sunlight simulations, an average climate data is applied in reference to all adopted cities. City Hefei is used for its climate, coordinates and annual statistical weather data provided by China Meteorological Data Service Centre in thermal and sunlight possibility simulation. Hefei locates in the middle of Anhui province, and it is the capital city of the province. Hefei locates in the central of Su-Zhe-Wan area in north-south span, and slightly to the west from centre on east-west span. This location provides the city typical subtropical monsoon climate with very clear four seasons and less direct impact from weather related to atmospheric movement over the ocean, as cities locates at the east coast in the area. The longitude and altitude of Hefei is used to calculate sun track and position and corresponding irradiation for sunlight simulations. For thermal studies, except for sun position and irradiation, its dominant wind direction and frequency C18 ENE 9 is also used. The annually peak wind direction is East Northeast, ranging from

106.88°-118.12°, with middle degree of 112.50°. The frequency of C18 ENE 9 shown on wind rose is explained as 9% of dominant direction, and is 18% of calm wind. Calm wind is defined as wind speed less than 0.5m/s at 10m above earth.

2.5.4   Model Scale and Simplification for Simulation

Rather than marco and micro, this research focus on meso-scale or intermediate scale of urban fabric. Regarding to building geometrical and morphological studies, large amount of researches are implemented at geological and urban scale. While regarding to micro climate topic, predominant studies concentrate on micro scale space surrounding the single target building or several buildings. This research keeps a foothold at meso-scale caring about micro-climate in between building clusters. The research interest is all the intermediate space or un-built space between residential buildings at ground level. Rather than directly focusing on green space or recreational space in the residential ward, evaluative analysis over all vacant space could provide supports for arrangement and distribution of outdoor amenity spaces. Sunlight simulations are also operated on the residential building facades as it directly or indirectly affect the indoor performance of the habitable rooms behind it; furthermore sunlight requirement at ground floor facade has higher priority than outdoor sunlight over ground.

For the comparability of the simulation data in each site, the simulation samples need to keep in similar scale. Hence, the residential wards in various sizes are divided into blocks of approximately 500×500m or below. The coding of the sample names, for example S260102, consists of S for site, 26 for residential ward number as seen in table A1 in appendix A, 01 for sub-division of the site if any, and 02 for second time sub-division if any. Sample name S34 refers to the ward 34 is modelled into S34 without any division.

Building model for simulation need to be simplified, to avoid unnecessary concern of non-significant factors and reduce calculation time. Building outline collected from general plan of the residential ward are use to build model, ignoring objects extruding from facade, including balconies, blinders, sun shadings, etc. The outline is the full shape of building footprint, based on which building blocks would be extruded to its

real height in model building tools. Because in finding the relationship between building distribution and environmental performance, building shape and geometries are included, further simplification of building shape into idealistic square or rectangular blocks as seen in previous research about building distribution is not considered. The specification of surface materials of buildings, roads and grounds are considered being the default materials in the simulation tools for each simulation. The impact of different material is not studied in this research to focus on the influences from building distribution and arrangement. Also, vegetation is not included in the research for the same reason. Furthermore, vegetation species and volume may vary case by case, that made case simulations lost comparability. Actually, vegetation will have great improve on environmental performance in all three aspects of acoustic, thermal and sunlight possibility, as seen in previous researches. Therefore, vegetation design could be considered as extra strategy applicable after determination of building distribution in the first design stage.

2.6 Statistical Methods Applied in This Research

This section explains the statistical analysis method used in this research including hierarchical cluster analysis and multiple linear regression (MLR). The mechanism of cluster and MLR expanded over the context of this research, as well as generalisation and model assessment of regression. The sample size is also discussed.

2.6.1   Hierarchical Cluster Analysis

Clustering means to categorise cases according to its data similarity measured by a certain type of distance. K-means and hierarchical clustering analysis is available in SPSS software.

K-means clustering only relies on Euclidean distance to cluster, and cluster centre is calculated based on average of distance. The initial cluster centres need to be decided arbitrary, as well as the amount of clusters, namely a successful clustering in K-means requires several tests. K-means method is more sensitive to cluster size and tends to general cluster of similar size. The advantage of K-means method is its effectiveness in calculation. However its result remains less accuracy compare to hierarchical

clustering, it is suggested for exploring data preliminarily. Because for this research, the amount of clusters is unknown before analysis and the data amount is small, the advantage of K-means could not be expressed but its short coming makes impact. Therefore hierarchical clustering analysis is adopted.

Hierarchical clustering initially defines all samples as separate clusters, and combines them one by one, until all samples are merged into one cluster. The result is recorded as a tree structure, in a dendrogram. Multiple cluster method could be selected from in SPSS, as well as multiple interval measures. In this research, between-group linkage (average-linkage) method and measure of squared Euclidian distance are adapted. The merit of hierarchical clustering is that the all possible ways of clustering for one set of data are covered in dendrogram after one time calculation. The amount of clusters and selected cluster structure could be decided based on the tree structure after calculation. Therefore, redefining cluster number does not need to recalculate distance in data. The shortcoming of hierarchical clustering is the calculation speed. However, data size in this research is not particularly large, so speed would no make huge difference. Also, it could be inferred that building distribution characteristics should form clusters of different size due to the market requirement of residential ward form is uneven. Hence, the insensitivity to cluster size of hierarchical clustering is suitable for this research.

Rescaled distances between clusters are basis of cluster division. During selection of proper calculation method for rescaled distance, two factors worth note: the monotonicity of the rescaled distance and the balance of distance range between concentration and expansion. Except for centroid linkage method, other calculation methods share rigid monotonicity in rescaled distance. The impact of distance range on cluster results shows as small or concentrate range leading to insensitivity in clustering, while large or expanded range leading to over sensitive clustering where sub-cluster may outstanding over main structure. Sorted the range of distances in order from concentration to expansion, the corresponding calculation methods are: nearest neighbour linkage, group-average linkage, furthest-neighbour linkage and centroid linkage, group-average linkage, Ward's linkage. Therefore, group-average linkage is adapted in this research.

The method of calculating distance, the most suggested is group-average method,

including between-group linkage and within-group linkage. Between-group linkage would calculate average distance between two clusters, while within-group linkage would include distance between all pairs of samples from two clusters and within the clusters for average. Between-group linkage is utilised according to the definition. Euclidean distance calculation which is applied in sample clustering relies on the point coordinates generated by sample matrix. In the condition of variable clustering, correlation coefficient is used as proximity measure.

The cluster division is considered based on the degree of rescaled distance between adjacent clusters. When two clusters combined have significant differences, the rescaled distance on dendrogram would be obviously large, and vice versa for two clusters of small differences. The determination of number of clusters accords to the criterion of select appropriate number of clusters from dendrogram is based on the theory by Demirmen. firstly, each cluster is significant, with steep increase of rescaled distance between them; secondly, number of cases in each cluster is not too large compared to other clusters; thirdly, the number of cluster makes sense in accordance to practice; fourthly, same cluster should appears in results by different method of clustering calculation.

Multiple calculation methods are tested for clustering. The combinations of method and calculation measures are: between group method with square Euclidean distance; within group method with square Euclidean distance, single linkage method with square Euclidean distance, centroid method with square Euclidean distance, and between group with correlation coefficient. Most combinations provide clustering result either invalid due to violation of hierarchical clustering principle, either biased by certain factor that hides other characteristics. Between group method with square Euclidean distance is ascertained for clustering in this research.

2.6.2   Multiple Linear Regression and the Generalisation Beyond the Sample

Multiple linear regressions is seeking the model in the form of equation by finding the linear combination of predictors that correlate maximally with the outcome variable (equation 3.1). The calculation method is least square method that is fitting a model to the data for which that sum of the squared differences between the plane and the

actual data points is minimised.

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_n X_n + \varepsilon_i \quad (3.1)$$

Hypothesis test before regression is aiming to assure the data based on is valid for the generalisation of the prediction model from the current sample to the whole population. When the assumptions of regression are met, the model regressed from the data sample could be able to generalise to the population of interest, namely, the parameters applied for the regression equations and their coefficients are said to be unbiased. When no outliers, if assumptions are some how violated, it is to say the model fit perfectly to the sample data, but may not able to generalise. However, even when the assumptions are met, the sample model may not be the same as the population model, but the likelihood of being the same is increased. Therefore, it is still necessary to test hypothesis assumption before regression.

The ten hypothesis required to be checked are: 1) variable types, 2) non-zero variance, 3) independence, 4) uncorrelation with external variables in predictors, 5) non perfect multicollinearity, 6) homoscedasticity, 7) independent errors, 8) normally distributed errors, 9) linearity, and 10) undue influential cases.

Details of the ten hypothesis are described as below:
1) Our research applies quantitative, continuous and unbounded predictors and outcome variables.
2) Variances of all independent variables are tested separately, being not zero.
3) Independence refers to all of the value from the outcome variable are independent.
4) To the current knowledge, applied predictors are not correlated to external variables which are not included in the regression model but do influence the outcome. However, further work is still needed to discover possible influential parameters to predict multiple outdoor environmental performances.
5) Correlation matrix between outcomes and predictors has been viewed for a preliminary look on multicollinearity. Strong correlation between predictors in regression model will cause problem, but less than perfect collinearity is actually unavoidable. The criteria of perfect collinearity is correlation coefficient between predictor larger than 0.8/0.9. Variance inflation factor (VIF) can also be used to assess multicollinearity, and value over 10 is worth concerns. Similarly, tolerance smaller

than 0.1, will cause serious problem in regression model. Seen in the correlation matrix table in the regression results, it is concluded that there is some extent correlation between predictors but no perfect linear relationship between predictors.

6) Homoscedasticity refers to the assumption that the spread of scores (variance) is proximately equal in different group of cases. This could be tested by Levene's test and chart of standardised predicted value (ZPRED) against standardized residuals (ZRESD), histogram of residuals, partial plot. Plots are also used for model assessment, so that will be discussed in following section.

7) Independent error refers to the residual of any two observations is uncorrelated, tested by Durbin-Watson test with criterion of equals to 2. If smaller than 1 or larger than 3, it may cause problem. Seen in regression summary tables, our regression results have test result close to 2.

8) Normally distributed errors mean the difference between the model and the observed data are most frequently zero and the difference greater than 0 only happens occasionally. This would be checked by residual histogram and normal probability plot. The standardised residuals are converted from un-standardised residuals into standard deviation unit, through dividing them by the estimate of their standard deviation. If normality exists, the standardised z-score should be normally distributed around a mean of 0 with a standardised deviation of 1, appearing in bell shape; and the probability plot should appear all points lie closely along the line. The regressions in this research all passed this hypothesis test.

9) Linearity means the relationship of interest is a linear one. Plot of standardised residual against standardised predicted value, it is used for checking assumption of linearity between x and y, random error and homoscedasticity have been met. When existing linearity and homoscedasticity, scatter appears in random array of dots evenly dispersed around 0. Partial plot is scatter plot of residuals of the outcome variable and each of the predictors when they are both regressed separately on other remaining predictors. It could be used to discover outlier and non-linearity between x and y and heteroscedastisity. If normal, the points should evenly spaced out around the line, rather than funnel shape or curve shape, that all regressions passed this test.

10) Existence of influential case is also examined. Removing case exerted undue influence over the parameters of the model, could improve the stabilisation of the model over the sample. Cook's distance helps to consider the effect of a case on model as a whole. There is no case has cook's distance over 1. Leverage value is also

checked, all case have value values between 0 and 1. Mahalanobis distance is examined and no case has value over 10. Also casewise diagnose by checking standard residual over 3 is conducted, and it validated that there is no outliers in the data. Based on these, it could be said that no outlier exists in the sample.

The table 2.3 summarised the hypothesis tests conducted before and after the regressions in this research.

Table 2.3 Hypothesis Test Conducted and the Results of Data for APSH Regression

	Content of Test	Calculation Approach	Criteria	Result	Violation
1	variable types			quantitative and continuous	No
2	non-zero variance		being not zero	being not zero	No
3	independence		value from the outcome variable are independent		No
5	no perfect multicollinearity	Correlation matrix	correlation coefficient between predictors < 0.8/0.9	Coefficients<0.8	No
		VIF	largest VIF value < 10; average VIF >> 1, regression is biased	VIF<10 and Tolerance>0.2	
		tolerance	<0.1serious problem <0.2 potential problem	no multicollinearity	
6	homoscedasticity	The Levene's tests	significance over 0.05	significance over 0.05	Slightly yes
		Plot of ZRESID agaist ZPRED	points appear in random array of dots evenly dispersed around 0	Shows a random array of dots	
		Partial plot	The cloud of dots evenly spaced around the line	evenly spaced around the line or with slightly funnel shape	
7	independent errors	Durbin-Watson test	Equals/close to 2; 1<Durbin-Watson<3	Equal/close to 2	No
8	normally distributed errors	Histogram of residuals	Normal distribution, mean of zero	Normal	No
		Normal probability plot of residuals	Points closely along the line		No
9	linearity	Plot of ZRESID agaist ZPRED	relationship of interest is a linear one	normal	No
		Partial plots between IV and DV	scatter appears in random array of dots evenly dispersed around 0		No
10	Influential cases	Cook's distance	Cook's distance<1	Normal	No
		Leverage value	have significance>0.05		
		Mahalanobis distance	Mahalanobis distance<10		

Based on the test, the model appears accurate for the sample and generalisable to the population.

2.6.3   Settings of Multiple Linear Regression and Model Assessment

The multi-linear regression is carried out in SPSS22. The method of entering the variables into the model is stepwise method, which means each time a predictor is added to the equation by selecting the one with the highest simple correlation with the outcome; meanwhile a removal test is implemented of the least useful predictor. Thus the possible redundant predictors can be removed in this constantly reassessing progress. for the stepping method criteria, the probability of F used for variable entry is 0.5, for variable removal is 0.1. Hieratchical entry is not used, as there is no prioritise indices for the regression equation. Method for missing data points is excluding cases listwise.

To assess the goodness of fit of the model, multiple correlation coefficient R is examined as it is the gauge of how well the model can predict the observed data. R square indicates the amount of variance in the outcome explained by the model. In this research, the acceptable threshold of R square is defined to 0.5, and 0.8 for good fitness.

Adjusted R square is applied to calculate the loss of predictive power when the model derived from the population from which the sample was chosen. The equation of calculating adjusted R square is of Wherry's. Adjusted R square have punish on the number of selected independent variables, so that it is checked in attempts of regression when adding in building distribution variables.

F-test is also conducted for significance of the whole regression model. The significance level is checked in the summary table of regression result. The criterion is below 0.05 to be considered significant.

In table of estimates, estimated b value could be read with t-test significance. T-test is conducted for the hypothesis that the value of b is significantly different from 0 and

that the independent variable contributes significantly to the ability of estimate values of the outcome. If significant ($p<0.05$), means the predictor has made significant contribution to the model. Confidence intervals could be provided for each unstandardised regression coefficients in coefficient table. The likely value of b for the population will locates with in the interval. The standardised coefficient beta makes it comparable that to what degree the predictors are important to the regression model.

The residuals represent the error in the model and normality in error is required in hypothesis test. If the model fits well to the observed data, all residuals should be small. Checking all standardised residuals of cases, no more than 1% of the sample cases has standardised residuals with absolute value over 2.58, and less than 5% with with absolute value over 1.96. Violation indicates that the model may be a poor representation of the actual data. But due to the small size of sample, slight violation is still acceptable.

2.6.4   The Discussion Regarding to Sample Size

It is suggested that the minimum acceptable sample size should be 5-10 times of number of independent variables. As the current sample size is 44, the proper number of building distribution variables used in regression is 4-8. However, for small samples, less independent variables is better for the stabilisation of the regression model. According to Green (1991), for aim of testing the overall fitness of the regression model, the minimum sample size is suggested to be 50+8k (k is the number of predictors). The sample of 44 cases used in this research, could not meet Green's minimum standard.

Shown in some regressions, the scatter plots of standardised predicted values by standardised residuals show funnel shape and open to left. This means the sample is biased and not fully covers the whole population. More samples on the side which is not sufficient are required to be supplemented.

2.7 Summary

This chapter firstly introduces the background and previous works of holistic design, decision making support at early design stage, proactive simulation, and global optimisation, in the context of architectural and urban design. With the discussion about the background and previous works, the research interest and aims are confirmed.

The chapter secondly introduces and compares the evaluative standards of the research interest performance domain from UK and China, answering the question of what is a good integrated environmental performance. Thirdly, the chapter expatiates the sampling selection and site information according to the aforementioned research interest, answering the question of what is researched.

At last the statistical tools applied in this research are explained and discussed regarding their hypothesis test and validity.

This chapter provides the information of where and how to start this research. It contributes to the whole research at 1. To understand of the previous works in holistic design, MD-MOO and integrated environment improvement in the early design stage; 2. To decide the research objectives and questions; 3. To confirm the standard of the objective; and 4. To confirm the tools to support the research.

Chapter 3 Individual Building and Neighbourhood Morphology Parameter Study

This chapter aims to understand the parameters which describe building distribution in a residential ward in the context of South-east China, from the aspect of single buildings and neighbourhoods. The conventional building distribution parameters applied in the design practise and architectural research are studied, with the innovative parameters extracted from the sample data in this research. A better understanding of parameters would support the following studies: environmental performances under the impact of building distribution patterns, the corresponding optimisation in each performance domain and the global optimisation aiming to improve the balanced multi-domain performance.

Statistical tools are applied to analyse the building distribution parameters of the sample site according to their data characteristics. Hierarchical clustering and multiple correlation analyses (one tailed) are used for the grouping of the parameters and for discovering their paired connections and interactions.

As a result, all adopted building distribution parameters including those of single buildings and neighbourhoods are clustered into 12 groups. Key parameters are selected from each group serving as design parameters in the following multiple linear regression and global optimisation. The paired connections between the distribution parameters reveal the residential ward planning rule and design habits in the SU-ZHE-WAN region of China. This enables the designers to look at how they usually organise design patterns in practise, to understand whether it might be possible to escape from a fixed habit to an optimised arrangement.

The structure of this chapter is: 3.1 Definition and Application Background of Individual and Neighbourhood Morphology Parameters; 3.2 Consolidation of Morphology Parameters for Simulation; 3.3 Data Collection and Transformation for Individual Building and Neighbourhood; 3.4 Cluster Analysis for Morphology Parameters to Define Key Parameters; 3.5 Level Grading for Key Morphology Parameters; 3.6 Correlation and Interaction between Key Morphology Parameter Pairs and 3.7 Summary. The flowchart of this chapter is shown in figure 3.1.

I. Background

1. Definition and Application Background of Individual Morphology Parameters

II. Data Collection

2. Consolidation of Morphology Parameters for Simulation

3. Data Collection and Transformation for Individual Building and Neighbourhood

III. Data Analysis

4. Cluster analysis for Morphology Parameters to Define Key Parameters

5. Level Grading for Key Morphology Parameters

6. Correlation and Interaction between Key Morphology Parameter Pairs

Figure 3.1 Content Structure of Chapter 3 Individual and Neighbourhood Morphology Parameter Study

Acronyms for Chapter 3

aCAH/AvgCornerHigh	average corner area high value
aCAL/AvgCornerLow	average corner area low value
aCAmean/AvgCornerArea	average corner area
aD/AvgDistance/Dmean	average distance to road of a site
aFHmean/AvgFrontHeight	average front façade height
aFL/AvgTotalFaçadeLength	average total façade length
aFLmax/AvgMaxFaçadeLength	average max façade length
aFLmean/AvgFaçadeLength	average façade length
aFLmin/AvgMinFaçadeLength	average min façade length
aFR/AverageFaçadeRatio	average façade ratio
aFSmean/AvgFrontStorey	average front façade storey
aIAmean/AvgIntervalArea	average interval area
aIDH/AvgIntervalDepthHigh	average interval length high value
aIDL/AvgIntervalDepthLow	average interval depth low value
aILmax/AvgMaxIntervalLength	average max interval length
aLFL/AvgLowFaçadeLength	average low-rise façade length
aLFR/AvgLowFaçadeRatio	average low-rise façade ratio
aOL/AvgOutlineLength	average outline length
AR/AspectRatio	aspect ratio
RS/AvgResiStorey	average residential storey
TCA/TotalCornerArea	total corner area
BD/BuildingDensity	building density
DL/DiagonalLength	diagonal length of site

Dmax	max distance of the front facade to the faced road in one site
DoCE	the design of computational experiments
FPA/FootprintArea	foot print Area
HRBA/HighriseResiArea	high-rise building area
HRBR/HighriseRatio	high-rise ratio
LMRBA/Low/medium-riseResi Area	low/medium-rise building area
MOO	multi-objective optimisation
OGBA/OvergroundBuiltArea	over ground building area
PR/PlotRatio	plot ratio
RBA/ResiBuildingArea	residential building area
RFPA/ResiBuildFootprintArea	residential foot print area
SF/ShapeFactor	site shape factor
sRC/Resicircumference	residential circumference
sRSA/ResiSuperfacialArea	residential superficial area
SVF	sky view factor
TIN	The triangular irregular network tool
TLA/TotalLandArea	total land area
TSD/TriangleSD	standard deviation of triangle area

3.1 Definition and Application Background of Individual and Neighbourhood Morphology Parameters

The previous review articles regarding multi-objective optimisation (MOO) applied in architecture and urban field lack analysis on application trend and adopted design variables and objectives (Ekici et al., 2019). Lack of support from previous research about optimisation design variables and objectives is always the challenging point in this research. Hence, this research attempts a pioneer study of building and neighbourhood morphology parameters.

In the application of architectural design and urban planning, morphology parameters are used to describe attributes of individual building geometry, building group relationships or city-scale patterns. For each interest scale the explanatory parameters are different. The micro scale parameters describe the geometry of a single building and building component. The meso-scale parameters include single building geometry parameters as well as parameters describing positional relationships between

buildings and building groups. The macro-scale parameters focus on the spatial relationships among large numbers of buildings from analytical and statistical viewpoints. The meso-scale parameters are the bridge to combine research focussed on a single building and an urban context.

Regarding the micro-scale studies, which are plentiful in the architectural design field, there is a considerable number of detailed parameters, representing the geometrical and other physical attributes of buildings, building components, and building materials. The geometrical metrics are often used in building form related studies, while metrics of physical attributes are mostly adopted in performance related research.

For macro-scale studies landscape metrics are more commonly used. These are algorithms that quantify specific spatial characteristics of elements using categorical maps including patches, classes of patches or entire landscapes, land-cover, land-use mosaics, etc. (Turner and Gardner, 1991). It is applicable when land-use maps are available for urban form analysis (Schwarz, 2010). Landscape metrics include two categories, those focussing on the composition of the map, and those quantifying the spatial configuration of the map. The spatial information is required to calculate the spatial configurations (McGarigal and Marks, 1995; Gustafson, 1998). However, this will not provide enough information about building blocks. Metrics representing meso-scale urban structure are needed for local residential ward research.

As in this research, residential wards are the research interest. This could be an urban texture with considerable scale, where urban scale morphology is worth considering; however single building form would also strongly impact the environment performances. Geometrical parameters describe the scale of length and area of buildings or the space in-between buildings. Morphological parameters describe positional relationship between buildings. The combination of both could represent a holistic image of a certain type of distribution pattern.

The combination of geometrical and morphology parameters adopted in this research is capable of describing the significance of the building distribution attributes of a site sample, however it is not sufficient to distinguish a sample by its distribution

characteristics from the remaining samples. In other words, the combination of building distribution parameters is a necessary condition to describe the distribution pattern and characteristics, rather than a sufficient condition which could be used to remodel a pattern in high-dimensional modelling.

3.2 Consolidation of Individual and Neighbourhood Morphology Parameters for Simulation

3.2.1 Convention and Innovation

A group of geometrical parameters are conventionally applied in residential ward planning by architects and urban planners. These include original geometrical parameters of length, height, and area, and derived geometrical parameters of ratio, factor, and coefficient.

The conventionally used original geometrical parameters include: total land area, foot print area, over ground building area, residential building area, residential foot print area, average residential storey, low/medium-rise building area, and high-rise building area. The derived geometrical parameters include plot ratio, building density, aspect ratio. These parameters loosely constrain the overall condition of a site on scale, intensity, and density. However, the distribution pattern is influenced by other social-economical factors and local design habits.

To provide more specific descriptions of a distribution pattern, several innovative parameters are created in this research. These include supplementary geometrical parameters of distribution pattern details and innovative morphological parameters of building distribution. Supplementary geometrical parameters are not usually used in the present design procedure. In this research they are collected and compared to describe building detailed distribution.

Supplementary geometrical parameters include front façade scale and proportion, front interval scale and proportion, façade height scale and derived parameters of overall site condition. First, parameters of front façade scale and proportion include average distance to road, average outline length, average total façade length, average

façade ratio, average max façade length, average min façade length, average façade length, average low-rise façade length, and average low-rise façade ratio. Second, parameters of front interval scale and proportion consist of average max interval length, average interval depth low value, average interval length high value, average interval area, average corner area, average corner area low value, average corner area high value, and total corner area. Third, parameters of façade height scale include average façade storey and average façade height. Finally, supplementary parameters of overall condition are residential circumference derived from conventional geometrical parameters, comprised of residential superficial area, high-rise ratio, diagonal length of site and site shape factor.

Innovative morphological parameters include indicators of residential building position relationship, which is a standard deviation of triangle area (TSD) and sky openness indicator, which is sky view factor (SVF). The site's sky-view factor (SVF) indicates the amount of visible sky at a given point outdoors (Chapman, Thornes and Bradley, 2001). Table 3.1 shows the summary of conventional and innovative parameters, together with their definitions, unit, range, collection approach and data source.

Other morphological parameters could be considered in future research. For example, parameters regarding relative positions between the open spaces, the low-rise buildings, the mid-rise buildings, the high-rise buildings, and the building-site boundary, etc. could be created or borrowed.

3.3 Data Collection and Transformation for Individual Building and Neighbourhood

3.3.1 Collection of Geometry Data

Values of the aforementioned original geometrical parameters are collected and consolidated from corresponding site plans of each sample site's construction drawing through AutoCAD. The collection of measured parameters value refers to figure 3.2 for the notations. The dimension tool in AutoCAD is used for measurement. Derived

geometrical parameters are calculated based on original geometrical parameters as shown in figure 3.2, following each of their definitions described in table 3.1.

The large variation of order of magnitude between parameters made it necessary to standardise the data to eliminate the influence of magnitude differences for the following statistical procedure of clustering, correlation, regression and optimisation. It is worth noting that shape of a patch is a spatial attribute extremely difficult to capture because of the infinite possibility of patch shapes. The shape complexity could appear simple, compact, irregular or convoluted. The shape metric usually holistically considers the shape complexity from an analytical view through calculating the parameters of selected metrics. The most commonly accepted measures of shape complexity are perimeter-to-area ratio or fractal dimension, which requires standardisation to a simple Euclidean shape. Other shape metrics were once proposed while not widely used (Gustafson, 1998). In this research the residential wards could be generally simplified to rectangular, hence the relative ratio parameter is suitable. Instead of the perimeter-to-area ratio, the diagonal-to-edge ratio is used for those rectangle patches.

Table 3.1 Details of All Building Distribution Parameters

Category		Parameter	Definition	Abbreviation	Unit	Range		Collection Approach	Data Source
						Lower Limit	Upper Limit		
Conventional	Original Geometrical	total land area	Description of site size.	TLA	10000m^2	2	42	Measure	AutoCAD drawing
		foot print Area	Footprint area of all buildings within one site.	FPA	10000m^2	0.4	9	Measure	
		over ground building area	Total building area which is above ground, except basement and underground car park.	OGBA	10000m^2	2	65	Measure	
		residential building area	Total building area of all residential buildings.	RBA	10000m^2	2	63	Measure	
		residential foot print area	Footprint area of all residential buildings in a site.	RFPA	10000m^2	0.3	7	Measure	
		average residential storey	Average storey of all residential buildings.	aRS	m	4	36	Calculation	
		low/medium-rise building area	Total residential building area of low-rise, medium-rise buildings (1F-10F).	LMRBA	10000m^2	0	35	Measure	
		high-rise building area	Total residential building area of high-rise buildings (>10F).	HRBA	10000m^2	0	63	Measure	
	Derived Geometrical	building density	Coverage ratio of a site reflecting density of a site at horizontal level, = FPA/TLA	BD	NA	0.07	0.4	Calculation	
		plot ratio	Reflecting density of a site or developing intensity at space level. = RBA/ TLA	PR	NA	0.6	3.6	Calculation	
		aspect ratio	Reflecting the ratio of height vs width of un-built space of at space level. Applied in this research: =total superficial area of all residential buildings/(TLA-FRA)	AR	NA	0.5	4	Calculation	

Innovative: Supplementary geometrical	Original geometrical: front façade scale	average distance to road	The distance of a row of front facing façades to the centre axis of a corresponding street. It is the average of values over all edges of the site boundary.	aD	m	13	61	Measure, calculation	
		average outline length	The length of one edge of the boundary. It is the average of values over all edges of the site boundary.	aOL	m	139	606	Measure, calculation	
		average total façade length	The sum of front façade length on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFL	m	86	383	Measure, calculation	
		average max façade length	The maximum of façade lengths on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFLmax	m	31	113	Measure, calculation	
		average min façade length	The minimum of façade lengths on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFLmin	m	12	82	Measure, calculation	
		average façade length	The average of façade lengths on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFLmean	m	18	92	Measure, calculation	
		average low-rise façade length	The total length of low-rise façades on one edge of the boundary. It is the average of values over all edges of the site boundary.	aLFL	m	0	219	Measure, calculation	
	Derived Geometrical: Front Façade Proportion	average façade ratio	Reflecting the percentage of façades on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFR	%	42	91	Calculation	

			=FL/OL					
		average low-rise façade ratio	The ratio of low-rise façades among all façades on one edge of the boundary. It is the average of values over all edges of the site boundary. =LFL/FL	aLFR	%	0	93	Calculation
	Original Geometrical: Front Interval Scale and Proportion	average max interval length	The maximum interval length on one edge of the boundary. It is the average of values over all edges of the site boundary.	aILmax	m	18	86	Measure, calculation
		average interval depth low value	An average value of the shorter side length of every interval on one edge of the boundary. It is the average of values over all edges of the site boundary.	aIDL	m	10	36	Measure, calculation
		average interval length high value	An average value of the longer side length of every interval on one edge of the boundary. It is the average of values over all edges of the site boundary.	aIDH	m	11	56	Measure, calculation
		average interval area	An average value of all interval area on one edge of the boundary. It is the average of values over all edges of the site boundary.	aIAmean	m^2	183	3729	Measure, calculation
		average corner area	The average of two corner areas on one edge of the boundary. It is the average of values over all edges of the site boundary.	aCAmean	m^2	0	6519	Measure, calculation
		average corner area low value	The lower value of the two corner areas on one edge of the boundary. It is the average of values over all edges of the site boundary.	aCAL	m^2	0	913	Measure, calculation
		average corner	The higher value of the two corner	aCAH	m^2	0	5837	Measure,

		area high value	areas on one edge of the boundary. It is the average of values over all edges of the site boundary.					calculation	
		total corner area	The summation of two corner areas on one edge of the boundary. It is the average of values over all edges of the site boundary.	aTCA	m^2	0	26073	Measure, calculation	
	Original Geometrical: Front Façade Height Scale	average front façade storey	The average of façade storeys on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFSmean	NA	2	23	Measure, calculation	
		average front façade height	The average of façade heights on one edge of the boundary. It is the average of values over all edges of the site boundary.	aFHmean	m	7	69	Measure, calculation	
	Derived Geometrical: Overall Site Condition	residential circumference	The summation of circumference of all residential buildings	RC	m	900	19000	Measure	
		residential superficial area	The summation of superficial area of all residential buildings. =RFPA×residential building height	RSA	m^2	17000	480000	Calculation	
		high-rise ratio	The proportion of high-rise residential building area to total residential building area. =HRBA/RBA	HRBR	NA	0	1	Calculation	
		diagonal length of site	The longer diagonal length of a site, reflecting scale of site	DL	m	203	897	Measure	
		site shape factor	A ratio reflecting the shape of a site. =DL/OLmin	SF	NA	1.5	4.5	Calculation	
Innovative morphological	Residential Building Position	standard deviation of triangle area	Triangle area is constrained by three adjacent centroids of three residential building footprints in	TSD	NA	60	1200	Measure, calculation	ArcGIS

	Relationship		the triangle network all formed by centroids. It is the standard deviation of all triangle areas of a site, used for the comparison of the fluctuation of area value between sites.						
	sky openness	sky view factor	The percentage of viewable weighted sky patch areas of a sky semi globe at a point.	SVF	%	50	77	simulation	Rhino+Grasshopper
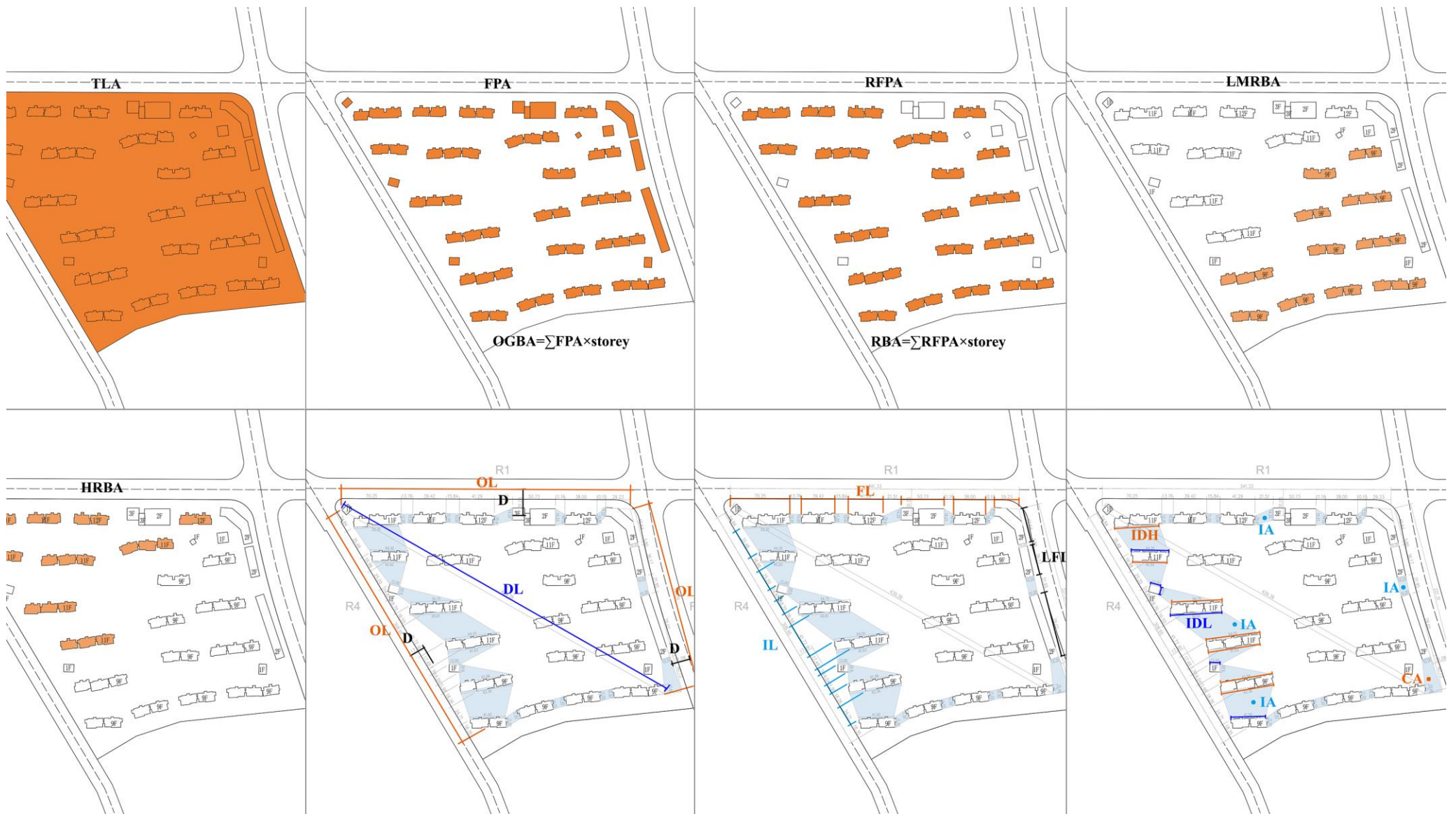
Figure 3.2 Illustrators of the Morphological Indicators over S06 Dimensional Map

3.3.2 Collection of Morphology Data

Indicators of morphology applied in this research are standard deviation of triangle (TSD) and sky view factor (SVF) describing distribution spacing and view openness to the sky, respectively. For large-scale cases, the dimensional value of morphology parameters could be collected (as is the case in some research) by a random sampling algorithm scripted in Matlab (Nault et al., 2017). Conventionally, SVF at each monitoring point can be acquired by assessments of fisheye images and calculations with RayMan (Matzarakis, Rutz and Mayer, 2010) at the approximate height of a grid setting. In this research of meso-scale sites, morphology information is calculated in ArcGIS and Rhinoceros+Grasshopper.

Distribution spacing is a created parameter describing the spacing between residential buildings. To define the residential building separation in a certain site, the centroid of each residential building footprint is collected with its two-dimensional coordinates. The network of the centroid which consists of triangles is generated with ArcGIS. Each of the vortices of one triangle is one of three centroids of building footprints from the most adjacent buildings. Distribution spacing is the area of one specific triangle within the network of centroids.

AutoCAD drawing is used to generate the list of centroid coordinate values of residential building footprints. The list containing centroid coordinates is imported into ArcGIS and converted into a shape file containing points.

The triangular irregular network (TIN) tool in ArcGIS is used for representing building morphology by generating a triangular network based on a centroid point matrix. TIN is vector-based digital geographic data. It is constructed by connecting a set of vertices with edges to form a network of triangles.

The generation of a triangle network applies the Delaunay triangulation method. The algorithm makes sure no vertex locates inside any of the circumcircles of the triangles in the network. By this method, the minimum interior angle of all triangles would be maximised. A thin and narrow triangle is avoided in the generation process.

Each site would generate one triangle network on one flat surface with an elevation of zero (Figure 3.3). The area information of all triangles in the network will be extracted to tables for further analysis.
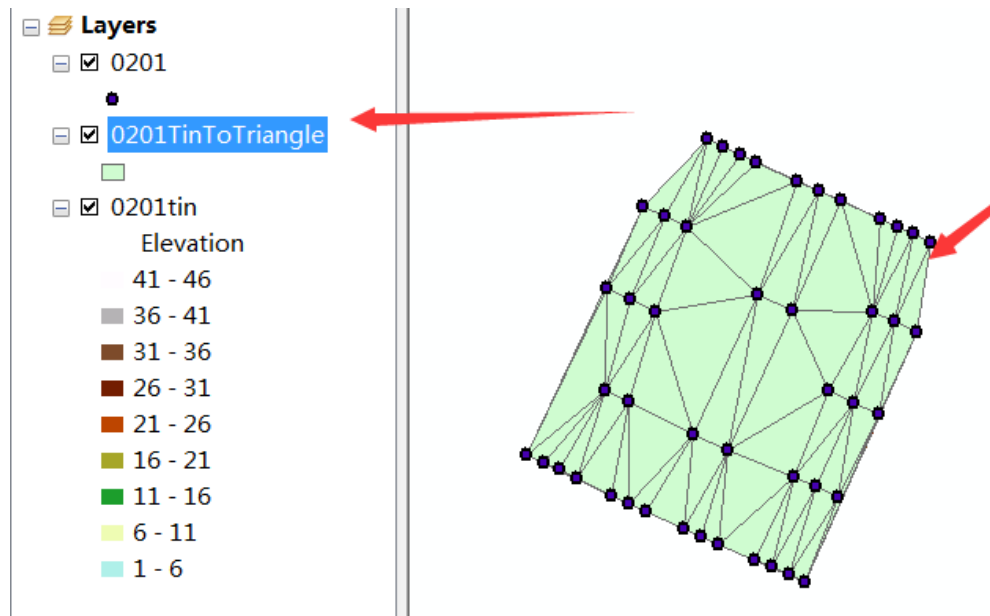
Figure 3.3 Triangular Irregular Networks by Residential Building Centroids (S0201)

ArcGIS is also capable of calculating minimum distance between adjacent points in a matrix. This is a possible alternative parameter for the representation of distribution spacing. The standard deviation of minimum distance could also describe the clustering attributes of buildings. This could be added into the list of parameters in future study.

In this research, the SVF data is collected through the simulation of 3D models of each site. Rhinoceros is applied as a model building platform and Grasshopper is applied as a programming and simulation platform. Calculation of SVF is carried out in the Ladybug view analysis component on the Grasshopper platform.

Simplified residential buildings of blocks are input as obstacle geometries. The viewpoints are set over ground at 0.001m on grid points with grid intervals of 1m (figure 3.4). The output includes an SVF map and a list of SVF values on all grid points (figure 3.5).
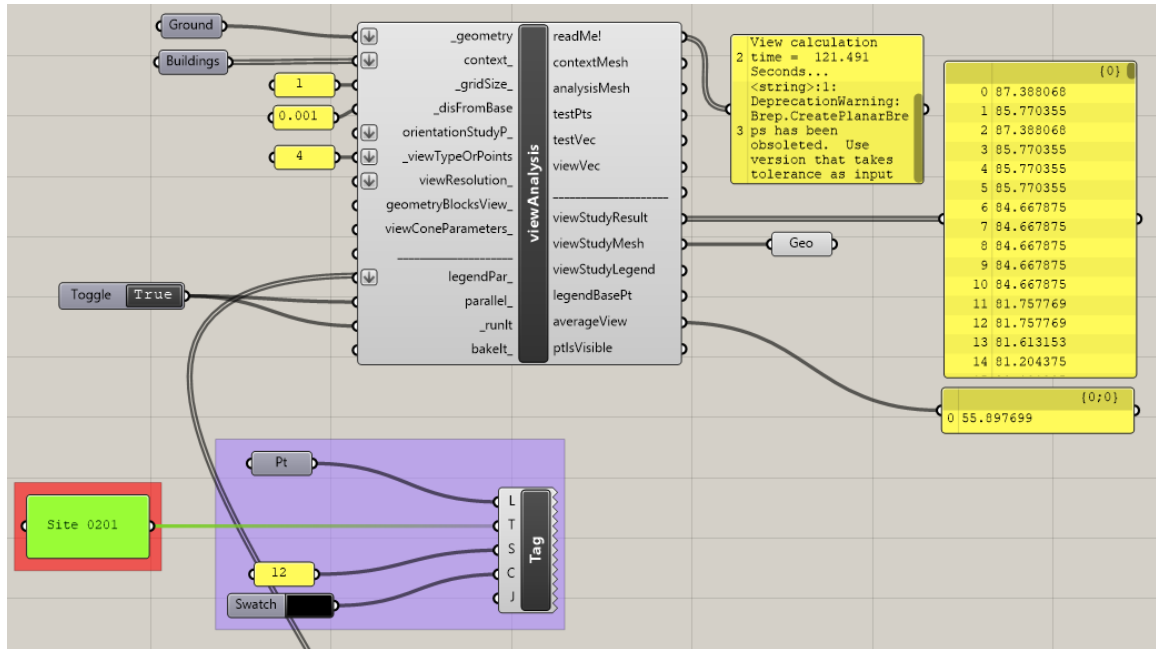
Figure 3.4 SVF Calculation Settings in Grasshopper
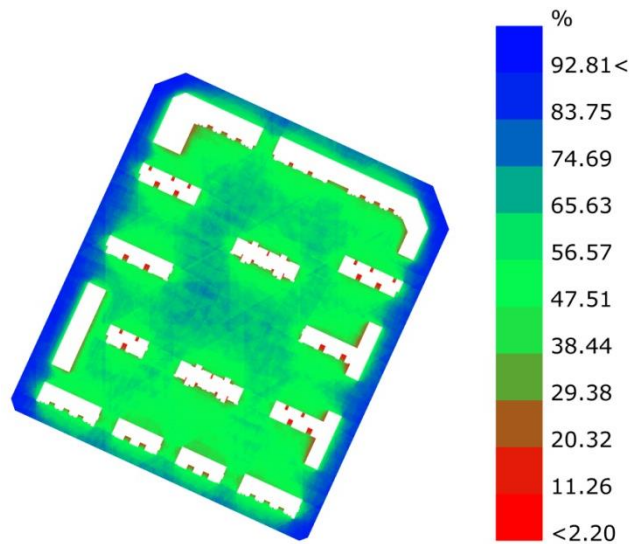
Site 0201
Sky View Analysis
Figure 3.5 SVF Map (S0201 as Example)

3.4 Cluster Analysis for Morphology Parameters to Define Key Parameters

In this section, parameters collected via the aforementioned method are analysed regarding to their inner connection which crowds them into groups. Statistical analysis of clustering by variables is used for grouping parameters. Two attempts at clustering are operated: one attempt includes all building distribution parameters describing summary condition, in-site condition, and front façade conditions; and the

other attempt only includes the parameters describing front façade conditions. The clustering of all building distribution is in service for acoustic, thermal and sunlight performance analysis across a site. The clustering of parameters regarding front façade condition is in special use for the acoustic performance analysis of front and behind façades.

The most influential and representative parameters are selected from each parameter groups as key independent parameters for the further regression and optimisation regarding environmental performances.

3.4.1   Clustering for All Distribution Parameters

All building distribution parameters including the parameters describing overall site conditions, in-site conditions and front façade conditions are analysed in this section. According to the calculation of their between group linkages based on squared Euclidean distances, a dendrogram is generated presenting the cluster structure of the groups. Through the observation of the structure, and with consideration of the possible numbers of clusters and meaning of each possible cluster, 12 groups are divided accordingly. To eliminate the impact of data scale differences, all data is standardised before clustering.

The clustering results are summarised in table 3.2, including the groups of parameters and the connotations of each group. Cluster 1-7 refer to overall and in-site conditions, and cluster 8-12 refer to front façade conditions. Seen from the connotations of all groups, the main categories of attributes of a residential ward are site scale, built amount, high-rise building amount, density, openness, set back amount, and site shape. Similarly, the main categories of attributes of the site frontier boundary are amount of low and long façade, short façade, large intervals, small intervals, and depth of intervals.

Table 3.2 shows that a group of parameters are under significant influence of the site scale (cluster 1) and the existence of high-rise buildings (cluster 3). These parameters are clustered into the same group although the nature of each parameter is different. There are 8 parameters describing the overall site conditions under the control of the site scale. The footprint area and low/medium-rise residential building area are

representative parameters relating to other distribution attributes; however, their explanation power is covered and highly impacted by the site size. Similarly, TriangleSD as a descriptor of evenness of distribution is highly bound to the existence of high-rise buildings. Comparable bound conditions happen to the built amount (cluster 2) and front interval amount (cluster 10), but are not as severe as those of clusters 1 and 3. The other clusters which have fewer parameters means less correlation on other parameters, and parameters in these clusters are more independent to others.

The phenomenon of parameters with various natures of distribution attributes clustered into the same group shows the binding relationship between the parameters in design practise in the SU-ZHE-WAN region. Design habits of planners in this region could be extracted from this data analysis. However, based on this data set, the regression of performance indices to building distribution parameters would definitely be influenced by this parameter-binding relationship from the design habits. The expected parameter may not appear in the regression equation due to the masking effect of other related influential factors. The influential parameters would be more dominant in regression progress.

The characteristic of building distribution presented by the sample data bares the influence of local design habits. This impact would appear in regression and optimisation of in-site environmental performance regarding building distribution. Therefore, the extraction of rules by regression using the current data set is biased in relation to its design habit and will fit the current design convention.

If the elimination of the influence from local design habits is required, more real practise samples from various contexts are needed. The computational generated idealistic models could also be used under the guidance of the design of computational experiments (DoCE). Through diluting the building distribution characteristics with various source samples, more generic relationships between environmental performance and the building distribution patterns would be achieved. However, regarding the application as guidance for future design, a consideration of local design preferences could make it more adaptive.

Here, by discovering and expounding local design habits, designers and planners will be more aware of the potential influential factors in design and could therefore avoid rigid customary constraints in design for new possibilities of innovative, optimised distributions.

As the requirement of independent parameters is applied to multiple linear regression and multi-objective optimisation, the selection of representative parameters is operated. In consideration of the physical meaning of parameters, each representative is selected from each cluster. The principles of selection are: 1. Conventional parameters have top priority; 2. Parameters with physical meanings which are easier to comprehend and to acquire at the early design stage have second level priority.

Therefore, according to the clustering results, the selected key parameters are: TotalLandArea (site scale), ResiBuildingArea (built amount), TriangleSD and PlotRatio (existence of high-rises), BuildingDensity and AverageFaçadeRatio (density), SVF (openness), AvgDistance (set back), ShapeFactor (site shape), AvgLowFaçadeRatio (low and long façade amount), AvgMinFaçadeLength (short façade amount), AvgIntervalArea (large interval amount), AvgCornerLow (small interval amount), AvgIntervalDepthHigh (interval depth amount).

The key parameters would be adapted in regression and optimization of acoustic, sunlight and thermal performances across the site against building distribution. This pioneer study of distributional parameter is essential for next step regression and optimisation.

3.4.2   Clustering for Front Façade Distribution Parameters

Similar to the grouping for all distribution parameters, this work is also carried out only for front façade parameters. Front façade condition would be used in the assessment of the traffic noise barrier effect in chapter 4. For the summary of the grouping, see table 3.3.

The connotation of the group shows that the aspects describing front façade conditions are large interval, small interval, interval depth, façade height, set back

distance, scale of outline, existence of low and long façade, and existence of short façade.

The key parameters could be selected from each group as: AvgIntervalArea (Large Interval), AvgCornerLow (Small Interval), AvgIntervalDepthLow, AngIntervalDepthHigh (Interval Depth), AvgFaçadeHeight(Building Height), AvgDistance (Set Back), AvgOutlineLength, AvgTotalFaçadeLength (Scale), AvgFaçadeRatio (Existence of Low Façade), and AvgMinFaçadeLength (Short Façade).

Table 3.2 Groups of All Distribution Parameters and Key Parameters in Each Group

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12
Connotation	**Site Scale**	**Built Amount**	**Existence of high-rise**	**Density**	**Openess**	**Set back**	**Site shape**	**low and long façade Amount**	**short façade Amount**	**large interval Amount**	**small interval Amount**	**interval depth Amount**
Parameter name	ResiBuildFootprintArea	OvergroundBuiltArea	AvgFrontStorey	BuildingDensity	SVF	AvgDistance	ShapeFactor	AvgLowFaçadeLength	AvgMinFaçadeLength	AvgCornerArea	AvgCornerLow	AvgIntervalDepthLow
	Resicircumference	ResiBuildingArea	AvgFrontHeight	AvgFaçadeRatio				AvgLowFaçadeRatio		TotalCornerArea		AvgIntervalDepthHigh
	FootprintArea	ResiSuperfacialArea	TriangleSD					AvgMaxFaçadeLength		AvgCornerHigh		
	AvgOutlineLength	HighriseResiArea	PlotRatio					AvgFaçadeLength		AvgIntervalArea		
	DiagonalLength		AspectRatio							AvgMaxIntervalLength		
	TotalLandArea		AvgResiStorey									
	AvgTotalFaçadeLength		HighriseRatio									
	Low/medium-riseResiArea											

Table 3.3 Grouping Of Front Façade Distribution Parameters

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Conno tation	**Large Interval**	**Small Interval**	**Interval Depth**	**Building Height**	**Set Back**	**Scale**	**Existence of Low Façade**	**Short Façade**
Parame ter Name	Avg Corner Area	Avg Corner Low	Avg Interval Depth Low	Avg Façade Storey	Avg Distance	Avg Outline Length	Avg Low Façade Length	Avg Min Façade Length
	Total Corner Area		Avg Interval Depth High	Avg Façade Height		Avg Total Façade Length	Avg Low Façade Ratio	
	Avg Corner High						Avg Max Façade Length	
	Avg Interval Area						Avg Façade Length	
	Avg Max Interval Length						Avg Façade Ratio	

The key parameters would be adapted in regression of front façade acoustic performance to building distribution. After several attempts, the global optimisation does not include the front façade acoustic performance for simpler calculation structure. Hence, the multiple linear regression based on the selected key parameters of the front façade distribution is organised in Appendix G, instead of in the main body of chapter 4.

As indicated in the table, cluster 1 and cluster 7 contain several related parameters, which suggests that many parameters are under the impact of large interval factor and low-rise façade factor.

A group of large interval (cluster 1) consists of parameters concerning large corner opening and large interval opening. By the definition of intervals, the large corner opening makes a great contribution in the amount and size of intervals.

The group of low and long façade (cluster 7) consists of parameters referring to low

façade and long façade. Due to the design habit, the low-rise commercial building is usually situated at the boundary of a residential ward. It tends to be arranged longer to enclose the boundary. Therefore, in many cases, the front façade with a large or even a maximum individual length is always the low-rise façade.

For a further expansion of front façade parameters as a suggestion for future work, a group focussed on the homogeneity of intervals could be added. This group should contain parameters, i.e., façade length standard deviation, interval length standard deviation, and number of intervals to the size of the outline. This group may help to distinguish the differences between small-separate and large-concentrated openings on acoustic and wind environments.

3.5 Level Grading for Key Morphology Parameters of Building Distribution

Since the key parameters of building distribution are selected based on variable clustering results and application convenience for the designer, it is necessary to comprehend those parameters on a deeper level. Because further regression and optimisation may require sample site analysis based on distribution characteristics, studies on the range, grading and conformation proportion in the grading of each parameter would be supportive for potential categorising requirements in further research. Furthermore, the proportion between grades would support the expansion of the sample size for a sparse sample section among the whole population, which is the preparation work for optimisation.
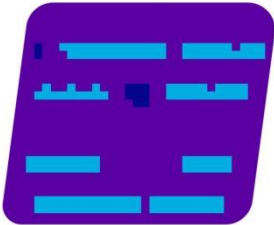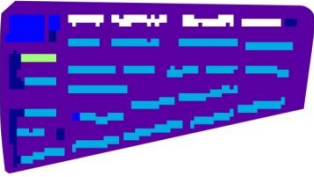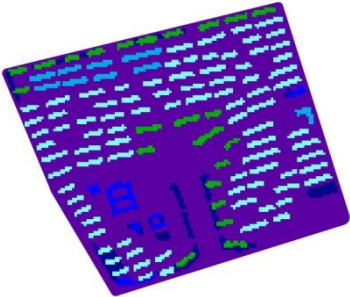
In this section, gradings of nine key parameters are discussed, regarding the threshold determination for grades and proportion between grades. The nine parameters are total land area (site scale), average residential storey, triangleSD and plot ratio (existence of high-rises), building density and average façade ratio (density), SVF (openness), average distance to road and maximum distance to road (set back), and shape factor (site shape).

3.5.1 Grading of Site Size factor: Total Land Area

Regarding to site size, the cases are divided into three sections based on clustering by cases on total land area: small site (29,730-72,000 m^2), medium site (72,000-154,000 m^2) and large site (154,000-419,580 m^2). Among the total number of cases, small sites account for 13.6%, medium site for 63.6% and large site for 22.7%. For a summary of the section see table 3.4.

Compared with grading by other parameters, the intersection shows that small sites tend to have high building density and low plot ratio with even distribution, particularly small TSD (table 3.6). This correlation will be discussed in detail in Section 3.6.

Table 3.4 TLA Grade Summary

Grade	Proportion	Building height maps	Representative Case
Small site 29,730-72,000 m^2	13.6%		S31 TLA: 29,730 m^2
Medium size 72,000-154,000 m^2	63.6%		S1101 TLA: 100,009 m^2
Large size 154,000-419,580 m^2	22.7%		S09 TLA: 419,580 m^2

3.5.2 Grading of High-Rise Factor: Average Residential Storey
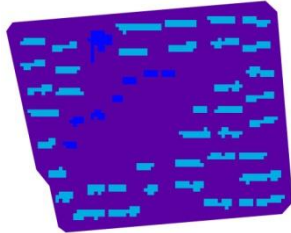
As noted in Chinese Residential Design Specification, GB50096-2011, 1-3F is defined as low-rise residential buildings, 4-6F is defined as multi-storey buildings,
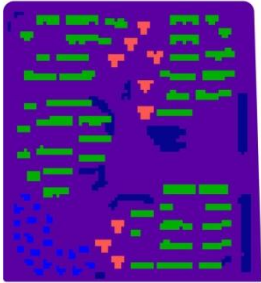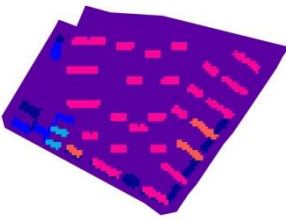
7-9F is medium/high-rise, and over 10F is defined as high-rise residential buildings. With reference to this and based on the case clustering by average residential building storey, the sites are divided to three sections: of low/medium-rise (4-8F), of medium/high-rise (8-18F), and of high-rise (over 18F). The proportions of each grading are 31.8% for low/medium-rise, 31.8% for medium/high-rise (8-18F) and 36.4% for high-rise (over 18F). For a summary of this grading sees in table 3.5.

A relationship between average storey and other distribution parameters shows in the grade intersection in table 4.3.10. In the high-rise section, 68.7% of cases tend to have high plot ratio, 87.5% of cases have low building density and all cases appear clustered based on high TSD value. All cases in the low/medium-rise section tend to have low plot ratio, 85.7% of cases have medium/high building density and 78.6% of cases appear in even distribution by low TSD value.

This shows potential design rules or habits related to building height and distribution patterns. The organisation of building distribution is highly influenced by the configuration of buildings in various heights. Whether it is possible to discover novel distribution pattern other than the present ones, affected by the underlining design rules between distributions and building height configuration, is one of the objectives of this research.

Table 3.5 aResiStorey grade summary

Grade	Proportion	Building height maps	Representative Case
Low/medium-rise 4-8F	31.8%		S0301 aResiStorey: 4.91

Medium/high-rise 8-18F	31.8%		S10 aResiStorey: 12.31
High-rise Over 18F	36.4%		S08 aResiStorey: 35.45

High-rise ratio (over 10F) which is comparable to storey grading result is added as an innovative geometrical parameter. The sites are divided into 3 sections based on high-rise ratio: low (0-0.5), medium (0.5-0.85) and high (0.85-1). This shows that there is positive correlation of high-rise ratio with average residential building storey, high-rise residential building area, plot ratio and TSD, and negative correlation with building density and average façade ratio, while 75% (3 out of 4 cases) of sites with highest TSD are with high-rise ratio of 1. Also, sites located at the low end of ranking by high-rise ratio tend to appear more in a square shape judged by small shape factor value close to 1.

3.5.3   Grading of High-Rise Factor: TriangleSD

TSD is the descriptor of to what extent residential buildings are close to each other. Only the separations between residential buildings are considered, for example building separation distance and in-site open space. The on-edge open space shows no impact on TSD as it is not between buildings. This makes TSD a presenter of the inside distribution of buildings, independent from the front façade conditions and whole site conditions. This characteristic is especially useful for distinguishing sites with large variations on the distribution of open space in-site and at site surroundings but with similar building density.
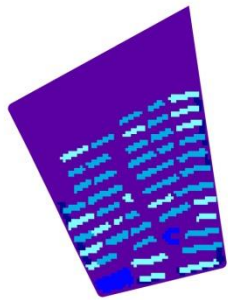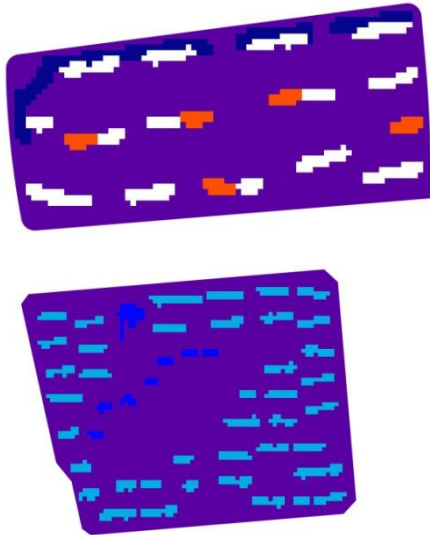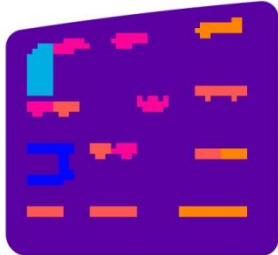
Based on the clustering of cases by triangleSD as calculation parameters, three sections are divided: even (63-390), mix (390-903), and clustered (903-1184). The proportions of each grade are 36.4% for even, 54.5% for mixed, and 9.1% for clustered. It is clear that at the SU-ZHE-WAN region, the pure clustered sites are few; most of them are mixed sites of even and clustered. The summary of the grade is shown in table 3.6.

As can be seen from the grading intersection, the even section consists of low/medium-rise buildings. The distribution pattern is that multiple low/medium-rise units form long building rows (name long façades) by sharing walls, and the long building rows separate evenly from each other in a matrix. In this even grade, there exists many long façades within and on the edge of the site.

The clustered grade consists of high-rise buildings of short façade. The distribution pattern appears as: 1. Some parts of high-rise units are independent to others, while some parts of high-rise units share walls by two, but still form a relatively short façade; 2. There are clear building convergences and the convergences are separated from each other.

In the mixed grade, it is necessary to separate the condition for high-rise and low/medium-rise buildings. For sites mainly consisting of high-rise buildings, the appearance of distribution is 1. High-rises share walls by two or three; 2. The separation between building rows are close to even. For sites mainly consisting of low/mid-rise buildings, the appearance is 1. Parts of the low/mid-rise unit are independent while parts share walls but still form short façades; 2. Buildings slightly cluster forming convergences, and these convergences are set apart from each other.
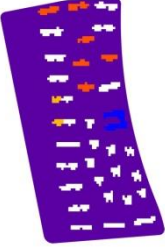
Table 3.6 TriangleSD Grade Summary
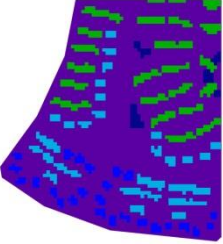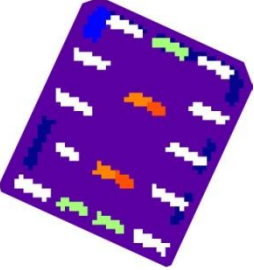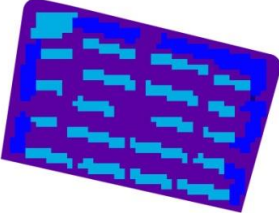
Grade	Proportion	Building height maps	Representative Case
even 63-390	36.4%		S1802 TSD:122
mix 390-903	54.5%	 	S260402 TSD:829 Of mid/high-rise buildings S0301 TSD:768 Of low-rise buildings
clustered 903-1184	9.1%		S1902 TSD:1054

Compared to grades in other distribution parameters, a tendency for the evenness of buildings under the influences of other distribution parameters appears, by which we can uncover the rules of design applied in practise, locally formed under complex constraints of economic and social reasons.

The high-density sites all appear even and located in an even grade. Three out of four cases in the clustered grade are with a high plot ratio. All other sites with high plot

ratio are located in the mixed grade, and none are located in the even grade. From this, we can conclude that evenness tends to be a characteristic of low/mid-rise building distribution, while clustered appearance tends to be attributable to high-rise buildings.

This could be explained by the attempts to maximise the total floor area for economic reasons when building height range is determined, with consideration of compensation due to sunlight and daylight requirements. Because high-rise buildings project longer and larger shadows, their distribution could not be even and matrix like, but more random and clustered.
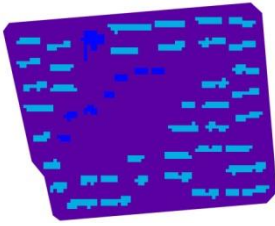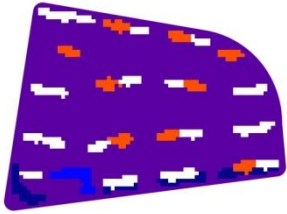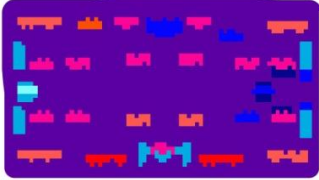
One of the aims of optimisation in the next stage is to define the good combination of high and low/mid-rises in a proper distribution. As low/mid-rises tend to set as even, how to individually arrange high-rises around low/mid clusters and how to mix low/mid clusters into high-rise clusters is of significance for further study. It is also possible to break design habits about evenness to create innovation organisation.

3.5.4   Grading of High-Rise Factor: Plot Ratio

By case clustering based on plot ratio, the cases are divided into three grades: low PR (0.66-1.86), medium PR (1.86-2.7), and high PR (2.7-3.55). The proportions of each grade are 70.5% for low PR, 25% for medium PR, and 4.5% for high PR. For a summary, see table 3.7.

It is clear from the table that the majority of the sites are of a low to medium plot ratio. This is in accordance with the population density in the SU-ZHE-WAN region. High density and super density communities are still not necessary for this region and are not preferred by the market.

Table 3.7 Plot Ratio Grade Summary
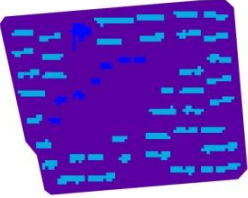
Grade	Proportion	Building height maps	Representative Case
Low PR 0.66-1.86	70.5%		S0301 PR:0.66
Medium PR 1.86-2.7	25%		S260401 PR:2.43
High PR 2.7-3.55	4.5%		S09 PR:3.03

3.5.5   Grading of Density Factor: Building Density

Through the clustering of cases by building density as a calculation parameter, three grades are divided: low density of 0-0.17, medium density of 0.17-0.23 and high density of 0.23-0.32. The threshold values of 0.17 and 0.23 are 53.12% and 71.9% of the maximum value of 0.32, respectively. The proportions of each grade are 45.5% for low density, 38.6% for medium density, and 15.9% for high density. The grade is summarised in table 3.8.

The intersection with other gradings shows that high density sites mainly consist of low/medium-rise buildings; high-rise buildings are rarely seen in this grade. Similarly in the medium-density grade, only a few high-rise buildings exist in those sites. However, the low density grade could freely consist of sites with low, medium and high-rise buildings.

Table 3.8 Building Density Grade Summary

Grade	Proportion	Building height maps	Representative Case
Low density 0-0.17	45.5%		S0301 BD:0.138 Of low-rise buildings
			S260201 BD:0.097 Of medium/high-rise buildings
Medium density 0.17-0.23	38.6%		S1501 BD:0.195 Of low-rise buildings
			S0201 BD:0.188 Of medium-rise buildings
High density 0.23-0.32	15.9%		S1102 BD:0.320 Of low-rise buildings

3.5.6   Grading of Density Factor: Average Façade Ratio
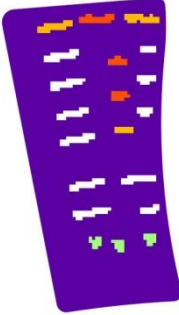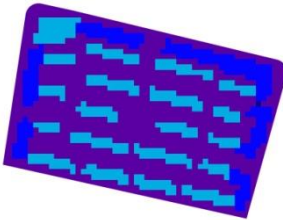
Average façade ratio is listed in the cluster of density factors in the previous section. It appears in the sites where a dense site tends to have a high level of average façade ratio.

Based on case clustering by average façade ratio, two sections are divided: low level (42-65%) and high level (65-90.36%). The proportions are 68.2% for low level and 31.8% for high level. For a summary of the grading of the average façade ratio, see table 3.9.

Compared by intersections with other grading, sites with high building density all appear to have a high average façade ratio. Furthermore, the sites with highest building density also overlap those with the highest façade ratio. It could be suggested that high density sites tend to have enclosed boundaries.

Small sites also show a tendency of enclosure at boundary, and 66.7% (4 out of 6 cases) of small sites have the highest average façade ratio. This also corresponds with the correlation between small sites and high-density sites.

Table 3.9 aFaçadeRatio Grade Summary
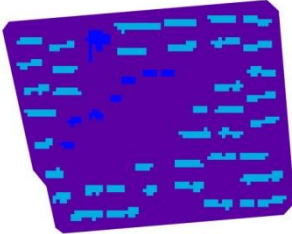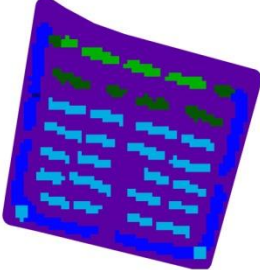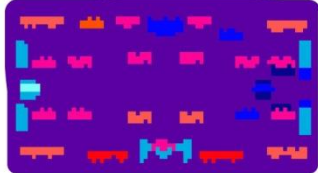
Grade	Proportion	Building height maps	Representative Case
Low (42%-65%)	68.2%		S260101 aFaçadeRatio : 42.5%
High (65%-91%)	31.8%		S1102 aFaçadeRatio: 90.36%

3.5.7   Grading of Openness Factor: SVF

Based on case clustering by SVF, three grades can be divided: high SVF (68-76.4), low SVF (52-68) and poor SVF (50-52). The proportions are 22.7% for high grade, 72.7% for low grade, and 4.5% for poor grade. For a summary, see table 3.10.

Compared to other parameters, high accordance between SVF and plot ratio appears. All high plot ratio sites are located in the poor grade of SVF. Almost all medium plot ratio sites are located in the low SVF section, except for S28 which has one skyscraper that boosts its plot ratio. Most cases of low plot ratio show high SVF.

Table 3.10 SVF Grade Summary

Grade	Proportion	Building height maps	Representative Case
High SVF 68-76.5	22.7%		S0301 SVF:76.46
Low SVF 52-68	72.7%		S1101 SVF:60.82
Poor SVF 50-52	4.5%		S09 SVF:50
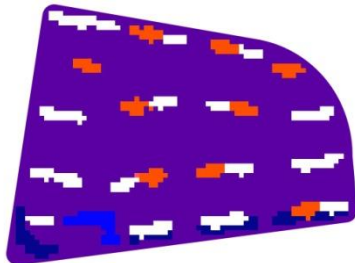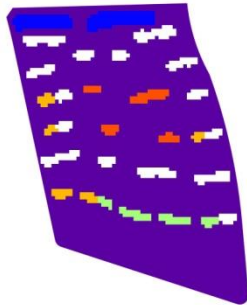
3.5.8 Grading of Set Back Factor: Dmean

Dmean is the most independent parameter among all distribution parameters significantly correlated with Dmax (0.828) and Dmin (0.572). Dmean has slightly correlated to building density at -0.429. This means that on average, sites with high density may not allow too much set back distances.
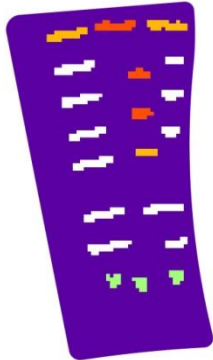
Based on case clustering by Dmean, three sections are divided: small (below 29m), medium (30-46m), and large (46-61m). The proportions are 47.7% for small grade, 42.0% for medium grade, and 11.4% for large grade. For a summary of the grading, see table 3.11.

Dmean barely shows relationships with any other distribution parameters. It appears that 50% (4 out of 8 cases) of large size sites have large Dmean, and 80% (4 out of 5 cases) of sites with large Dmean are of a large size. It could be suggested that larger sites tend to be generous in setting back front row buildings due to, 1. The requirement of avoiding traffic noise from the wider artery street; 2. The requirement to set pedestrians apart from traffic in front of retail stores which is more suitable to accommodate along longer outlines.

Table 3.11 Dmean Grade Summary

Grade	Proportion	Building height maps	Representative Case
Small (below 29m)	47.7%		S0260401 Dmean:13.96m
Medium (30-46m)	41.9%		S260502 Dmean: 35.40m
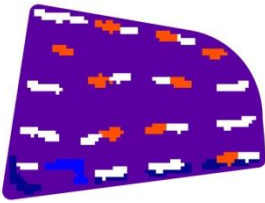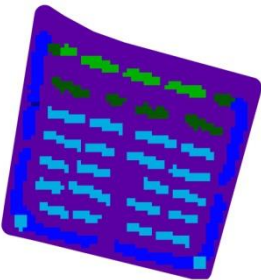
Large (46-61m)	11.4%		S260101 Dmean: 60.51m
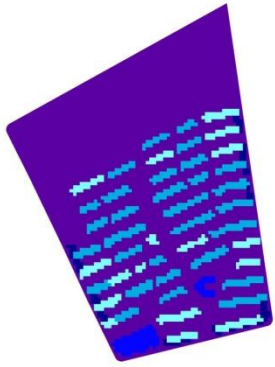
3.5.9   Grading of Set Back Factor: Dmax

Sites are divided into three grades based on case clustering by Dmax, small (14-36m), medium (36-65m) and large (65-163m). The proportions are 36.4% for small grade, 34.1% for medium grade and 29.5% for large grade. For a summary, see table 3.12.

Dmax is related to Dmean, but not related to other distribution parameters. The essence of Dmax is the set-back distance of residential buildings from highest traffic load, railway or nature body. It should be related to conditions outside of the site, which are not optimisable in the research.

Table 3.12 Dmax Grade Summary

Grade	Proportion	Building height maps	Representative Case
Small (14-36m)	36.4%		S0260401 Dmax:14.00m
Medium (36-65m)	34.1%		S1101 Dmax: 51.00m

Large (65-163m)	29.5%		S1802 Dmax::162.27m
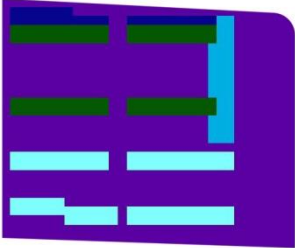
3.5.10   Grading of Site Shape Factor: Shape Factor
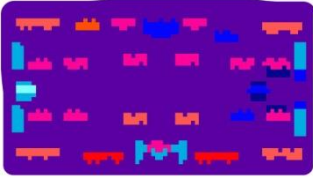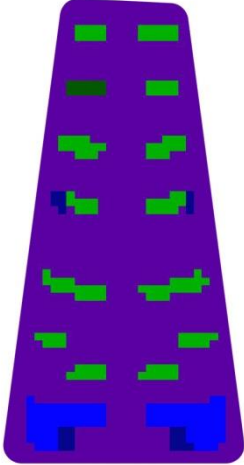
Shape factor is calculated as the diagonal length divided by the shortest outline length of the same site. The larger the value, the narrower and longer is the shape of the site.

The cases are divided into four grades based on case clustering by shape factor: square (1.5-2.0), rectangle (2.0-2.4), narrow (2.4-4.47). The proportions are 47.7% for square, 22.7% for rectangle and 29.5% for narrow. The summary is shown in table 3.13.

For the narrowest sites, SVFs tend to be large and more clustered as judged by the corresponding TSDs. Shape factor is not in accordance with site size and high-rise building area.

Table 3.13 ShapeFactor Grade Summary

Grade	Proportion	Building height maps	Representative Case
square (1.5-2.0)	47.7%		S12 ShapeFactor: 1.51

rectangle (2.0-2.4)	22.7%		S09 ShapeFactor: 2.28
narrow (2.4-4.47)	29.5%		S29 ShapeFactor: 4.47

3.5.11 Other Gradings and Discussions

Connected to average façade ratio, average max façade length is also analysed. Sites could be divided into three grades based on case clustering by max façade length: short (31-65m), medium (65-94m) and long (94-112.5m). In the high façade ration grade, 93% (14 out of 15 cases) of the sites are located in the medium/long grade of max façade length. This means the consistency between high façade ratio and long max façade length.

Average front façade storey also shows connections with the front façade ratio. The low storey grade of the average front façade storey shows connections with small site, high building density and low average storey by the intersection of gradings. This verifies that the smaller, denser the site, the more enclosed it tends to be. It appears that dense and enclosed sites with a low/medium-rise building is a fixed design strategy for the SU-ZHE-WAN region.

The clustering by average interval area shows that sites with an interval area over 1000m^2 covers the majority of 88.9% (16 out of 18 cases) of sites with a high grade of medium to high-rise building area and of an average residential building storey.

3.5.12 Summary

This section analysed the grading of sites according to the characteristics of each key parameters from each distribution parameter group extracted in Section 4.2. The summary of the grading of key parameters is shown in table 3.14.

Table 3.14 Summary of Grading by Each Key Parameter

Cluster	Parameter	Grade and Proportion				
Site Size	TLA	Small site: 29,730-72,000 m^2	Medium size 72,000-154,000 m			Large size 154,000-419,580 m^2
High-rise	aResiStorey	Low/medium-rise 4-8F		Medium/high-rise 8-18F		High-rise Over 18F
	TSD	even 63-390		mix 390-903		clustered 903-1184
	PR	Low PR 0.66-1.86			Medium PR 1.86-2.7	High 2.7-3.55
Density	BD	Low density 0-0.17		Medium density 0.17-0.23		High density 0.23-0.32
	aFaçadeRatio	Low 42%-65%			High 65%-91%	
Openness	SVF	High SVF 68-76.5	Low SVF 52-68			Poor 50-52
Set back	Dmean	Small below 29m			Medium 30-46m	Large 46-61 m
	Dmax	Small 14-36m		Medium 36-65m	Large 65-163m	
Site Shape	Shapefactor	Square 1.5-2.0			Rectangle 2.0-2.4	Narrow 2.4-4.47

The proportion of grades by each parameter shows the unevenness of sample data. This may be due to local design habits as samples are randomly selected from SU-ZHE-WAN region live projects. To eliminate the bias of sampling for further optimisation, it is suggested to expand the sample size in grades which have low proportions for each key parameter. By comparing the intersections of sites in different grades of various parameters, some correlation or coexistence appears between parameters. This will be analysed and discussed in detail in Section 3.6.

In the following qualitative analysis, the description terms regarding to level values would refer to this grading for each building morphology parameter.

3.6 Correlation and Interaction between Key Morphology Parameter Pairs

The variable clustering shows that some distribution parameters with different description natures are grouped into the same cluster. The grades by one key parameter may also overlay the grading of another parameter. Therefore, further multiple correlation analyses (one tailed) are carried out for all building distribution parameters. Through comparing correlation coefficients, interactions of parameter pairs are discovered. Except the parameters clustered into the same group in Section 4.2, some parameters which are not expected to be connected show binding relationships, i.e., innovative parameter to conventional parameter.

This may be due to a two-stage reason. Firstly, limited by economic and public preferences, the planners in the SU-ZHE-WAN region have formed certain design habits when arranging residential wards; secondly, including development intensity and density, the conventional design parameters are loosely constraining building distribution.

It is worth noting that some innovative parameters extracted from the sample set have already been biased by conventional design habits. Therefore, the novel distribution

parameters would have a high correlation with conventional design parameters, even though they are expected to be independent parameters.

The pairs of distribution parameters are discovered through the comparison in the multiple correlation matrix of parameters. For convenient reasons, the large and complex correlation matrix is not shown in this section. The pairs showing clear correlations by their definitions are not listed here. Only the pairs indicating unexpected connections are discussed.

The pairs appear from three aspects: impact of high-rises, impact of low/medium-rises, and boundary phenomenon. The pairs are, 1. impact of high-rises: high-rise building factor and plot ratio, plot ratio and SVF, high-rise building factor and triangleSD; 2. impact of low/medium-rises: low-rise building factor and total land area, low-rise building factor and building density; building density and TSD 3. boundary phenomenon: high-rise building factor and interval length, TSD and Interval length.

3.6.1   Pair of High-rise Impact: High-Rise Building Factor and Plot Ratio

The high-rise building area shows a clear influence power on the residential building area by a correlation coefficient of 0.852 and on an average residential storey by 0.818. As parameters presenting high-rise existence in cluster 3 in Section 3.5, these two parameters both show strong correlation with other parameters under the impact of high-rise buildings.

Plot ratio is strongly influenced by the existence of a high-rise buildings area at the level of 0.808. It also correlated to an average residential storey at 0.87, and residential building area at 0.615. It is obvious that the more high-rises in a site, the higher the intensity of the community with a higher plot ratio will be.

3.6.2   Pair of High-rise Impact: Plot Ratio and SVF

Plot ratio shows a high-level influence on SVF at a correlation coefficient of -0.797, specifically that a high plot ratio site would tend to have poor sky view performance, and this will negatively impact daylight and sunlight conditions both outdoors and indoors. Similarly, SVF is correlated with a high-rise building area at a level of -0.604, which is not as high as with plot ratio, but shows the impact of high-rises.

The connection between high plot ratio and poor SVF is concurrent with the fact that high plot ratio planning needs more adjustment on building distribution for satisfactory daylight and sunlight performances.

3.6.3   Pair of High-rise Impact: High-Rise Building Factor and TriangleSD

Parameters regarding high-rises all show correlations with triangleSD, which describes in-site building distribution and should not link with high-rises in expectation. TriangleSD is correlated with average residential storeys at a level of 0.753, a high-rise building area at 0.577 and a plot ratio at level of 0.542.

These links show that sites with more high-rise buildings tend to be arranged in clusters rather than evenly distributed, by having larger value of triangleSD. This may be due to the maximisation of economical repay. Application of cluster distribution on low/medium-rises is actually decreasing development intensity without the necessity for doing so. However, for high-rises, larger separations for daylight and sunlight requirements are necessary so that the amendment of distribution patterns into clusters does not cause a deduction in repay, but improves the outdoor environment by which increasing the additional value of real estate products

As seen from three pairs of high-rise impact, conventional parameters about the existence of high-rise buildings are correlated to conventional parameter-plot ratios and innovative parameters-SVF and TSD. The close relationship between plot ratio and high-rise parameters has made plot ratio commonly used as a representative parameter describing the number of high-rise buildings in current residential design systems, instead of parameters directly describing the number of high-rises, i.e., average storey or high-rise building area.

Similarly, it could be expected that SVF and TSD could also supersede high-rise parameters in next stage regression and optimisation. Particularly, SVF and TSD present cluster openness and cluster evenness (however, they are dominated by the high-rise cluster) respectively mentioned in Section 3.5. This also shows that parameters of openness and evenness have the potential to describe the condition of the high-rises.

Changing the descriptive system of building distribution from conventional categories of the attributes of residential wards to new categories of other unfamiliar aspects, i.e., openness and evenness, could provide a different emphasis on distribution for designers. The new angle enables designers to develop new design approaches for optimised organisations rather than following conventional habits.

3.6.4   Low-rise Impact: Low-Rise Building Factor and Total Land Area

Total land area is correlated with the low/medium-rise building area by 0.65, and it is not significantly correlated to the high-rise area. This reflects that the large size site tends to be filled by low/medium-rise buildings rather than high-rise buildings.

This could be explained by the current population density in the SU-ZHE-WAN region. Compared to super-dense cities i.e., Hong Kong and Tokyo, the price of land still allows the existence of low/medium-rise community with a dispersion of

high-rise buildings, rather than a full high-rise community, since the public prefers low density and low intensity communities. Therefore, when a site is large enough to accommodate low/medium-rises at enough intensity for economic repay, communities with more low/medium-rises are still much more preferred by the market and the developer.

Because low/medium-rises result in a larger footprint area than high-rises, and an inner connection between low/medium-rise and total land area, residential footprint area correlates to total land area at the very high level of 0.902. The correlation level of low/medium-rise buildings and residential footprint area is 0.844. The predicting power of the residential footprint area in describing distribution is covered by site scale factor.

3.6.5   Low-rise Impact: Average Residential Storey and Building Density

Building density is calculated as building footprint area divided by total land area. However, building density only shows a correlation level at 0.284 with footprint area and at -0.105 with total land area. This means by fixed building density, the limitation on footprint area and total land area is very loose. In other words, the variation range of the design constrained by building density to the planner is quite large. The parameter with attributes of a rigid limit on development intensity but of loose constraint on distribution pattern is most convenient for the planners and makes the project design-friendly.

It shows that building density is correlated to low-medium-rise building areas at level of 0.403, to a high-rise building area at -0.498, and to an average residential building storey at -0.616. Namely, sites consisting of more low/medium-rises tend to have higher building density. This is obvious because low/medium-rises produce more footprint area which is the numerator in the calculation of building density. Another

reason is that low/medium-rises require a small separation distance for daylight and sunlight requirements which in fact allows high density.

3.6.6   Low-rise Impact: Building Density and TSD

Comparatively, there are unexpected coexisting relationships between building density and TSD. Building density has a correlation coefficient of -0.707 with TSD. In other words, the denser sites tend to be more evenly distributed rather than clustered.

This could be understood as the result of economic consideration. The denser sites do not set environmental comfort as priority, therefore clustered distribution is not necessary which will sacrifice building number and density for extra indoor and outdoor comfort.

As seen from the three pairs related to low/medium-rises, total land area, building density and TSD could be used to describe distribution attributes regarding the existence of low/medium-rises. Conventionally, high building density is considered a sign of high low/medium-rise proportions. Similarly, large total land area and low TSD could also be applied as signs.

3.6.7   Boundary Phenomenon: TSD and Interval length

Except for the relationship with general site parameters, TSD also shows connections with front interval conditions. TSD correlates with average interval length at a level of 0.602 and a maximum interval length at 0.463. These results mean that the majority of clustered distributed sites could not be in a high level of enclosure at the boundary; otherwise, it is not possible to allow the existence of a large interval length.

Relevant to front conditions, minimum distance to road and minimum corner area are also discussed. The correlation between minimum distance to road and TSD is 0.422. The correlation between minimum corner opening area and TSD is at 0.437. Namely, the more clustered distributed sites have more set back distance from the site boundary and have more openings at the corners on the minimum level. It means a clustered site has a higher allowance of set back distance and corner openings on all of its edges and corners.

3.6.8   Boundary Phenomenon: Building density and front façade conditions

Building density also correlated with average façade storey at -0.72, which means denser sites tend to have lower front buildings. Compared to the coefficient with the average residential building storey at -0.616, building density has a closer relationship with front building storey than in-site building storey. It could be suggested that, in high-density cases, high-rise buildings which tend to have higher influence on average storey are arranged inside the site, rather than set on the boundary.

Building density also shows a relatively close correlation with front façade ratio at 0.722, with average low-rise façade ratio at 0.583 and with average façade length at 0.524. It could be suggested that sites with higher density tend to have a higher ratio of front façade, especially low-rise and longer façade. In other words, sites of higher density tend to be enclosed at boundary, even wrapped up by low-rise commercial buildings.

Indicated by the pairs referring to boundary conditions, clustered sites tend to have large intervals, high density sites tend to have lower front building stories, and high-density sites tend to have enclosed boundaries and small set back distance to roads.

3.6.9   Other Pairs between Un-key Parameters

Some parameters are not key parameters selected from categories of distribution attributes. However, their coexistence pairs of distribution parameters are still worth analysis.

The overground building area correlates with the minimum distance to road at the level of 0.501. Except for residential buildings, overground buildings include kindergartens, churches, service buildings and commercial mixed-use office buildings if there are any within the site. Among these, commercial and office buildings are usually located at the boundary of the site with set back distances to the road. The other building types are usually located inside the site. The correlated relationship is in accordance with the fact that if the number of commercial and office buildings (account for largest proportion in overground building except for residential) is large, the requirement for a larger distance in front of the shops at the site boundary rises, namely a larger set back distance.

Dmean is the most independent parameter among all distribution parameters because it only significantly correlates with Dmax (0.828) and Dmin (0.572). Dmean has a slightly correlation to building density at -0.429, which means sites with high density may not allow too much set back distance on average.

Shapefactor is a very independent parameter that it only correlated to minimum outline length at -0.546. Shape factor is calculated as the diagonal length divided by minimum outline length.

Residential footprint area is under a stronger influence from a low/medium rise building area with a correlation coefficient of 0.844, rather than the high-rise building area. Similarly, the residential superficial area is under a strong influence from high-rise buildings with a coefficient of 0.853, rather than low/medium-rise buildings.

This is also in accordance with the superficial area which is correlated to average residential storeys at a level of 0.666 and to residential footprint area at 0.422; therefore, the superficial area is more controlled by the building storey.

Other relations of the physical meaning of the parameters that should be related are not analysed in detail. These pairs are listed below. Average interval length is correlated to average interval area and average/total corner area. Average façade ratio relates to average/maximum façade length, average/total low-rise façade length and low-rise façade ratio. Front façade storey is related to residential building storey, and other parameters related to the existence of high-rise buildings. Average façade ratio is negatively correlated to the total interval area at -0.717. Multiple correlation relationships are all significant between total land area, overground building area, residential building area, average outline length, footprint area, residential footprint area, and low/medium-rise residential building area. This corresponds with cluster 1 of all parameters under the influence of site size.

3.6.10   Summary

To summarise, building density is a parameter flexibly constraining the planner with a rigid value of development intensity. It also shares coexisting phenomenon with TSD, avg façade storey, and front façade ratio, due to the design habit in the SU-ZHE-WAN region.

In this chapter, a total of 31 parameters are analysed and compared. Among all the possible significant correlation pairs between them, only 8 pairs are selected for study, because they are not naturally correlated by definitions. The 8 pairs represent three aspects: high-rise impact, low/medium-rise impact and boundary conditions.

The analysis of the selected 8 pairs uncovers hidden rules and design habits of local residential ward planners. The parameter pairs and the rules are listed in detail in table 3.15.

The rules extracted reflect regulation, and a mix of design habits. Therefore, these rules are the description of current design approaches. They could be applied as references in future design, which means the design is following the same track of design habits. If innovative distribution patterns are applied in future projects, a break and recombination of building distribution parameters to encourage the abandonment of fixed habits is essential.

Table 3.15 Summary of Coexistence Pairs and Design Rules

Category	Parameter1	Parameter2	Rules Interpretation
High-rise impact	High-Rise Building Area, Average Residential Storey	Plot Ratio	more high-rises in a site ←→ higher intensity of the community (higher plot ratio)
	Plot Ratio	SVF	high plot ratio site ←→ poor sky view performance
	Average Residential Storey	TSD	more high-rise buildings ←→ arranged in clusters than evenly distributed
Low-rise impact	Low/Medium-Rise Building Area	Total Land Area	large size site ←→more filled by low/medium-rise buildings than high-rise buildings.
	Average Residential Storey	Building Density	more low/medium-rises ←→ higher building density
	Building Density	TSD	denser sites ←→more evenly distributed rather than clustered
Boundary phenomenon	TSD	Interval Length	clustered distributed sites ←→ less enclosed at boundary
	Building Density	Average Façade Storey,	higher building density ←→ lower front façade storey
	Building Density	Front Façade Ratio	higher building density ←→ higher front façade ratio (more enclosed)

It is suggested for future work to adapt partial correlation when considering connections between building parameters. Partial correlation calculates correlation coefficients, which eliminates impact from parameters, therefore defining the

coefficient solely due to connections between two parameters. Partial correlation could also help in the discovery of unexplained correlation parts between innovative and conventional parameters, so it would be worth studying potential unnoticed innovative parameters.

3.7 Summary

This chapter studies the conformation of distribution parameters and their relationships. Rules of local residential designs are extracted and could be used for further environmental optimisation regarding building distribution.

The building morphology parameters are from three origins: conventional parameters for single buildings, innovative for single buildings and for neighbourhoods. Based on the characteristics of the value of each parameter, all distribution parameters are clustered into 12 groups. Each group indicates one attribution aspect of building distribution.

The implications of each group are, site scale, built amount, existence of high-rises, density, openness, set back, site shape, low and long façade amount, short façade amount, small interval amount, and interval depth amount.

Some aspects are very influential, including several parameters originally of different natures. One or two key parameters from each group are selected for detail studies, based on the principle of prioritising innovative parameters but with consideration of agreement with the convention of the parameter application system.

The selected key parameters are individually studied by their range, grading and intersection with the grading of other parameters. The grading of key parameters will be used in the following chapters' discussions as the basis of level division. The discussions of acoustic, thermal and sunlight performance regarding to building

distribution will be carried out according to groups determined by key parameter grading.

To confirm and articulate the connections shown in parameter grading intersections, correlation between all parameters are operated and special pairs are selected for analysis because they show strong correlated characteristics. Based on studies of parameter pairs, local residential design rules are extracted as in table 4.4.1. The rules concentrate on high-rise, low/medium-rise and boundary impacts.

These rules are indications of local design conventions for residential ward planning, which help designers to understand their own design progress. They also could be used to guide future designs following local habits. However, to acquire innovative design pattern surmounting design habits, further studies of integrated optimisation of reorganised distribution parameters in accordance with environmental performance is essential.

Chapter 4 Sample Simulation in Acoustic Domain and Performance Data Analysis

With an increase in population, more high-density residential quarters were created throughout the world to meet the new requirement. These were built in place of the more local-traditional, low-density residential wards. Residential development intensity is increasing further, sacrificing environment performance in the residential quarters – including the acoustic environment - because of the continuously growing urban traffic load.

Therefore, the aim of this chapter is 1) to define the proper performance indices of assessing noise propagation performance in the residential ward; 2) to discover the qualitative relationship between the level of assessing parameters and building morphological parameters; in other words, the rules of residential ward layout arrangement regarding noise propagation performance; 3) to define the quantitative relationship between assessing parameters and indicators of building distribution.

Noise mapping technique is employed with the software package CadnaA, for the simulation of an acoustic environment in residential wards. Statistical clustering analysis is applied for the purpose of extracting significant representative acoustic performance indices. Multiple linear correlations and regressions are performed to discover the mathematical relationships between building morphology parameters and acoustic performance indices.

In this chapter, Section 4.1 introduces the background of acoustic simulation. Section 4.2 compares the currently available acoustic simulation packages. Section 4.3 states the set up process of acoustic simulation model in CadnaA. Section 4.4 elaborates in detail on the sound source definition in CadnaA. Section 4.5 discusses the selection of representative performance indices for result assessment and analysis. Section 4.6 and

4.7 explain in detail about the qualitative and quantitative analysis of acoustic simulation results. For the flowchart of this chapter see figure 4.1.

Figure 4.1 Content Structure of Chapter 4 Sample Simulation in Acoustic Domain and Data Analysis, The expansion of box 4 in Overall Content Structure
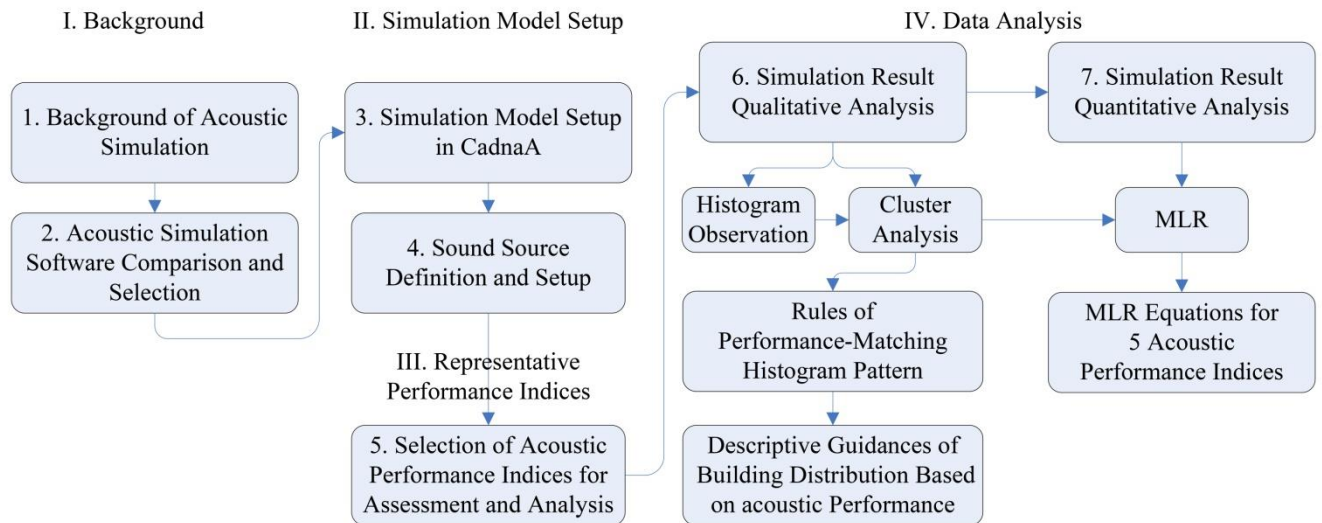
Acronyms for Chapter 4

aCAL	average corner area low value
aD	average distance to road
aFLmin	average min façade length
aFR	average facade ratio
aIAmean	average interval area
aIDH	average interval length high value
aIDL	average interval depth low value
aLFR	average low-rise façade ratio
aOL	average outline length
aTCA	total corner area
BD	building density
CRTN	calculation of road traffic noise
dendrogram	a tree diagram that is used to represent and categorize hierarchical relationships among objects, created as an output from hierarchical clustering
L(A)eq	the A weighted equivalent sound level
L10	a percentile at 10% of all data which suggests that in 10% of the time this value is exceeded during the whole record period and it is an indicator of the average peak value of the sound source.
MLR	multiple linear regression
P10, P20, P30, P40, P50, P60,	ten percentiles of all collected simulation data over one grid

P70, P80, P90

P25, P75	two quatiles of all collected simulation data over one grid
PR	plot ratio
RBA	residential building area
SF	shape factor
SPL	sound pressure level
SPL(A)	A weighted sound pressure level
SPL-IQR/IQR	interquatile range, range between P25 to P75
SPL-P10	10% percentile of SPL grid value of a site
SPL-P40	40% percentile of SPL grid value of a site
SPL-P70	70% percentile of SPL grid value of a site
STI	the speech transmission index
T30	reverberation time under T30 measurement
THR(65)	threshold ratio of grid points met acceptable threshold of less than 65dBA in one acoustic simulation
TLA	total land area
TSD	triangleSD, standard deviation of triangle area

4.1 Background of Acoustic Simulation

In this research, the acoustic simulation section is primarily focused on traffic noise control. Traffic noise control is a mature research field, within which plenty of theory and experiments have been carried out. To control traffic noise, efforts were made on noise control at sound source, sound propagation control and noise abatement methods.

Related to this research, previous works are primarily about simulation, evaluation, prediction and abatement of traffic noise, and they are closely related to residential area standard and regulation compilation. Most studies are focused on particular projects and the discussion of noise abatement methods. The research related to urban structure is mainly located on the soundscape field, which is more concerned with subjective evaluation and perceptive sound quality; therefore, social and psychoacoustic factors will have an influence on the research results. In addition, the abatement methods in previous studies are mostly about technique approaches such as

noise barrier and sound absorber installation. While these are most effective, the abatement methods rarely include building arrangement in primary design process.

The above techniques are mature methods combined with design considerations to amend sound quality. However, this research targets the optimisation of sound quality by architecture design and urban planning, so all of these strategies are not used in the modelling because they can be applied after the building distribution procedure in order to make further improvements. In general, if considering the building arrangements during the design process to improve the outdoor sound environment, the technical methods could be added to enhance local area performance. So, in this research, the technical abatement method will not be used, as it can be applied in the post-design process.

In residential areas, traffic is often the dominating noise source (Skånberg and Öhrström, 2002). Many efforts have been made in previous works of noise control at sound source, sound propagation control and noise abatement methods. However, building distribution morphology is relatively less involved in studies regarding the residential area, as most studies focus on the particular project of discussing noise abatement methods (Kang, 2006). Previous studies on noise exposure have rarely offered a quantitative relationship between noise level and urban structures (Weber, Haase and Franck, 2014). Therefore, there is a need to understand the limit of the conversional noise attenuation method, and search for new approaches to solve the gradually growing problem (Lam and Ma, 2012). It is reported by multiple works that human response is not only affected by the exposure level of the building an individual is in, but also by the surrounding environment (Klæboe, Engelien and Steinnes, 2006; Botteldooren, Dekoninck and Gillis, 2011). Achieving a general healthy acoustic environment of a soundscape is an available approach (Skånberg and Öhrström, 2002).

Some works make attempts on the effects of the built environment pattern on sound environment performances. This work looks for the influence of the physical

environment qualities of a quiet courtyard on residents' response to the noise environment. It is found that the annoyance is related to noise exposure, the quality of the courtyard and the form of the building. At the side of the quiet facade, the noise level is much lower in closed building blocks than in open ones (Gidlöf-Gunnarsson and Öhrström, 2010). Landscape metrics are used for the estimation and prediction of noise based on urban land-use/cover structure (Weber, Haase and Franck, 2014).

It is noted that noisescape characteristics at dwelling scale in residential complexes are strongly correlated with the building design and arrangement, while residences categorised by the noisescape characteristics are also distinguished by their morphological indices (Lam *et al.*, 2013). Lam also mentioned that the ability of noise barrier is limited. Due to high correlation between urban form and noisescape, noise reduction through urban design is possible. Lam mentioned that the noisescape can be managed through urban design, as it is found that noise-screened buildings are less noisy than unshielded ones. In Hong Kong, lower noise exposure exists in recently built residential complexes when compared with older buildings (Lam and Ma, 2012).

Researchers also note that if the inhabitants have access to a quiet facade, the traffic noise annoyance is reduced (Öhrström *et al.*, 2006). This suggests that it is acceptable to sacrifice the frontline facade by locating at a high traffic noise exposure to achieve a quieter back facade by confining the frontline building form. Noise levels at quiet facades would appear lower in closed building blocks than in open blocks (Salomons and Berghauser Pont, 2012).

Some works are more focused on the impact of single building form on sound performance. The distribution of traffic noise on high-rise building sites in Tehran is explored and it is concluded that the highest noise level occurs at the lowest front corner of the side panels closest to the motorway, and the lowest noise level occur at the back edge of the roof (Ranjbar, Gharagozlou and Nejad, 2012). Noise level at a quiet facade of a building will be substantially affected by the form of a building

block (Gidlöf-Gunnarsson and Öhrström, 2010).

This research focuses on sound propagation control through building distribution arrangement in general.

4.2 Acoustic Simulation Software Comparison and Selection

Several acoustic simulation software are available and validated. Generally, acoustic simulation software could be classified into room acoustic simulation tools and outdoor noise mapping tools. Also it could be distinguished by the method of modelling of the ray-tracing method, radiosity (sound energy) method and image-source, etc. Software is usually selected according to the objective of the task, location of the project and scale of the simulation. For large scale simulations, like meso and macro scale simulations with complicated sound source and influential factors, noise mapping is the most appropriate choice. Noise mapping is widely accepted for its cost efficiency in noise exposure assessment with proper accuracy(Lam *et al.*, 2009; Law *et al.*, 2011).

A noise map is used to present the geographical distribution of noise exposure. Its data source can be measured or calculated. It is suitable for larger urban areas compared to the various micro scale simulation techniques. It is especially powerful and effective in visualisation and assessment to sound performance. Nowadays, it is applied in various strategic ways in numerous fields of prediction and management, etc. (Kang, 2006).

The following are acoustic simulation software packages; their simulating range differs from vast urban noise mapping to indoor physical acoustic simulation. From the comparison of the characteristics of the software, one will be selected as the primary simulation tool in this research.

ODEON

This is a tool based on the image-source method combined with ray tracing method, suitable for interior acoustics of buildings, industrial environments and outdoor areas with complicated geometry. The required Input variables are geometry and surface-properties, while outputs will be the acoustics which can be predicted, illustrated and listened to. It is an Industrial edition for environmental acoustics where SPL, SPL(A), T30 and STI are available in simulation. It allows the modelling of point, line and surface sources, with the possibility to model large and complex sound sources. In its Auditorium edition for a large set of rooms, the acoustic parameters used are based on the reverberation curve. It contains various graphical and auralisation tools. The Combined edition includes all the features in the Industrial and Auditorium editions. However, Odeon is still more professional on indoor simulation.

CUSTIC

CUSTIC is a noise pollution modelling software. The programme calculates the noise level in full coverage of the space, considering all sound sources and the conditions of the atmosphere. The basis of the model is the linear sound propagation equation, which is used to model point source emissions from vehicles, industries and aircrafts. The Emission sources are categorized into point sources and line sources. CUSTIC is capable of noise mapping graphically, but no building model could be built in it, only the sound source is simulated. It is a simple version of noise prolusion mapping with less parameters.

ACOUSTICS MODULE OF COMSOL

COMSOL is a much more detailed simulation tool than those in urban scale, and it is concentrated on mechanical noise simulation. It is specifically designed for users of devices which produce, measure, and utilize acoustic waves. There are several application scenarios of speakers, microphones, hearing aids and sonar devices. It could also be used for muffler design, sound barriers, and building acoustics. It

contains powerful models for predicting acoustic pressure wave propagation in air, water, other fluids and solids. However it is not in the scope of this research.

RAYNOISE

LMS RAYNOISE is a computer-aided acoustic design and analysis system. Ray tracing methods are used to predict the sound field produced from multiple sources. It is suitable for simulations in a 3D space and in the far field outside. Complex interactions such as multiple reflections from different surfaces are operable. The results are presented in the form of 1/3 octave spectra and echograms, colour SPL maps and selectable sound quality metrics. RAYNOISE is mainly applied on indoor acoustic design. For meso- and macro-scale urban model simulation, detailed ray tracing method is very computationally expensive.

The software packages more widely used for noise mapping are LimA, CadnaA, IMMI, Predictor, Olive Tree Lab Terrain and SoundPlan. These simulation tools are very useful, especially at planning stages where measurements are not possible. Because CadnaA is easy-to-use, fits the research scale of this research and is accessible at the University of Sheffield, it is utilised as the acoustic domain simulation tool in this research.

CadnaA

CadnaA, fully named as Computer Aided Noise Abatement is a leading simulation program for noise and air pollution prediction. It is capable of calculation, presentation, assessment and prediction regarding environmental noise in multiple scales. The method applied is also ray tracing. Its calculation of noise emission is based on the international and national regulations, standards and guidance for traffic and neighbourhood noise. The suitable scenario of CadnaA is detailed noise mapping in a large-scale city. Its screen display, modelling space, setting interface and graphical calculation grids are parts of its highlights as well as the import and export of geometrical model data from third party modelling programs, the acceleration

ability of program controlled segmented processing, group variants, and table of results. Various project views are also available, such as 3d-model navigation.

4.3 Simulation Model Setup in CadnaA

4.3.1   Input and Output Form and Configuration of Simulation

A series of separate and interactive effects need to be considered in large distance sound propagation, including source characteristics, source-receiver distance, ground and air attenuation, wind speed and direction, temperature and relative humidity, barrier attenuation and acoustic screening, and surface reflection (Kang, 2006).
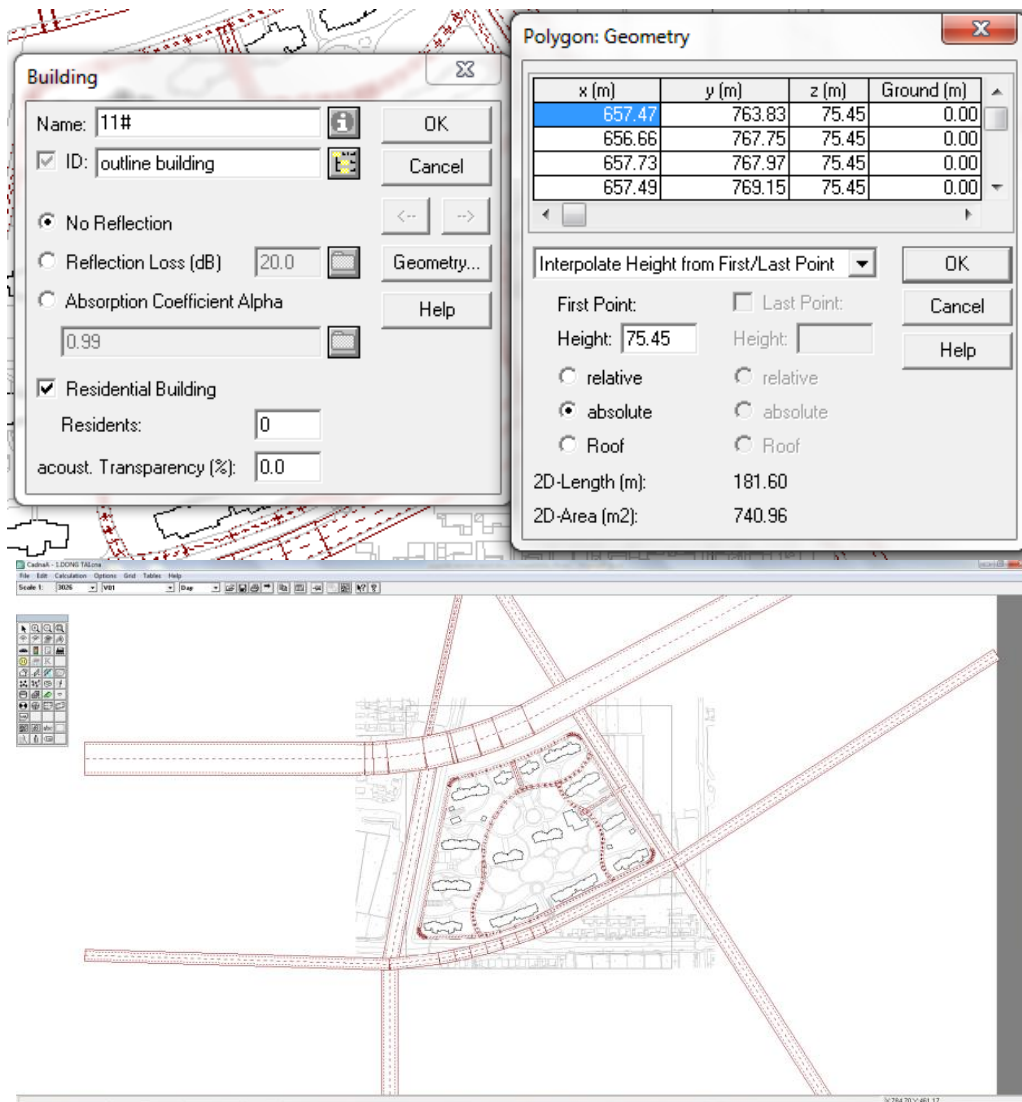
Figure 4.2 Building Geometry Modelling in CadnaA

The involved components in the simulations are sound source, building blocks (noise barrier), receivers and calculation grids. The simulation is organised under a structure of simplified building model generation, simulation configuration and result data consolidation and export.

For acoustic simulation, the building model needs to be simplified to reject unnecessary concern of non-significant factors and reduce calculation time. The outline of the building is the full shape of the building footprint, and will be extruded into a building block. Building outline used for meso-scale research is simplified by ignoring objects on a facade such as balconies, windows, blinders, etc. Vegetation is not included in this research, because the aim of the research focuses on how building distribution influences the outdoor environment. Vegetation may have a great influence on the acoustic environment, but vegetation species and volume may vary case by case, making it difficult to compare case simulations.

Building blocks are constructed as noise obstacles in the input. Building general plan of arbitrary closed polygons in format of AutoCAD DXF is imported into CadnaA as a footprint reference of building model generation. No terrain is modelled in this research. Buildings are extruded according to their heights in CadnaA and based on the absolute height. The reflection property of a building is presented by the absorption coefficient of the surface of building. It is empirically applied for all buildings in the simulation as 0.75 which is for residential buildings with the allowance of airing. The geometry setting panel and model is shown in figure 4.2.

The calculation grid is defined within a squared area covering all site boundaries (excluding area of road network around the site). The grid point data outside of the site boundary is later removed, to save only SPL in site for further analysis. This can be viewed in figure 4.3.

The spacing of the grid is set 3×3m, at a height of 1.75m. The appearance of the grid alters the form of the noise map, set as lines of equal sound level, areas of equal sound level, raster oversampling and without showing grid points value.

Figure 4.3 Grid Receivers Arrangement within Sample site

Except for the calculation grid, receivers are set in pairs along street facing buildings on both sides of the facade facing the traffic or in the noise shadow. Pairs of receivers are set 1m outside of street facing building facades on the roadside and away from roadside, and by default setting the sound reflections from their own facade surfaces are excluded. The separation distance of adjacent pairs along buildings is 3m. The receiver's height is defined as 1.7m above ground. Only pairs of receivers separated by buildings are kept. The pairs located at the interval area between buildings are removed because the difference of the two receiver values could not represent the noise barrier effect from buildings, as shown in figure 4.4. Because the following multi-domain multi-objective optimisation has not adopted the street facing facade

regression function, in order to reduce the complexity of the whole optimisation system, the simulation results and qualitative and quantitative analysis of street facing facades are located for the reader's reference in appendix G.
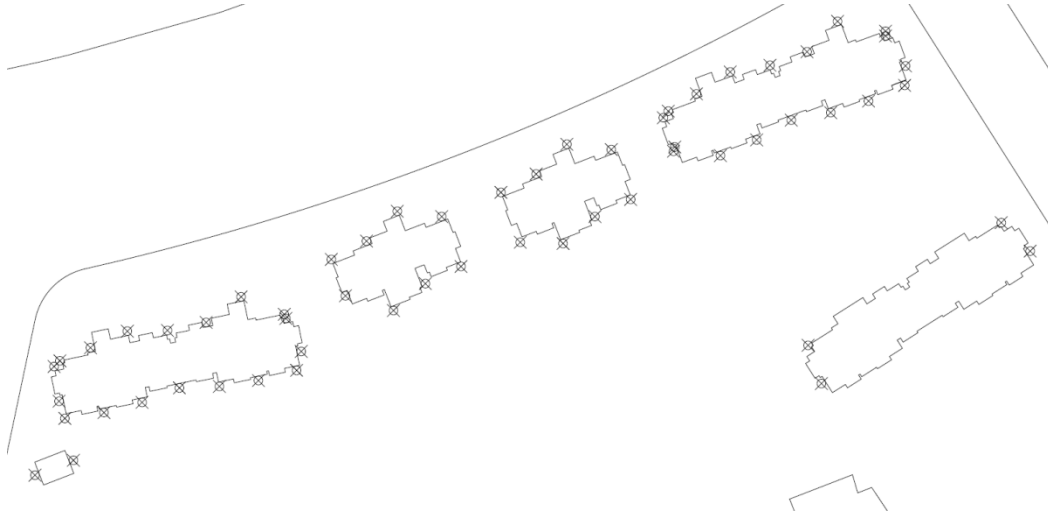
Figure 4.4 Paired Receivers Arrangement along Street Facing Facades

The output forms of simulation results over a site are 1) exported tables of SPL values at all grid points and at paired receivers set on the both sides of street facing buildings; 2) noise map generated on the grid over a site.

The output form of the simulation results are 2D horizontal noise maps (in form of iso dB-lines, noise contours, or raster oversampling) and SPL in ASCII format exported from grid or receivers. The SPL from paired receivers would be converted to SPL deference across street facing buildings.

The key configurations of the general model for the simulation are: 1) ground absorption being 0, referring to hard ground surface with no sound absorption; 2) max order of reflection at receivers being 3 times for a balance of proper precision of calculation and non-excessive time cost.

As the focus of this research is to discover how building distribution patterns impact SPL pattern, all irrelevant factors are excluded, i.e., ground absorption. Meanwhile,

the simulated noise map would not be present on its own, but used for comparison between simulated cases. So the conclusion would not be influenced by this idealistic simplification.

4.4 Sound Source Definition and Setup

4.4.1   Estimation of Traffic Noise Source Level

Calculation of Road Traffic Noise (CRTN) method has been validated to be reliable for predicting traffic noise in Hong Kong (Leung and Mak, 2008), and can provide accurate results under vertical urban settings with varying height (Mak, Leung and Jiang, 2010). LimA is utilised with the CRTN algorithm in the noise mapping of Hong Kong (Lam *et al.*, 2013).

The sound source applied in this research is line source representing traffic noise from roads outside of the site boundary. The line sound sources are generated from the street axis from the master plan. The parameters of linear sound sources are length and equivalent sound pressure level. Length is determined by the geometry of the road. Equivalent sound pressure level is calculated based on traffic count over 18h, traffic speed, ratio of heavy vehicles in daytime, evening and night on the particular road. Because this research is only focused on daytime simulation, evening and night setting will not be used and changed.

The traffic counts over 18h are equivalently estimated based on the designed traffic load of the planned street. Designed traffic load is the maximum allowance on a road. Therefore, the traffic noise of a certain road is defined as its highest potential noise level, which may be significantly higher than the real condition. Although the sites are real practices, to simplify the basis data collection, this estimation of traffic noise level is adopted instead of a live measurement at site. Because the following analysis is compared between cases within this research, searching for differences of SPL due

to variation of building distribution over estimation of sound source level does not cause invalidation in the analysis.

To determine the traffic counts of certain roads, road classification, road width, track numbers and population of the located city in each model are determined according to the Code for Transport Planning on Urban Road GB 50220-95. The traffic flow on each road is estimated by Urban Road Design Code CJJ 37-90.

The procedure of determining line sound source level is to:

1) Define city scale based on its population. As all cases are located in large scale cities, only the data of road network planning in large scale cities are consolidated in table 5.1 and 5.2.

2) Define road classification (expressway, artery road, minor artery road, and local road) based on its width and inquire corresponding speed limit and track number, according to table 4.1;

3) Inquire 18h accumulative traffic counts based on the number of tracks within the road and road grading according to table 4.2.

Table 4.1 Road Grading Referenced from Code for Transport Planning on Urban Road GB 50220-95

Item	City Scale	Population (Million)	Road Classification				
			Expressway	Arterial Road	Minor Arterial Road	Local Road	Local Drive
Road Width (m)	large cities	>200	40-45	45-55	40-50	15-30	<15
		≤200	35-40	40-50	30-45	15-20	<15
Design Speed of Vehicles (km/h)	large cities	>200	80	60	40	30	20
		≤200	60-80	40-60	40	30	20
Number of Tracks Within Road	large cities	>200	6-8	6-8	4-6	3-4	1-2
		≤200	4-6	4-6	4-6	2	1

Table 4.2 Road Traffic Count Calculation

Item	Road Classification			
	Arterial Road	Minor Arterial Road	Local Road	Local Drive
Coefficient of Track Categorisation α_c	0.8	0.85	0.9	1
Possible Traffic Capacity per	1690	1640	1550	1380

Track N_p (pcu/h)				
Design Traffic Capacity per Track (pcu/h) $N_m=\alpha_c \cdot N_p$	1352	1394	1395	1380
Traffic counts	1 track	/	/	/	1380
	2 track	/	/	2790	/
	3 track	/	4182	4185	/
	4 track	5408	5576	5580	/
	6 track	8112	8364	/	/
	8 track	10816	/	/	/
	1 track*18h	/	/	/	24840
	2 track*18h	/	/	50220	/
	3 track*18h	/	75276	75330	/
	4 track*18h	97344	100368	100440	/
	6 track*18h	146016	150552	/	/
	8 track*18h	194688	/	/	/

Sound source emission required for the calculation of noise impact in the vicinity of roads is converted from traffic accounts as determined above. CadnaA provides various standards to generate traffic noised model. The standard applied in this research is the British Road Noise Model CRTN (Great Britain. Department of the Environment, Great Britain. Ministry of Housing and Local Government, Great Britain. Ministry of Public Building and Works, Great Britain. Ministry of Transport, Great Britain. Department of National Heritage, Great Britai, 1975).

Emission of sound source is calculated based on input counts of vehicles per 18 hours (6:00-24:00) daily, and the road type determined the proportion of heavy vehicles. This input will be equivalently converted into emission level of L10 in dB(A) for three daily time periods of day, evening and night. L10 is a percentile at 10% of all data which suggests that in 10% of the time this value is exceeded during the whole record period and it is an indicator of the average peak value of the sound source. A speed limit is also input according to the designed speed of a certain road. Road width considered from curb to curb is also required to determine the location of line source. According to CRTN paragraph 4, the source of traffic noise (emitting line source) is

defined 0.5m above road surface and 3.5m inward from the nearside of road curb. Road setting is shown in figure 4.5.

Although the sites are located in China, CRTN criterion (UK) is still applied for estimation of traffic noise level. Because the result of an individual case would not be presented on its own, but would be used for comparison between cases, the final conclusion would not be influenced. An acoustic simulation bird view effect is shown in figure 4.6.
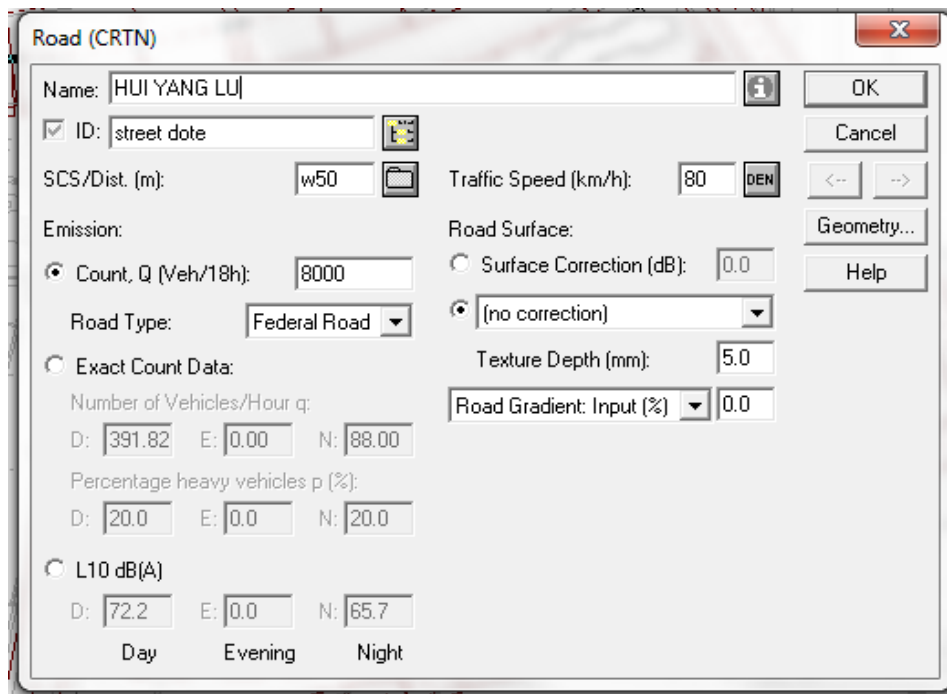
Figure 4.5 CRTN Road Setting Panel in CadnaA
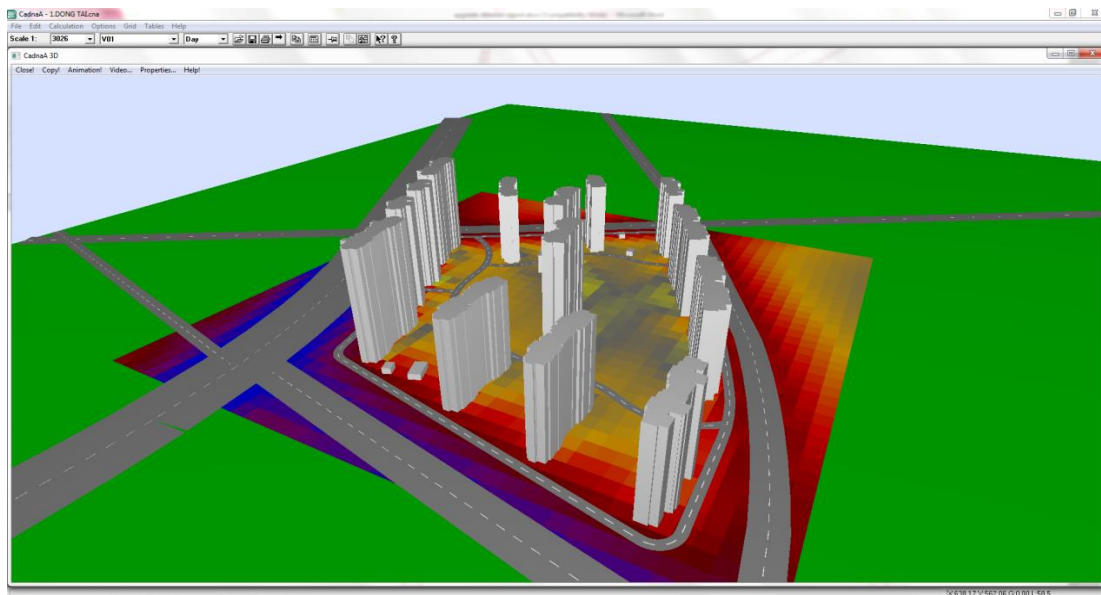
Figure 4.6 Bird View of the Noise Map of Sample Site

4.5 Selection of Acoustic Performance Indices for Assessment and Analysis

4.5.1 Parametric Study Result of Sound Pressure Level

The parameters applied in the research regarding acoustic performance of over-site condition and front facade attenuation are SPL(L_{Aeq}) at grid points and SPL difference across street facing building facades. The measure is equivalent continuous A-weighted sound pressure level which is the average sound energy at A-weighted level within a certain recorded time period.

Statistical measures of these two parameters are compared and selected as representative indicators of the data distribution. Analysis approach of statistical measure selection is variable clustering by between-group linkage method. It helps to define distance between statistical measures based on their characteristics.

Observation on the curve of P10-P90 of each case is also conducted to affirm turning points or key percentiles from the whole data distribution of all cases. Selections of representative percentiles refer to the key measures representing turning points and the separation of those measures in variable clusters.

4.5.2 Grouping and Selection of Statistical Measures of SPL

Through variable hierarchical clustering of ten percentiles and two quartiles, the groups are divided as below: low range (P10, P20, P25, P30), mid range (P40, P50, P60) and high range (P70, P75, P80, P90). The threshold measures of ranges are P40 and P70.

As seen in curves of percentiles of all cases (figure 4.7), taking clusters of fair acoustic performance as an example, key turning points are P10, P40 and P70. Cases in the fair performance group, where the majority of cases were located, could represent the majority. Similarly, the turning points of clusters with good-very good and poor performance are also close to P40 and P70, however are slightly higher or lower. Therefore, P40 and P70 especially distinguish power in separating different clusters of performance. Because they are highly dependent on sound source level, P80 and P90 are not included as they do not contribute to the current research.

P70 is the seventh percentile of SPL grid data. It is the turning point on the curve of percentiles of data between a fair and a noisy zone of a site. The gradient of the curve on the left of P70 would be greatly steeper than on the right of P70; in other words, noise deduction below P70 is significantly faster than above P70. As discussed in cluster analysis, in the histogram of grid SPL there exists a valley bottom indicating the peak value of noise barrier effects due to street facing buildings. The valley bottom is located around 70dBA for fair performance sites, yet with small variations slightly higher than 70dBA and less than 75dBA for poor performance sites. On the corresponding curve of data, for sites with fair-good performance, 70dBA is exactly the value of P70, while for good-very good performance and poor performance 70dBA is the value between P70-P80 and P60-P70, respectively. In other words, for good-very good performance sites the steep drop of noise level started at 70dBA occurs between P70-P80, so P70 value lower than 70dBA. Conversely. for poor performance sites, the P70 value would be higher than 70dBA. Therefore, P70 is a good distinguishing parameter capable of separating very good, fair and poor performance sites.

P40, as show on curve of percentiles of data, is the turning point of a quiet zone and a fair zone of a site. The gradient of the curve below P40 becomes moderate compared to between P40-P70 for the majority of sites with fair-good acoustic performance. Similarly, as displayed on the histogram of grid SPL data, 60dBA is the value of P40

for fair performance sites, 60-65dBA and 55-60dBA for poor and very good performance clusters, respectively. Therefore, P40 is a proper variable with which to distinguish different clusters from a mid-range SPL viewpoint.

P10 represents the quietest area SPL without considering extreme values resulting from special building shapes and arrangements. Therefore, P10 is adopted as an indicator of low range level.
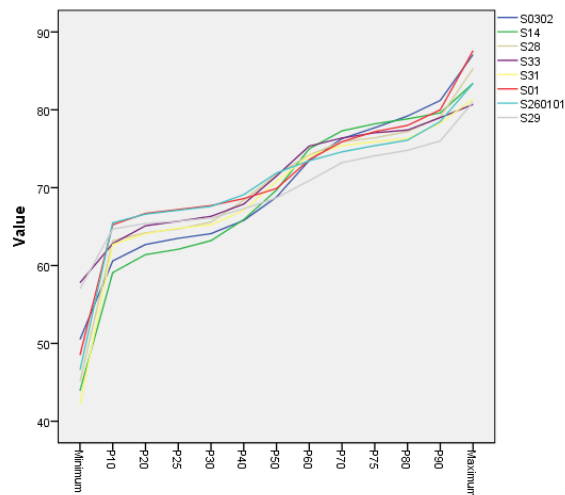
Figure 4.7 Curves of P10-P90 of Majority Case

Except for percentiles of grid SPL, to describe data distribution characteristics, interquartile range and ratio of grid points meeting critical value (65dBA) are also involved. Interquartile range is a presenter of a range of data without extreme value. Extreme values are suggested to be removed before analysis because, for acoustic study in this research, extreme values are results of special conditions that do not contribute to the generalisation of the analysis. The ratio of grid points meeting critical value (65dBA) is an indicator of the ability of forming an acoustic-acceptable zone through a building distribution approach. Its calculation is the ratio of grid points achieving a critical threshold value mentioned in regulation and guidance which is 65dBA converted for this research background, out of the total grid point amounts.

Table 4.3 Summary of Over-Site Acoustic Performance Variables Applied in Research

Parameter	Abbreviation	unit	range		Definition	Reason
P10	SPL-P10	dBA	44	70	10% percentile of SPL grid value of a site	Substitution measure of minimum value, indicating level of low range
P40	SPL-P40	dBA	51	74	40% percentile of SPL grid value of a site	Turning point of curve between low and mid range, marking attenuation due to inner building distribution
P70	SPL-P70	dBA	61	81	70% percentile of SPL grid value of a site	Turning point between mid and high range, marking attenuation due to street facing buildings
interquartile range	SPL-IQR	dBA	7	25	The range between P25 to P75	An indicator of data spread without impact of extreme values
SPL65Ratio	THR(65)	%	0.95	78.5	Ratio of grid points met acceptable threshold of less than 65dBA	An indicator of the ability of forming acceptable zone by building distribution. The achieved ratio below critical threshold value mentioned in regulation and guidance

To summarise, P10, P40 and P70 are selected as representative statistical measures of SPLs on grids, describing low range, mid range and high range SPL, respectively. Interquartile range and threshold ratio at 65dBA are also adopted (table 4.3).

4.6 Simulation Result Qualitative Analysis

4.6.1   Cluster analysis of sites by SPL

Hierarchical cluster analysis is conducted for sample cases by P10-P90 (including P25 and P75). The clustering is on the basis of between-group linkages, calculated by squared Euclidean distance between cases.

The result of hierarchical clustering indicates that sample cases are grouped into four clusters: very good (cluster D), good (cluster B), fair (cluster A) and poor performance (cluster C). Each cluster is discussed in detail considering the common pattern of the curve and histogram of acoustic data (Figure 4.8).
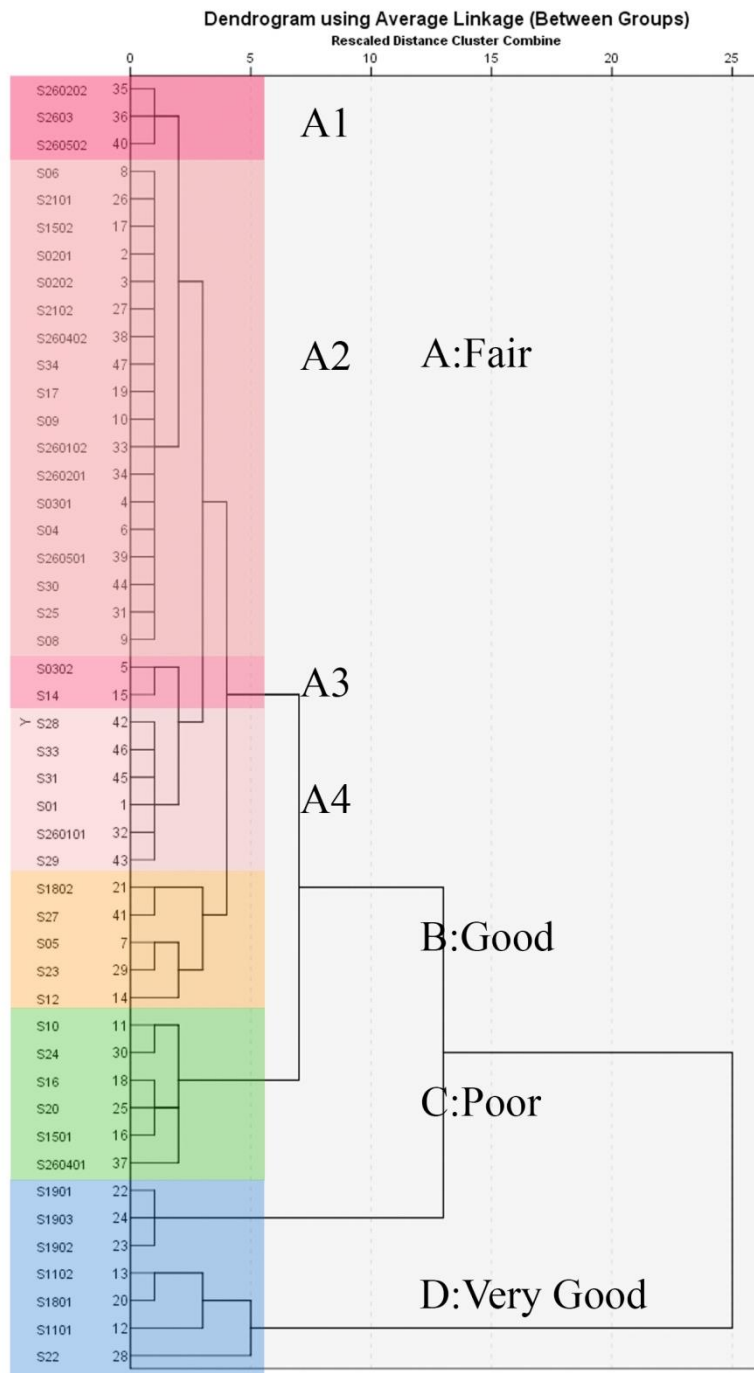
Figure 4.8 Dendrogram of Clustering Analysis of Case by P10-P90 of Grid SPL

In general, the curve of acoustic data distribution appears in a gradually increasing trend with several turning points, unlike the dramatic leap appeared in curves of mean thermal data distribution. The turning points are between P30-P40 and P70-P75.

Based on the comparison of frequency distributions (histograms) of all sites and relevant noise map, generally the histogram has two kurtosis; namely it is a bimodal

distribution. The higher peak is dominated by the sound source level and the distance of front facade to sound source. The location and height of valley bottom indicates the noise barrier effect from the street facing buildings. The lower peak is influenced by the characteristics of inner building distribution.

Referencing Chinese acoustic regulation for civil architecture design GB 50118-2010, the allowed noise Level for indoor residential environment (living room) is less than 45dBA for first grade architecture and 50 dBA for second grade.

As defined in the Environmental Quality Standard for Noise GB3096-2008, residential ward is classified as a first-grade functional zone which requires a relatively quiet environment; or as second grade if mixed with other civil functions which requires a relatively quiet environment adjacent to a residential area. The third-grade zone refers to industrial and warehouse functions. The allowed equivalent A-weighted sound pressure level from environment for first grade zone is limited to 55dBA in day-time (06:00-22:00) and 45dBA in night-times (22:00-06:00). For second grade and third grade zone it is limited to 60dBA (daytime), 50dBA (night time) and 65dBA (daytime) and 55dBA (night time), respectively.

In this simulation the sound source level is over estimated due to the adoption of designed traffic counts (maximum load) for the calculation of equivalent sound source levels. Therefore, the traffic noise around a residential ward would be stronger than in practise, and the simulation would find it difficult to meet the requirements of current regulations. For the purpose of easy application of analysis in this research, the criteria of allowed environment noise is defined as 65dBA for daytime and 55dBA for night time, referencing the criteria for the third-grade zone.

Limited by the length of the thesis, for the reader's reference the qualitative analysis processes are located in appendix D, qualitative analysis process of acoustic simulation results.

4.6.2 Consolidation of Clusters by SPL P10-P90 and Summary

The data distribution of each cluster as indicated by the histogram pattern is consolidated in table 4.5. Several rules of histogram patterns and acoustic performance characteristics are summarised below.

It is clear that only clusters of very good and poor performance have strong sound source levels, indicating that even in strong traffic noise, a quiet residential environment is achievable by arranging building distribution.

In the table, clusters with enclosed boundary and less enclosure would have a valley bottom of 1-2% and 3%, respectively. Except for 5% of valley bottom height, the location of valley bottom also tends to skew left below 70dBA for sites with open boundary.
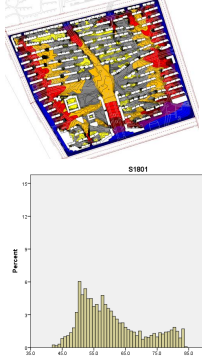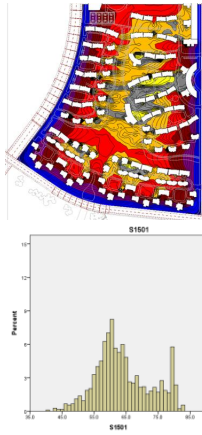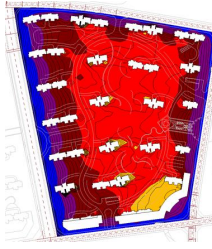
Cluster of sites with a higher level of inner noise attenuation ability would have a left-skewed low peak below 60dBA, and a lower height of various value noted as *variou*s in the table 4.4. Conversely, a cluster of sites lacking in inner noise attenuation would have concentrated level around 60-65dBA of height over 10%.

Table 4.4 Histogram Characteristics of All Clusters Based on SPL P10-P90
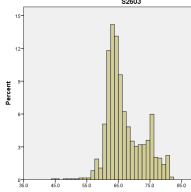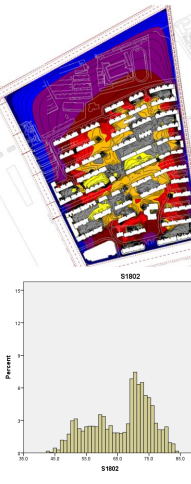
Performance	cluster	Sub	Low peak		Valley bottom		High peak	
			height	Location (dBA)	height	Location (dBA)	height	Location (dBA)
Very good	cluster D	/	6%	50	1%	70	Various	80
Good	cluster B	/	Various	60	3%	70	Various	75
Fair	cluster A	A1,Type1	13%	60-65	3%	70-75	Various	75
		A1,Type2	10%	60-65	2%	70-75	Various	75
		A1,Type3	12%	60-65	1-1.5%	70-75	Various	75
		A1,Type4	8%	60-65	5%	65-70	Various	70-75
		A2	>10%	65-70	5%	70	Various	70-75
		A3	Various	55	3%	65	Various	70-75
		A4	Various	60-65	2%	70-75	Various	75
Poor	cluster C	/	12%	70	5%	75	Various	80
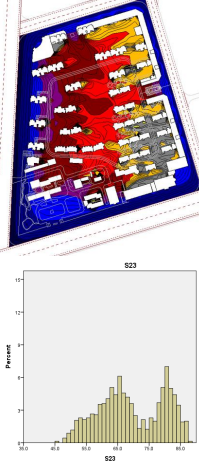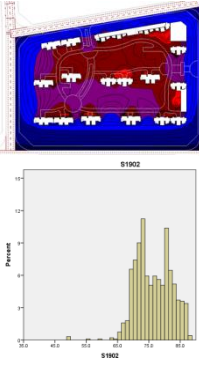
The clustering analysis of sites by SPL P10-P90 is consolidated in the table 4.5

.

Table 4.5 Consolidation of Clustering Analysis of Samples Based on SPL P10-P90

Cluster	Sub-groups	Distribution Description	Example	Noise map and histogram	Histogram Characteristics	Conclusion
D: Very Good	/	Small distance from street, enclosed site boundary (boundary attenuation between 70-75dBA) high density and relatively long facade (inner attenuation between 40-60dBA)	S1102, S1801 S1101 S22	Attenuation: over 30-40dBA; Range: 45-75dBA S1801	Location: low peak (50 dBA) valley bottom (70 dBA) high peak (80 dBA) flat and thick tails on both sides of low and high peak. clear and deep valley bottom	1) highly enclosed site boundary and small distance to road cause dramatic noise attenuation from 75dBA to 70 dBA; 2) high building density and facade length for potent noise attenuation at inner area of the site.
B: Good	/	Noisy site, noise limited barrier effect at P70 (70-75dBA); Relatively long continuous facade, large building separation and low-medium density causes limited inner attenuation at P10-P40 (55-65dBA)	S10 S24 S16 S20 S1501 S260401	Attenuation 20-30dBA; Range: 55-75dBA S1501	Location: low peak (60 dBA) valley bottom (70dBA), high peak (75dBA) short and thin tails on both sides of low and high peak. clear and deep valley bottom	1) P70 represents noise drop by barrier effect of the street facing buildings, P10-P40 represents the noise attenuation ability of inner building distribution. 2) Large proportion of area in satisfactory noise level (55-65dBA) could also achieve general good acoustic performance 3) For inner noise attenuation, long continuous building facade is positive factor while large building separation and low-medium density is negative factor. When combined it leads to satisfactory inner attenuation. 4) Less boundary enclosure results in taller valley bottom appeared on histogram of grid SPL data.
A: Fair	A1	1) Small distance to road, lack of noise barrier effect (open boundary and low-medium density) 2) very small site of enclosed boundary 3) large	S260202 S2603 S260502 S06 S2101 S1502 S0201 S0202 S2102 S260402 S34 S17 S09 S260102 S260201	Attenuation 20-30dBA; Range: 60-80Dba S2603	Location: low peak (60-65 dBA) valley bottom (70-75dBA), high peak (75dBA) narrow spread and pointy low peaks, and taller valley bottom caused by lack of inner noise	1) Short in height of valley bottom in histogram indicates thorough noise barrier effect at street facing buildings. 2) High density, shorter building separation distance is good for inner noise attenuation ability; 3) Enclosure works well for noise attenuation, but small size would weaken its power; 4) Set-back distance from street facing buildings to road may not

		distance to road, open boundary, low-medium density.	S0301 S04 S260501 S30 S25 S08		attenuation	contribute to inner area acoustic performance if no front façade barrier existed; it only avoids noise stress of street facing facades. For better inner acoustic performance, enclosed site boundary with buildings however sacrificing street facing facade being exposed in strong noise, could be more effective than approach of setting-back all buildings from road.
	A2	Medium-large distance to road resulting attenuation at 70-75dBA. Open boundary, low density, large building separation distance resulting attenuation at 65-70dBA	S0302 S14 S28 S33 S31 S01 S260101 S29	S260101 	Location: low peak (65-70 dBA) valley bottom (70dBA), high peak (70-75dBA) left skewed high peak with thin left tail and thick right tail, tall valley bottom, narrow , pointy and right-skewed low peak with thin tails on both sides; poor on-edge and in-site noise attenuation ability	5) Sound source is very influential on overall noise level if boundary not highly enclosed;
	A3	Wide distance between buildings and road resulting attenuation at 70-75dBA Enclosed boundary, long continuous facade, medium-high density resulting attenuation at 40-65dBA	S1802 S27	S1802 	Location: low peak (55 dBA) valley bottom (65dBA), high peak (70-75dBA) flat and left-skewed low peak, deep and left-skewed valley bottom caused by good inner noise attenuation; left-skewed high peak with dramatic drop on left and thick right tail due to free noise propagation between wide distance of building and road.	1) For site with small and large separation distance between buildings and roads, the turning point on curve indicating noise barrier effect at street facing façade is P70 and P50 respectively. 2) Large distance to road to allow noise level to drop to 65-70 when meeting front row building facades 3) When combining large distance to road with enclosed boundary, it would lead to very good inner attenuation without exposure front façade in strong noise, which could be the best approach of noise attenuation without considering economic reasons.
	A4	Half-enclosed , low-medium density,	S05 S23 S12	S23	Location: low peak (60-65 dBA)	Enclosed boundary is very helpful for inner acoustic performance of medium density site lack of inner

		normal distance to road resulting inner attenuation at 60-65dBA			valley bottom (70-75dBA), high peak (75dBA) left-skewed low peak with flat left tail as in cluster D of best acoustic performance; Clear valley bottom.	noise blocking ability.
C: Poor	/	Strong source level, less enclosed boundary, small site resulting insufficient boundary attenuation. low density, short building facades resulting insufficient inner attenuation	S1901 S1903 S1902	S1902	Location: low peak (70 dBA) valley bottom (75dBA), high peak (80dBA) greatly right-skewed and pointy low peak, tall valley bottom, thick right tail of high peak,	The distribution worth avoiding from acoustic performance point of view: small site of less enclosed boundary and low density, especially when exposed to strong sound source.

The rules extracted from clustering analysis of SPL P10-P90 could be consolidated in three levels, macro, meso and micro:

Macro level (graphical distinguishing of acoustic performance)

Turning points on curve:

1. P70 represents noise drop by barrier effect of the street facing buildings, P10-P40 represents the noise attenuation ability of inner building distribution.

2. For sites with small and large separation distance between buildings and roads, the turning point on the curve indicating noise barrier effect at street facing façade is P70 and P50, respectively.

SPL histogram:

3. Valley bottom height: Less boundary enclosure results in taller valley bottom appearing on histogram of grid SPL data. Short in height of valley bottom, as displayed on the histogram, indicates a thorough noise barrier effect at street facing buildings.

4. Low peak: lack of inner noise attenuation ability leads to pointy and narrow low peak of considerable height in percentage.

5. High peak: wide distance between street facing buildings and corresponding road leads to thick left tail of high peak.

Meso level (various achievable strategies for good acoustic performance)

6. Best combination for overall good performance: When combining large distance to road with enclosed boundary, it would lead to very good inner attenuation without exposure front façade in strong noise. This could be the best approach of noise attenuation without considering economic reasons.

7. Worst combination for overall good performance: The distribution worth avoiding from an acoustic performance point of view: small site of less enclosed boundary and low density, especially when exposed to a strong sound source.

8. Generally good acoustic performance: achievable by large proportion of area in satisfactory noise level (55-65dBA), even though no quiet area (below 55dBA) existed.

9. Inner area acoustic performance: Set-back distance from street facing buildings to road may not contribute to inner area acoustic performance if no front façade barrier existed; it only avoids noise stress of street facing facades. For better inner acoustic performance, an enclosed site boundary with buildings - sacrificing

street facing facade being exposed in strong noise, (therefore requiring the installation of other attenuation approaches - could be more effective than approach of setting-back all buildings from road.

Micro level (influence of distribution approach on acoustic performance)

10. *For inner noise attenuation:* long continuous building facade is a positive factor while large building separation and low-medium density is a negative factor. When combined it leads to satisfactory inner attenuation.

11. *Large distance to road:* allowing noise level to drop to 65-70 when meeting front row building facades.

12. *Open boundary and Sound source:* Sound source would be highly influential on overall noise level if boundary not highly enclosed.

13. *Enclosure and small size:* Small size would weaken noise attenuation power inside the site of enclosed boundary.

14. *Enclosure and small distance to road:* Highly enclosed site boundary and small distance to road causes dramatic noise attenuation from 75dBA to 70 dBA;

15. *Enclosure and low-medium density:* Enclosed boundary is very helpful for inner acoustic performance when a site lacks inner noise blocking ability.

16. *High density and shorter building separation distance:* Contributes to inner noise attenuation ability.

17. *High density and long continuous facade length:* Results in potent inner noise attenuation.

The combinations of distribution configurations mentioned above could result in a satisfactory to a very good acoustic performance, as listed in table 4.6 according to their series number.

Table 4.6 Building Distribution Parameter Combinations
Suggested for Improved Over-Site Acoustic Performances

Distribution Parameter	Configuration	Overall Noise Attenuation			Histogram Appearance			
		Inner area	Street facing buildings	On-edge open space	Flat and left-skewed low peak	Pointy and left-skewed low peak	Deep valley bottom	Left-skewed high peak with thick right tail
Size	Large site	13,			13		13,	
Homogeneity	Even distribution	10			10			
Density	Dense separation	10, 16, 17			10, 16, 17			
	Loose separation	15				15		
Distance to road	Large	6		6, 11				6, 11
	Small		9, 14				9, 14	
Boundary enclosure	Open boundary							
	Highly enclosed	6, 9, 13, 15	6,9,13, 14, 15		6, 13	15	6, 9, 13, 14	
Facade length	Long facade	10, 17			10, 17			

4.7 Simulation Result Quantitative Analysis

4.7.1   Regression Parameter Consolidation

To conduct regression of building distribution with acoustic performance indices, firstly selections of performance and distribution parameters are required.

As discussed in chapter 4, a series of building distribution parameters are selected from each group of all parameters according to their clustering results. The selected building distribution indicators applied are listed in table 4.7.

Table 4.7 Representative Building Distribution Parameters for Regression

Group	Representatives	Abbreviation
Site scale	Total land area	TLA
Built amount	Residential building area	RBA
Existence of high-rises	TriangleSD	TSD
	Plot ratio	PR
Density	Building density	BD
	Average facade ratio	aFR

Set back	Average distance to road	aD
Site shape	Shape factor	SF
Low and long facade amount	Average low facade ratio	aLFR
Short facade amount	Average min facade length	aFLmin
Large interval amount	Average interval area	aIAmean
Small interval amount	Average lower corner area	aCAL
Interval depth amount	Average higher interval depth	aIDH

To summarise the overall condition of SPL on grid, statistical measures are used. Based on the parametric study in section 4.5.2, P10, P40 and P70 are selected as critical measures among all percentiles. The interquartile range and ratio of achieving criteria are also included. Details of the variables refer to table 5.3

4.7.2   Regression Result of P10 of Sound Pressure Level on the Grid

P10 is the first percentile of all sound pressure level data on a calculation grid set in a certain site. It suggests that 90% of grid points have a level higher than the value at P10. Therefore, P10 represents the lowest level both in low range and among all data, meanwhile removing the impact of the minimum value which is unstable and unpredictable due to multiple distribution reasons. To achieve a quiet residential ward, the P10 value is expected to be as low as possible as a result of inner and on-edge noise attenuation.

Two attempts are conducted for the regression of P10 with building distribution parameters: with distribution representative parameters and with all parameters (for a validation of the regression with representative parameters). The regression with representative parameters shows a holistic coverage of all related distribution factors of sound source level, site size, site shape, density, distance to source and enclosure at boundary.

The regression with representative parameters has an explanation power of 0.852 to the variation of SPL, and adjusted R square of 0.828, namely shrinkage of 2.82%

when generalising into the whole population. The explanation power and generalisation degree are both desirable.

Regression coefficients are listed in table 4.8. As shown, to achieve lower P10, it is desirable to have a higher building density and total land area, average distance to road and higher average minimum facade length, and a lower sound source level and shaper factor. It could be suggested that a larger square-shaped site in a high building density with a large separation distance to road and an enclosed boundary tends to have lower P10. This agrees with the best configuration combination in a qualitative analysis of clustering in section 4.6.2

Table 4.8 Details of Regression of SPL-P10 with Representative Parameters

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-4.009	17.263		-.232	.818
building density	-87.467	8.528	-.828	-10.257	.000
total land area (m^2)	-.372	.056	-.462	-6.591	.000
L10SoundSource	1.101	.236	.419	4.657	.000
ShapeFactor	1.561	.478	.225	3.266	.002
AvgDistance	-.130	.045	-.287	-2.918	.006
AvgMinFacadeLength	-.066	.031	-.154	-2.097	.043

Building density dominates the explanation power over all other parameters. The second dominative distribution factors are site size and source level.

It is noteworthy that boundary enclosure is represented directly by the average facade ratio (grouped in density cluster in variable clustering analysis), the average low facade ratio (grouped in long and long facade amount cluster) and the interval area (grouped in large interval amount cluster), and indirectly by average minimum facade length (grouped in short facade amount cluster). However, the involved parameter about boundary enclosure is minimum facade length including regressions with a selected statistical measure of SPL.

Minimum façade length refers to the short north and south facing facades, and gable façade facing east and west. However, gable walls cannot be much longer than 6m on

residential buildings due to the two-sided indoor natural daylighting requirement. Therefore, this results in long commercial façade wrapping outside of front row residential buildings, especially on the east and west sides.

SPL is more correlated to a minimum facade length rather than a facade ratio, a maximum facade length or an interval length.

4.7.3   Regression Result of P40 of Sound Pressure Level on the Grid

P40 is the fourth percentile of all SPL data on grid. It is the first turning point shown on the curve of data percentiles. The transformation of the curve gradient shows the inner noise attenuation ability by building distribution. A lower P40 is expected for a quiet in-site sound environment.

Regressions are conducted with representative distribution parameters and all parameters. The regression with all parameters is adopted for analysis because it reasonably involves more parameters and becomes more explanatory to the variation of SPL. Except for building density, total land area, sound source level and average minimum facade length in regression with representatives, average outline length (grouped in cluster of site size in building distribution parametric study in chapter 4), total corner area (grouped in cluster of large interval amount), shape factor (grouped in cluster of site shape) and average low interval depth (grouped in cluster of interval depth amount) are included in regression with all parameters. The seven parameters holistically cover the distribution factors. L10 of sound source level reflects the strength of source; average outline length combined with shape factor reflects the expectation of a large and square site; average minimum facade length reflects the boundary enclosure level; total corner area reflects the enclosure at the corner or the degree of high level enclosure (as only highly enclosed site tends to wrap the corner with buildings); average low interval depth reflects the limitation on passage of noise

penetration boundary enclosure and building density reflects the density of inner building distribution.

Regression of P10 with all distribution parameters has R square of 0.857 and adjusted R square of 0.828, namely the shrinkage in generalisation is 3.4%. The explanation power and the generalisation of the equation are both acceptable.

Coefficients of the regression are listed in table 4.9.

Table 4.9 Details of Regression of SPL-P40 with All Available Parameters

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	20.334	11.138		1.826	.076
BuildingDensity	-37.054	7.095	-.440	-5.223	.000
AvgOutlineLength	-.025	.004	-.493	-6.930	.000
L10SoundSource	.775	.143	.370	5.401	.000
AvgMinFacadeLength	-.093	.026	-.274	-3.565	.001
TotalCornerArea	0.264×10^{-3}	.000	.271	3.553	.001
ShapeFactor	1.057	.370	.191	2.859	.007
AvgIntervalDepthLow	-.127	.047	-.199	-2.687	.011

The influences of distribution parameters in the equation of P40 are relatively even. The most dominant parameters are building density, site size (represented by average outline length) and source level, followed by the three parameters related to boundary enclosure.

4.7.4   Regression Result of P70 of Sound Pressure Level on the Grid

P70 represents the seventh percentile of SPL grid data. It is the turning point on the curve of percentiles of data for the majority of sites with fair-good performance. P70 indicates the potent noise attenuation started from noise meeting an enclosed boundary to the inner area of the site. Seen on the histogram and curve of SPL grid point data of all sites, the P70 with value of 70dBA would generally be the location of the valley bottom of fair performance sites showing the peak of noise barrier effect from street facing buildings. As the good performance sites tend to have a lower P70

value, while sites of poor performance tend to have a higher P70 value, to achieve quieter overall acoustic environment, lower P70 is desirable.

The regression of P70 with representative distribution parameters is adopted for further analysis. The explanative power of the equation is 0.642 and adjusted R square is 0.604. The shrinkage of generalisation is 5.9%.

Regression coefficients are listed in table 4.10. The most dominant parameter is total land area. The equation shows that to achieve small P70, the following conditions are required: small distance to road (namely, barrier buildings close to road or earlier blocking of noise close to source), large total land area (referring to large site scale), less noise sound level and large average minimum facade length (referring to higher boundary enclosure).

Table 4.10 Details of Regression of SPL-P70 with Representative Parameters

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	26.384	16.825		1.568	.125
AvgDistance	.125	.039	.391	3.175	.003
TotalLandArea	-.394	.060	-.697	-6.576	.000
L10SoundSource	.633	.227	.342	2.789	.008
AvgMinFacadeLength	-.064	.031	-.212	-2.038	.049

4.7.5   Regression Result of Threshold Ratio of Sound Pressure Level on the Grid

The acoustic environment for a residential ward is desired to be as quiet as possible or as low an SPL value as possible. According to the aforementioned adjusted criteria of acceptable noise level in a residential ward, below 65dBA is considered the acceptable threshold. The threshold ratio (abbreviated as THR(65)) is the proportion accounted for by grid points with value less than 65dBA among all grid points, namely the proportion of acceptable area from the view point of outdoor acoustic assessment. The ratio is expected to be high for an overall good performance.

Two regression attempts are conducted: with representative distribution parameters and with all the distribution parameters. The two parameters involved in regressions, building density and source level are identical. The difference is that instead of total land area appearing in regression with representatives, a combination of average outline length and shape factor is adopted in regression with all parameters. Comparing to total land area, the combination reflects not only site size but also site shape information, that to achieve large THR(65) not only a large but also a square-shaped site is required. Therefore, regression with all parameters is adopted for further analysis.

The R square of the equation is 0.674 and adjusted R square is 0.640 with generalisation of 5.0%. Coefficients of the regression are shown in table 4.11. As seen in the table, a high ratio of value less than 65dBA requires high building density, a small source level, a high avg. outline length, and a small shape factor (square). The most dominate parameter is building density.

It is noted that the regression of THR(65) is not related to boundary enclosure and distance to road. The possible reason is that they both account for part of the practical scenario. In other words, an enclosed boundary or a large distance to the road could result in a large proportion of acceptable area, as agreed in the rule extracted from clustering analysis in section 4.6.2. The exclusion of these two factors does not mean they are irrelevant to the regression of THR(65), but it does not have to be enclosed or widely separated from the road to have a high ratio of area below 65 dBA.

Table 4.11 Details of Regression of SPL-THR(65) with All Available Parameters

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	250.632	67.131		3.733	.001
BuildingDensity	222.083	34.792	.607	6.383	.000
AvgOutlineLength	.115	.021	.521	5.351	.000
L10SoundSource	-3.550	.880	-.390	-4.034	.000
ShapeFactor	-4.597	2.269	-.192	-2.026	.050

4.7.6　Regression Result of Interquartile Range of Sound Pressure Level on the Grid

Interquartile range (IQR) is the range between two quartiles (P25 and P75). It shows the general noise attenuation ability of a site without impact from extreme values. The noise deduction from noise level before meeting boundary buildings to the level of an inner quiet area is represented by IQR. Therefore, IQR is expected as high for a quiet site.

Regressions with representative and all distribution parameters acquire identical results. The R square of regression is 0.758, and adjusted R square is 0.740, with generalisation shrinkage of 2.4%, shown in table 4.12.

The regression involves parameters from the three main distribution factors as concluded in clustering analysis: density, distance to road, and boundary enclosure. The most dominant parameter is building density. To acquire large IQR, it is required to have high density, long minimum façade length (more enclosed boundary) and large distance to road. In other words, large IQR is resulted by high inner attenuation ability, high on-edge ability and more noise deduction outside boundary buildings.

Table 4.12 Details of Regression of SPL-IQR with Representative Parameters

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-3.215	1.715		-1.874	.068
Building Density	59.711	6.392	.853	9.342	.000
AvgDistance	.103	.027	.341	3.868	.000
AvgMinFacadeLength	.066	.024	.234	2.703	.010

4.7.7　Regression Result of Maximum of Sound Pressure Level on the Grid

Interesting relationships could be extracted from the regression of maximum with representative parameters, although regression of maximum is not instructive for future residential ward design as it is dominated by sound source level. The regression

has an R square of 0.727 and adjusted R square of 0.714. The standardised regression coefficients are 0.809 and 0.251 for L10 sound source and average interval area, respectively. Similar results show in regression of maximum with all distribution parameters. The involved parameters are L10 sound source and total corner area.

Surprisingly, the regression points out that high maximum levels of SPL are correlated not only to high sound source level but also to a large interval area or a large corner area. This could be suggested that high maximum SPL tends to appear together with a larger interval area which includes sites of more interval area or more total corner area. However, facing a strong sound source level, the site boundary is in more need of enclosure than it is to a large interval area. These two parameter correlations would cause strong traffic noise leak into the inner area of the site, which should be especially avoided in design from an acoustic point of view.

Design habits of commercial buildings in residential wards from aspect other than environmental consideration could explain the counter-intuitive phenomenon. The commercial buildings with low heights and long façades are mostly designed along narrow drives and walks with considerable pedestrian flow.

The scale and aspect ratio along narrow drive and walk are fit to pedestrian requirements: less traffic, slow traffic speed, less noise and smaller aspect ratio, which means smaller building height divided by road width. On the contrary, wide roads with heavy traffic, or strong L10 sound source levels, tend to prevent pedestrians along the road due to a large amount of traffic, high speed, strong traffic noise and air contamination. The sufficient pedestrian flow ensures commercial activity along narrower road is more prosperous when compared to wide roads. This commercial requirement promotes the design of low-rise commercial buildings with long facades located along narrow drives and walks.

As a result, wide roads with high traffic noise tend to have few commercial buildings along them. This agrees with the result of regression of maximum SPL: residential wards wrapped by high traffic have a lack of commercial buildings as noise barriers along their edges.

In awareness of the phenomenon in current design, long facade commercial buildings or residential buildings is suggested located along the road, especially roads with high levels of traffic noise from an acoustic point of view. However, the suggestion would cause two conflictions: 1) the confliction of a lack of pedestrian flow and the need for commercial buildings along wide roads with heavy traffic; 2) the confliction of exposure facade in strong noise and immediate noise barriers close to the source by street facing buildings in a small distance to the road.

To solve the conflictions, designs could allow large set-back distances from building facades to the road, or could increase the application of noise attenuation installations on the noise-exposed facades. With the proper amount of set-back distance, noise exposure on facades could be mitigated. Furthermore, regarding commercial buildings, the set-back distance allows for the existence of open spaces as pedestrian zones in front of commercial buildings which could itself encourage pedestrian flow, not to mention fusing landscape design in the open space.

4.7.8   Consolidation and Discussion of Regressions of SPL on Grid

Comparison of Standard Coefficients of Regression

Table 4.13 Comparison of Standardised Coefficients of Over-Site Acoustic Regressions

Parameter Group		Sound source	Site scale		Site shape	Set back	Enclosure		Interval depth	Density
Models	Expect	L10	TLA	aOL	SF	aD	aFLmin	aTCA	aIDL	BD
P10	Small	.419	-.462		.225	-.287	-.154			-.828
P40	Small	.370		-.493	.191		-.274	.271	-.199	-.440
P70	Small	.342	-.697			.391	-.212			
THR(65)	Large	-.390		.521	-.192					.607
IQR	Large					.341	.234			.853

The standardised coefficients from regressions of SPL on the grid with multiple distribution parameters are consolidated in table 4.13. The influential power of the sound source level slightly reduces from P10 to P40 and P70, indicating that a sound source presents a stronger impact in a quiet area. Building density is the most influential parameter among all parameters; only P70 is not related to building density. P10 is more sensitive to building density than P40. Interquartile range show the same degree of sensitivity on building density compared to P10. Enclosure at boundary is also one of the wide influential parameters in SPL grid regression. From a low-level acoustic environment represented by P10 to a high-level environment represented by P70, and therefore on the interquartile range, enclosed boundary improves the acoustic performance. Only the ratio of acceptable area below the threshold value 65dBA could not be explained by enclosure, namely boundary enclosure is not necessary for achieving a large proportion of area below 65dBA.

Because different regression objectives of representative distribution parameters or all distribution parameters are used in regression, P10 and P70 are involved in total land area in regression indicating their correlation with distribution factors of site scale. But in regression of P40 and threshold ratio at 65dBA, average outline length and shape factor are involved as a pair, indicating impact of site scale. The discrepancy uncovers that the expected form of good overall acoustic performance is not only a large site area, but is also a square shaped rather than a long-narrow sliced site.

It is worth noting that average distance to road is the only parameter showing conflict expectation for a good acoustic performance from a low level (inner area) to a high level (close to boundary area) among all other parameters. For improving low level performance, wide aD is desirable, while for high level performance, short aD is desirable. In other words, if the design aims to acquire a very quiet inner area without considering other zones as main attenuation objectives, wide aD will contribute. If the design aims to achieve a not too noisy high-level area, also considering a quiet inner area if possible, locating a street facing building close to the road with a short aD

would be effective. These two approaches could be selected according to the different objectives of design.

Interval depth is only adopted in regression of P40. It shows that the existence of noise passage would impact the average noise level in the inner area. The deeper and narrower the passage, the less noise would leak into the site from the boundary.

Application of Regression Equations

Table 4.14 Equivalent Change of Building Distribution
to One Unit Improvement of Over-Site Performance Variables

Independent Variables			P10	P40	P70	THR(65)	IQR
Cluster	Name	Range	-3dBA	-3dBA	-3dBA	+10%	+3dBA
Sound source	L10	73-83	-2.725	-3.871	-4.739	-2.817	
Site scale	TLA	2-42	8.065		7.614		
	aOL	139-606		120.000		86.957	
Site shape	SF	1.5-4.5	-1.922	-2.838		-2.175	
Set back	aD	13-61	23.077		-24.000		29.126
enclosure	aFLmin	12-82	45.455	32.258	46.875		45.454
	aTCA	0-26073		-11363.636			
Interval depth	aIDL	10-36		23.622			
Density	BD	0.07-0.4	0.034	0.081		0.045	0.050

For easier application of the regression result to design practise, equivalent compromise of building distribution for one unit change on the performance index is calculated. Checked with range of each index, all equivalent compromises are valid and achievable in practise. Building density is the most effective controller where a slight increase of building density would cause a great improvement in low-level, mid-level and high-level SPL performance. Average distance to road is also effective and operable in practise: an increase of 23m or a decrease of 24m of distance between a street facing building facade and a road could achieve 3dBA attenuation in P10 and P70, respectively. The increase of minimum facade length at the boundary is also operable in a scale of approximately 40m, which is a reasonable scale for a commercial building wrapped around a boundary. If applied on a residential building

facade, a 40m increase equals to approximately 6 shared-wall apartment units, considering the south-facing width of an apartment being 6-7m according to local residential building design habit.

Relatively, the changes on distribution about site scale (total land area, average outline length) and enclosure at a corner (total corner area) for one unit performance improvements are costly in practise, although they are achievable.

All regression equations are listed as below in table 4.15 for reference.

Table 4.15 Regression Equation Comparison between Regressions of Over-Site Acoustic Performance

Model	Regression Equation
SPL-P10	SPL-P10=1.101×L10-87.467×BD-0.372×TLA+1.561×SF-0.130×aD-0.066×aFLmin-4.009
SPL-P40	SPL-P40=0.775×L10-37.054×BD-0.025×aOL+1.057×SF-0.093×aFLmin+0.000264×aTCA-0.127×aIDL+20.334
SPL-P70	SPL-P70=0.633×L10+0.125×aD-0.394×TLA-0.064×aFLmin+26.384
THR(65)	THR(65) =-3.550×L10+0.115×aOL-4.597×SF+222.083×BD+250.632
SPL-IQR	SPL-IQR = 0.103×aD+0.066×aFLmin+59.711×BD-3.215

It is worth noting that a minimum of SPL could not form a valid regression in a significant relationship with distribution parameters, regardless of whether it is with representative or all distribution parameters. It could be explained that the minimum value is unpredictably influenced by detailed building shape and spatial relationship between each other, which could not be predicted with the average measures of the building distribution.

4.8 Summary

This chapter operates the relationship exploration between building morphology parameters and on-site traffic noise attenuation performance indices, from qualitative and quantitative aspects.

With the help of statistical tools, the representative statistical measures of the acoustic performance metric SPL are defined, which describes the sound atmosphere of the simulated site. They are used to MLR regression and further MD-MOO optimisation.

Based on a combination of the cluster analysis of acoustic simulation data of key statistical performance indices, and the observation analysis of the histogram collections of all simulated samples, several descriptive rules of histogram patterns are concluded in matching with certain noise map forms, which could be directly applied as a judging basis of acoustic performance from an SPL distribution histogram.

Descriptive guidance of building distribution is suggested from cluster analysis of acoustic performance data. This could help in qualitative assessment and guides of design schemes on building morphology regarding corresponding SPL performance.

Multiple linear regression equations are achieved for 5 SPL statistical indices which are representatives of the acoustic climate of the site. These equations provide the mathematical relationships of the building morphology parameters and on-site acoustic performance indices considering multiple viewpoints of low level, medium level, high level, satisfactory ratio and degree of evenness. These MLR equations are also the sensitive analysis of design variables in the following meta-model construction for multi-domain multi-objective optimisation. The regression models could be directly applied in design, but it is also a choice to utilise the self trained meta-model for acoustic domain.

This chapter is one of the three parallel chapters of domain performance simulation and data analysis. They are the basis of further multi-domain multi-objective optimisation.

To summarise the general design rules for desired acoustic performance over the whole residential ward, it is suggested that:

The best distribution for overall good performance is when combining large distance to a road with an enclosed boundary which would lead to very good inner attenuation without exposure of the front façade in strong noise. It could be the best approach of noise attenuation without considering economic reasons.

A set-back distance from street facing buildings to road may not contribute to inner area acoustic performance if no front façade barrier exists. Thus, a compromised approach to achieve a quiet inner area is an enclosed site boundary with buildings, sacrificing street facing facade which is exposed to strong noise (therefore requiring the installation of other attenuation approaches). It could be more effective than the approach of setting back all buildings from the road.

Chapter 5 Sample Simulation in Sunlight Availability Domain and Performance Data Analysis

In the densely growing residential wards of high-density urban structures in Asia, a lack of direct sunlight is widely reported. To enhance and balance the sunlight availability of the dominate function room and outdoor environment, the most effective approach is to carefully arrange building layout and shape in the early design stage.

The aim of this chapter is 1) to define the proper performance indices for a sunlight availability assessment; 2) to explore the qualitative rules between building morphology characteristics and sunlight performance; 3) to define the mathematical relationships between building morphology parameters and sunlight performance indices. The three aims would be applied in the following multi-domain multi-objective optimisation for the integrated performance. The statistical tools of clustering analysis and multi-linear regression are used for qualitative and quantitative studies. The simulations of sunlight availability are operated in Rhinoceros and Grasshopper based on Ladybug and Honeybee algorithms.

Three performance evaluative indices are introduced, namely annual possible sunlight hour (APSH), winter possible sunlight hour (WPSH), sunlight hour in winter (SHW) for outdoor environments. Simulations are executed for these three indices for the samples. The parametric comparisons are carried out regarding the explanative power and emphasis over sunlight performances between index pairs, which are between sunlight duration on the ground and at window centroid, between APSH and WPSH on the ground, and between WPSH and SHW on the ground. Qualitative rules are excavated about their variation tendencies and the corresponding clustering from APSH and WPSH result distribution based on the various characteristics. Finally, multi-linear regressions between APSH/WPSH and building distribution parameter executed in SPSS are analysed and discussed in detail.

The structure of this chapter is: 5.1 Background of sunlight availability simulation; 5.2 Comparison of simulation software; 5.3 Simulation model setup; 5.4 Selection of performance indices for sunlight availability; 5.5 and 5.6 Qualitative analysis for APSH and WPSH simulation data; 5.7 Comparison of qualitative analysis results from APSH and WPSH; 5.8 and 5.9 Quantitative regression of APSH and WPSH against building morphology parameters; and 5.10 Summary. For the flowchart of this chapter see figure 5.1.
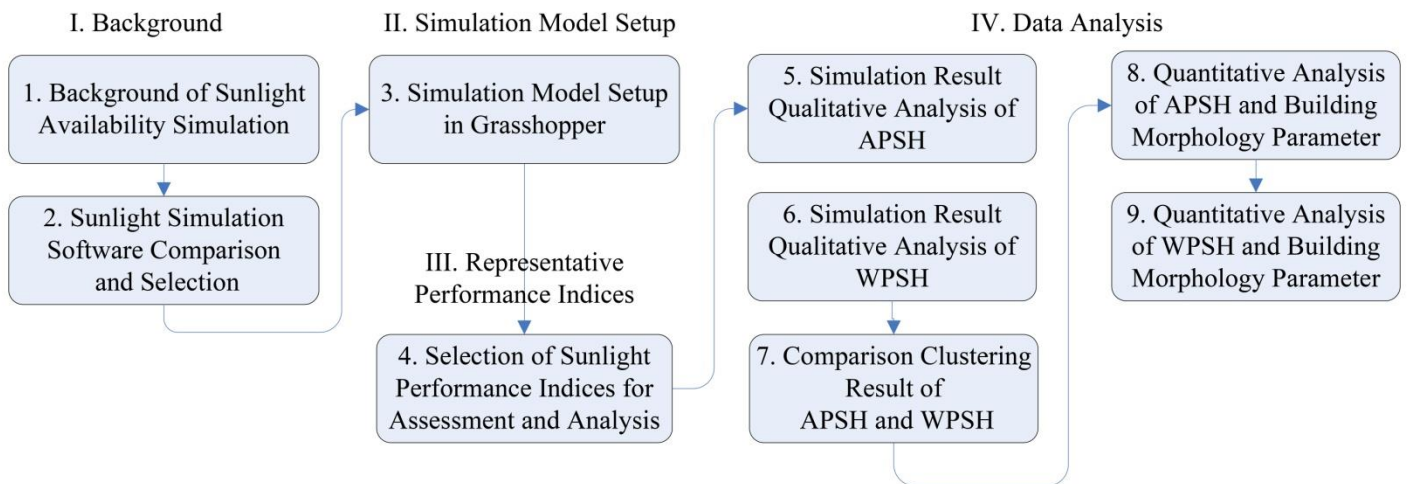
Figure 5.1 Content Structure of Chapter 5 Sample Simulation in Sunlight Availability Domain and Data Analysis, The expansion of box 5 in Overall Content Structure

Acronyms for Chapter 5

aCAH/AvgCornerHigh	average corner area high value
aCAL/AvgCornerLow	average corner area low value
aD/AvgDistance/Dmean	average distance to road of a site
aFLmin/AvgMinFaçadeLength	average min façade length
aFR/AverageFaçadeRatio	average façade ratio
aIAmean/AvgIntervalArea	average interval area
aIDH/AvgIntervalDepthHigh	average interval length high value
aIDL/AvgIntervalDepthLow	average interval depth low value
aILmax/AvgMaxIntervalLength	average max interval length
aLFL/AvgLowFaçadeLength	average low-rise façade length
aLFR/AvgLowFaçadeRatio	average low-rise façade ratio
APSH	annual possible sunlight hour
APSHF, WPSHF, SHWF	APSH, WPSH or SHW value at 1.6m facade level height grid
APSHG, WPSHG, SHWG	APSH, WPSH or SHW value at ground level height grid
APSH-IQR	the difference between two quartiles of APSH grid value of a site

APSH-P20	20% percentile of APSH grid value of a site
APSH-P50	50% percentile of APSH grid value of a site
APSH-P70	70% percentile of APSH grid value of a site
AR/AspectRatio	aspect ratio
APSH-THR(0)	ratio of grid points with sunlight duration equal to 0 hours
APSH-THR(413)	ratio of grid points with sunlight duration below 413 hours
BD/BuildingDensity	building density
dendrogram	a tree diagram that is used to represent and categorize hierarchical relationships among objects, created as an output from hierarchical clustering
FPA/FootprintArea	foot print Area
HRBA/HighriseResiArea	high-rise building area
HRBR/HighriseRatio	high-rise ratio
LMRBA/Low/medium-riseResi Area	low/medium-rise building area
P10, P20, P30, P40, P50, P60, P70, P80, P90	ten percentiles of all collected simulation data over one grid
P25, P75	two quatiles of all collected simulation data over one grid
PR/PlotRatio	plot ratio
RBA/ResiBuildingArea	residential building area
RFPA/ResiBuildFootprintArea	residential foot print area
SF/ShapeFactor	site shape factor
SHW	sunlight hour in winter
SVF	sky view factor
TLA/TotalLandArea	total land area
TSD/TriangleSD	standard deviation of triangle area
WPSH	winter possible sunlight hour
WPSH-IQR	the difference between two quartiles of WPSH grid value of a site
WPSH-P30	30% percentile of WPSH grid value of a site
WPSH-P50	50% percentile of WPSH grid value of a site
WPSH-P70	70% percentile of WPSH grid value of a site
WPSH-THR(0)	ratio of grid points with sunlight duration equal to 0 hours in winter
WPSH-THR(83)	ratio of grid points with sunlight duration below 83 hours in winter
Z-APSH	normalised APSH value
Z-SHW	normalised SHW value
Z-WPSH	normalised WPSH value

5.1 Background of Sunlight Availability Simulation

Regarding to daylight and sunlight studies in residential design context, indoor daylighting, outdoor sunlight radiation and energy cost is always the main stream. The indoor daylighting assessed by daylight factor, illuminance, solar radiation and visual comfort, is task orientated, which is most widely applied in natural lighting buildings or in sun-shading design with certain on-surface-task, i.e. reading, writing, applying computer or fixing, etc. There are also daylighting performance simulation for heavilu obstructed residential buildings ((Li *et al.*, 2006)). Outdoor daylighting is usually applied on open space design, many involving vegetation arrangements, some aiming to activity zone or facility design (Tsou, Chow and Lam, 2003).

Solar radiation is a popular interest of research as part of the indoor energy consumption, presented by heating and cooling load. It is a thermal objective applied in indoor context, while for outdoor the most researches are related to power generation efficiency of photovoltaic solar panels. Hence, solar radiation is not directly used in this research as a sunlight or thermal assessment metric.

Sunlight availability is evaluated by possible effective sunlight hour according to design guidance in UK and China (GB50096-2011). In China for residential planning, sunlight hour simulation under the obstruction by surrounding buildings is mandatory in design procedure. The focus would be building location, site layout and shading interaction between buildings. However, the design procedure is in a repeat iteration of scheme, test simulation, analysis, amendment scheme, and re-tests. The sunlight simulation would not provide any suggestions for amendment. The improvement of the schemes is mostly empirical based. Therefore, following the conventionally applied assessment method and explore optimisation opportunities is suitable for this research.

The possible sunlight hour implicates the fact that the estimation in the interested time span is only based on the theoretical solar vector available in the time span, with a deduction of sun vector due to overcast weather. The ratio of overcast days is estimated based on historical weather data of the city in which the site is located. The hours during the daytime of a day is daylight hour, while only daylight hour with solar radiation over 120Wh/m^2 is defined as effective sunlight hour. In this research, the effective sunlight hour is referred as sunlight hour and sunlight duration for the purposes of convenience.

Sunlight duration is evaluated by three measures under various critical standards, including annual possible sunlight hour (APSH), winter possible sunlight hour (WPSH) and sunlight hour in winter (SHW); at two different heights, on ground or ground floor window centroid. Therefore, the simulations of sunlight duration in Rhinoceros and Grasshopper consist of calculation on APSH (ground and window centroid), WPSH (ground and window centroid) and SHW (ground and window centroid). APSH and WPSH criteria is suggested by UK design guidance; SHW is obligatory in Chinese design regulation.

The assessment criteria for each evaluative parameter at ground level are borrowed from the criteria conventionally assessed on the façade. The critical values for APSH, WPSH and SHW are no less than 413h year round, no less than 83h in winter and no less than 270h in winter, respectively.

5.2 Sunlight Simulation Software Comparison and Selection

Multiple simulation tools are compared in this section. They have various simulation focuses and advantages.

INVENTOR TOOLS - FX64 LAMBDASPECT

It is the light simulation software for Autodesk Inventor Design. It is used for simulation of behaviour and interaction of light, based on the laws of physics and

light wave spectrum setting. It has extends the existing Inventor setting for material definitions, for further satisfying the requirements of additional optical material properties. However, it is not comprehensive enough to fit the research interest here.

INSPIRER

It is a package for image rendering and physically accurate lighting simulation. It is widely applied by architects, lighting designer, automobile and aerospace manufacturers. Applied in architecture context, it is capable of supporting optimised usage of natural light, visual effect rendering provided by outdoor lighting fixtures, shadow casting by buildings under sunlight or artificial light and simulation of the lighting effects for illumination characteristics. With its rendering tool, it is also capable of virtually reproduce the space or objects according to the provided light condition. The bi-directional Monte Carlo ray tracing method is used. Real-time ray tracing allows a very fast generation of photorealistic rendered image. However, this is a commercial software which having no accessibility from the university.

OPTIS' SIMULATION SOFTWARE FAMILY, SPEOS® AND OPTISWORKS

It is a comprehensive optical solution available for various types of products interactive with light. It is the only software package capable of human perception and lit appearance. The fast non-sequential ray-tracing engine is adopted, which is capable of one-shot processing with multi-threaded and distributed calculation. The simulation is physical-based, considering optical properties, spectral behaviour, and physical properties of materials, light sources and environments. One strong advantage of OPTIS is that it is integrated in AutoCAD, which could dramatically improve design effectiveness. A single-data model is used for seamless data transfer. The OPTIS GUI could provide a fast learning curve based on simulation result, which improving productivity in design procedure. It is also a commercial software package.

RAYFRONT - THE LIGHTING DESIGN TOOLKIT

It is an independent toolkit that provides a graphical user interface to the lighting simulation software Radiance which is the industry standard ray-tracing engine for physically correct lighting simulations. Rayfront is capable of daylighting and artificial lighting simulation. There are no limits for geometry size and complexity in Rayfront.

THE VIRTUAL LIGHTING SIMULATOR

It is a web-interface visualising tool for quickly check the effect of daylighting and artificial lighting designs under several key parameters. It contains large rendered image storage, computed by lighting simulation with Radiance. By searching and selecting key parameter value from the pop-up menu, a quick displaying of images is available. The database is arranged in two main modules.

Radiance

Radiance is a suite of programs for the analysis and visualisation of lighting in design. It is the one of the most widely used and validated software in lighting and daylighting fields. It is not only used by architects and engineers to predict illumination, visual quality and appearance of innovative design spaces, but also by researchers to evaluate new lighting and daylighting technologies. The rendering method of Radiance is Hybrid deterministic/stochastic (Monte Carlo) ray tracing. Direct illumination and specular reflections are calculated deterministically, while indirect diffuse contributions and their gradients are calculated and cached at surface points, used to estimate neighbouring values. Hence the advantage of Radiance over this simple lighting calculation and rendering tools is that there are no limitations on the geometry or the materials that may be simulated.

The simulation information would provided by AutoCAD files or input script files to specify the scene geometry, materials, luminaries, time, date and sky conditions (for daylight calculations). Its modelling capability includes general geometry, complex

light sources, light model, colour model, shadings, reflection and refraction, opaque and translucent materials, transport materials, surface maps, mist, smoke and clouds.

The output would be calculated values of spectral radiance, irradiance and glare indices. Simulation results could be displayed in colour images, numerical values and contour plots via RGBE HDR image files.

Radiance is so popularly applied in educational and research background as an open source. It is available as plug-in for Grasshopper in Rhinoceros. An easy call-in of the Radiance module to the visual programming platform Grasshopper for sunlight simulation of the building geometry models from Rhinoceros, is very labour efficient.

5.3 Simulation Model Setup in Grasshopper

5.3.1   Model Setting and Input for APSH, WPSH

To apply sunlight hour simulation, the start-up job includes the selection of location of site, definition of sun vectors, site model generation and settings.

The sites are all selected from Chinese Architectural Design Zone 4, sharing similar climate conditions. Therefore, to eliminate delicate differences in simulation results due to the impact of location differences, in this simulation all sites are set at the location of Hefei. Hefei located in the middle of SU-ZHE-WAN region, has a milder monsoon climate with characteristics of an inland climate, compared to cities on the edge of SU-ZHE-WAN region and close to the south-east coast where the climate is under the impact of the ocean.
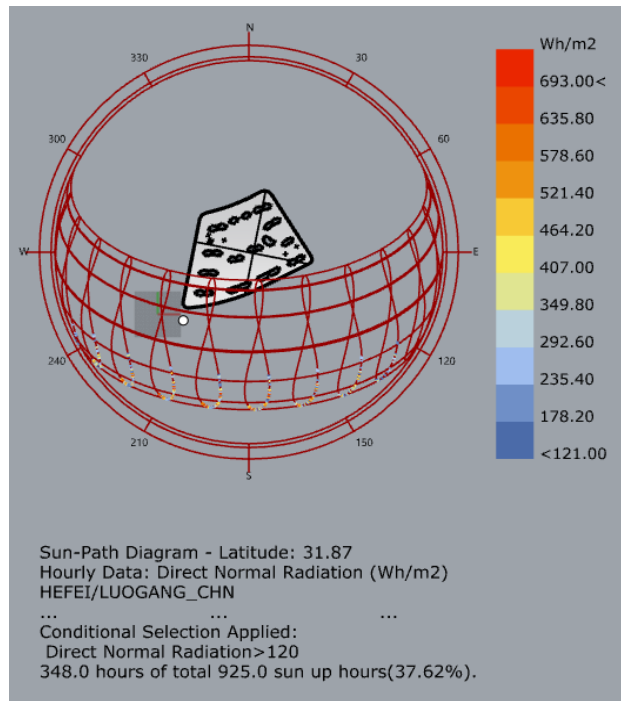
Figure 5.2 Sun Path at Hefei for WPSH
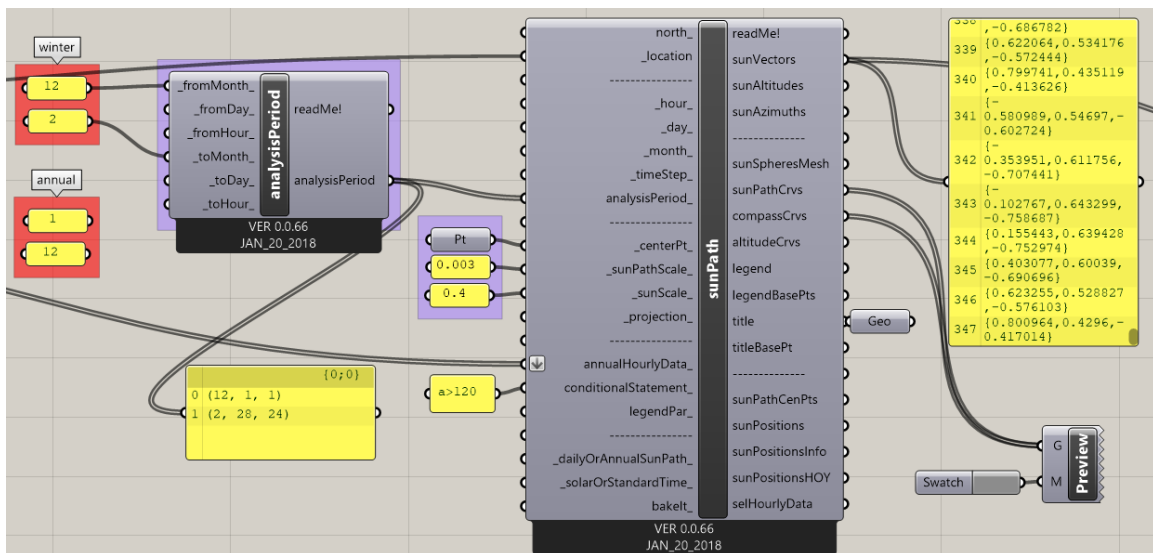
Figure 5.3 Sun Vector Setting for APSH and WPSH simulation

Sun vector is defined by tools of sun path in Grasshopper. Location input defines the latitude and longitude to calculate overall sun path movement. The analysis period limits the section of sun path which is used for simulation according to the corresponding output APSH or WPSH. The analysis period is January-December and December-February, respectively, for APSH and WPSH simulation. The input of annual hourly data applies solar radiation at this location, with a conditional statement of no less than 120 Wh/m^2 (Figure 5.2). Through this conditional input, solar vectors

providing effective sunlight are screened, which is 348 for winter and 1653 year-round. Through the above steps effective sun vectors are prepared for following simulation and are listed in note pad (Figure 5.3).

The model for simulation consists of the building model, the ground and 1.6m surfaces, and façade window panels at a height of 1.6m.

The site ground and building geometric model is built in Rhinoceros 6, based on an original AutoCAD site layout plot with information of building position and height. The building model is comprised of all structures above ground on the site, and includes residential buildings, commercial and public buildings, and entrances of underground parking. Accessories on building façades are excluded. The buildings are extruded out of their outlines of external façades (Figure 5.4). Ground surfaces and a surface at 1.6m height are necessary because the calculation grid is fitted on the selected surface in Rhinoceros 6. Surface normal is pointing up to allow it to receive solar radiation in simulation. The calculation grid surfaces are trimmed by building foot print area. This allows the removal of the calculation grid point within the building outline, which will have a calculation value of zero. If not excluded, those zero value points would have an impact on the statistic result on overall grid value.

To allow the calculation at the window centroid on the south-facing façade, referring to the tested window belonging to the main habitable room as mentioned in BS8206, a band of calculation area is generated for each building. Because window location and size is unknown, it is not necessary to have a precise calculation for each window at the beginning stage of design so simulation for insolation hour on each window centroid is simplified. A surface band on the façade at window height (from 1.1m to 2.1m) is used to present potential locations of windows. Test points are generated on the surface band in the middle at a height of 1.6m with an interval of 3m. These test points comprise the window centroids at 1.6m for insolation hour simulation, and the interval is less than 5m according to BS8206 for insolation simulation. These are used as representatives of potential centroid locations. In later analysis, the ratio of these

test points with values exceeding the criteria threshold is used to indicate the insolation quality on potential windows. As mentioned in BS8206 and GB 50180-93, only windows orientated to 90°of due south are required to test insolation hour. Therefore, only residential building façades orientated to 90°of due south are applied to create the surface band. By this method, a duplicate of separate receivers set at the centroid of each window panel on south-facing façades is acquired in a systematic way without specifying window position (Figure 5.5).
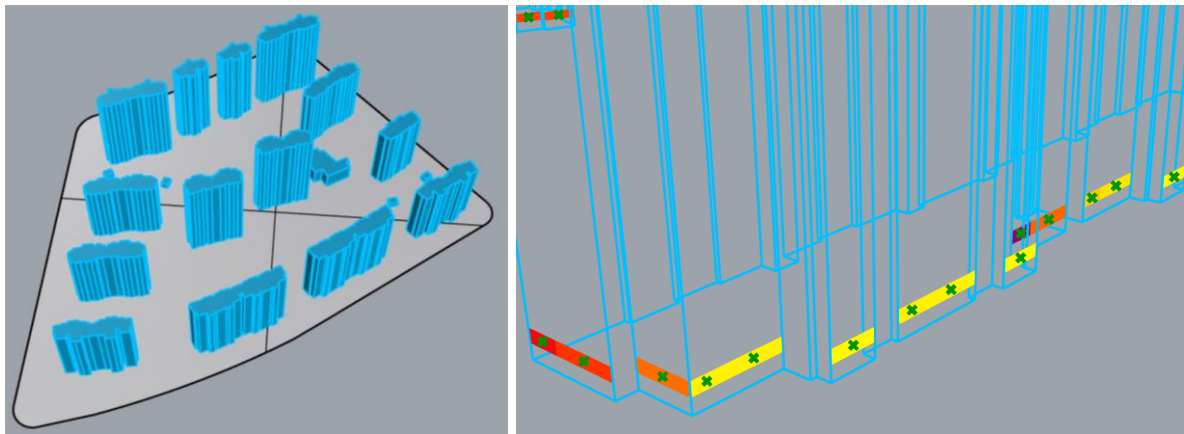
Figure 5.4 (Left) Model space of Rhinoceros + Grasshopper
Figure 5.5(Right) Grid Points on Façade Band

The simulation is executed in tool Sunlight Hour Analysis. The required inputs are context (building model), geometry (test surface), grid size (3m), the distance from base (0.01m), and sun vectors (Figure 5.6). The low bound and high bound of figure legends are set to be 413-1653 and 83-348h, respectively, for APSH and WPSH, according to the corresponding design criteria.

5.3.2   Model Setting and Input For SHW

As for simulation of sunlight hour, the selection of location of site, site model generation, and most of the settings are identical to that of APSH and WPSH.

In definition of sun vectors, the conditional statement of 120 Wh/m^2 is removed, so from December to February, 925 sun vectors are available. Correspondingly, the low bound and high bound of the legend for SHW is set to 270-925h.

5.3.3 Simulation Output of APSH, WPSH and SHW

Output of the simulation includes sunlight hour maps and a data list for all 45 sites. Maps consist of APSHG, APSHF, WPSHG, WPSHF, SHWG and SHWF. The maps show the shadow patterns of a specific building distribution under the same climate with different time spans. The legends are set to the range from the lower threshold to the maximum value, these are 413-1653h, 83-348h and 270-925h for APSH, WPSH and SHW, respectively. The white blocks indicate the building footprints. The area in blue indicates an area below the acceptable threshold of sunlight.
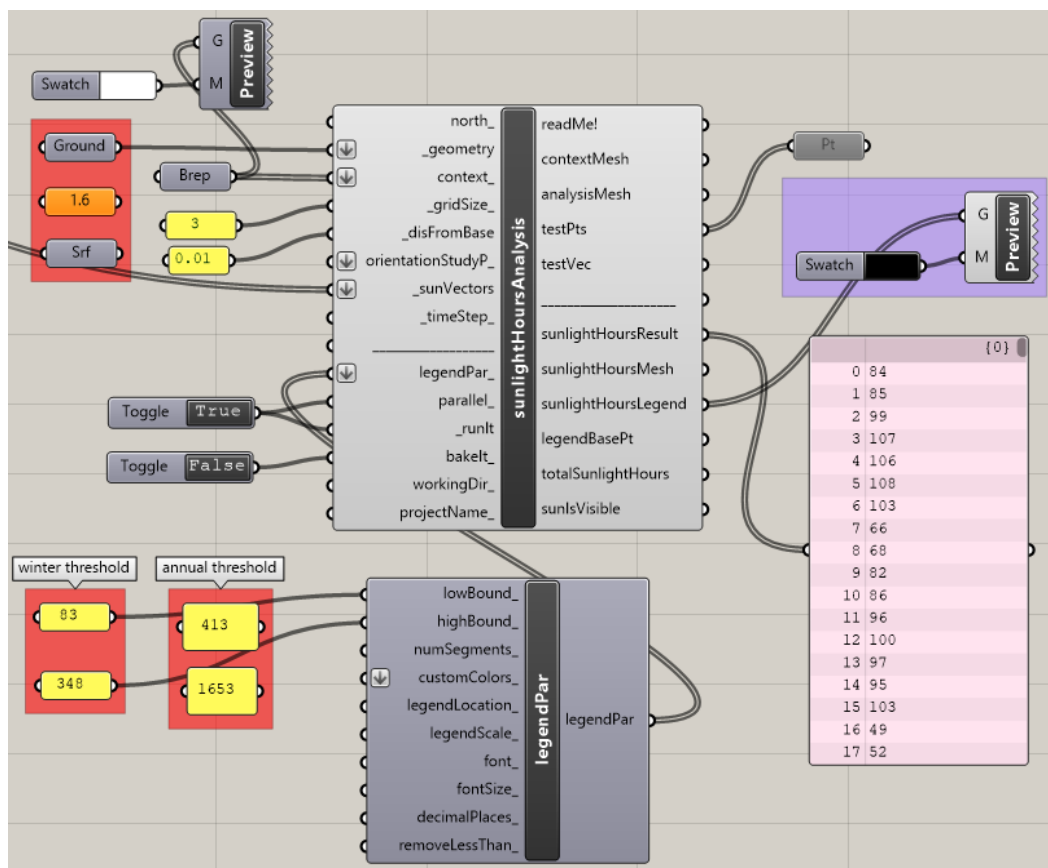
Figure 5.6 Simulation Settings for WPSH

The data list includes the grid point data of the aforementioned 6 simulation combinations, together with 3 sets of data at points on the façade band under APSH, WPSH and SHW settings, exported to SPSS 22 for statistical analysis and multi-linear regression (Figure 5.7).
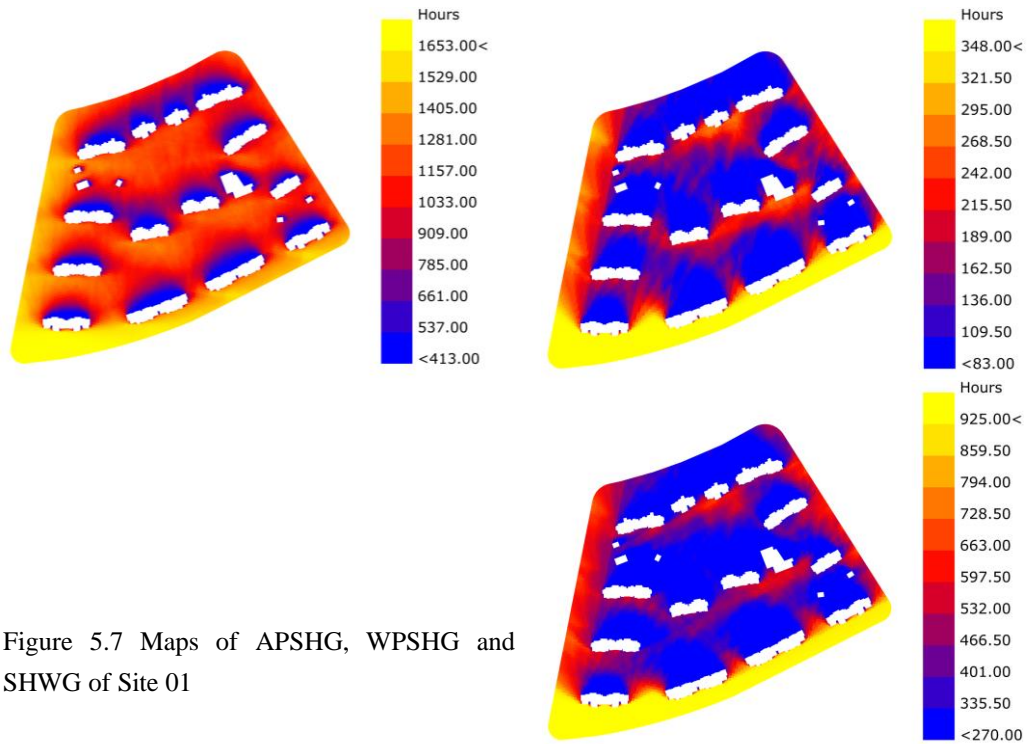
Figure 5.7 Maps of APSHG, WPSHG and SHWG of Site 01

5.4 Selection of Sunlight Availability Performance Indices for Assessment and Analysis

5.4.1 APSHG VS APSHF

APSH is simulated on the grid at both ground level (0m) and ground floor window centroid height (1.6m). Conventionally, APSH is calculated at ground level, but here the window centroid is also applied as the test point, in reference of sunlight and daylight level prediction method defined in Chinese and British design regulation and guidance: in Chinese residential design regulation, the test point is set at the ground floor window centroid of the main bedroom, to enable the sunshine hours collected the façade to predict the indoor sunlight potential. The same applies to daylight level calculated at the window centroid as a prediction of indoor daylight potential.

Taking site 05 and site 10 as examples, a rank-sum test is executed to examine whether the two sets of data, APSHG and APSHF, have significant differences or are identical. The test is carried out using APSH data from site 05 and site 10, which are representative sites of sunlight performance from each of their groups with U-shape

and W-shape histograms of data distribution. With the help of the Wilcoxon test executed in SPSS 22, the 2-tailed asymptotical significance is less than 0.05 from table 5.1, which indicates that the APSH value calculated at ground level and at 1.6m on the façade has a significant difference from a statistical viewpoint. The extent of this difference will be discussed in the following analysis.

Table 5.1 Result of Wilcoxon Test of APSHG and APSHF on S05, S10

Test Statistics[a]

	APSHS05F - APSHS05G
Z	-126.195[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Test Statistics[a]

	APSHS10F - APSHS10G
Z	-160.686[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

In the sunlight hour simulation of APSH for one particular site, all pre-set parameters including site location, available sun vector definition, context buildings, calculation period and 3m grid resolution are identical for ground level grid and 1.6m grid calculations. The only difference between the two types is the height of the calculation grid. Height difference will cause three types of discrepancies between simulation data on ground level and a 1.6m grid: rough consistency at the sunlight blocked section, fluctuation at the partly blocked section and a right skew on the histogram at the unblocked section. For brevity reasons, in this research the APSH value calculated on the ground level grid will be named APSHG. The same applies to APSHF for APSH on 1.6m grid at façade height.

At a particular location, due to the sun azimuth and elevation angle, there must be areas covered by shadows cast by obstacles blocking sun vectors. In this research, the area heavily shaded all year round behind the obstacle building is named as the shaded section of the site. Similarly, the partly shaded section and unshaded section refers to the area covered in shadow in the winter period and partly covered or

un-covered in seasons other than winter, and the area between sun vector incidence direction and building south-facing façade, respectively.

As seen in figure 5.8, when raising the calculation grid, due to the angle of sun vector, the shaded area and partly shaded area will decrease with the height of the calculation grid. Additionally, the gradient or changing rate of the partly shaded area is higher than the shaded area. The area of the un-shaded area is not affected by changing the height of the grid.
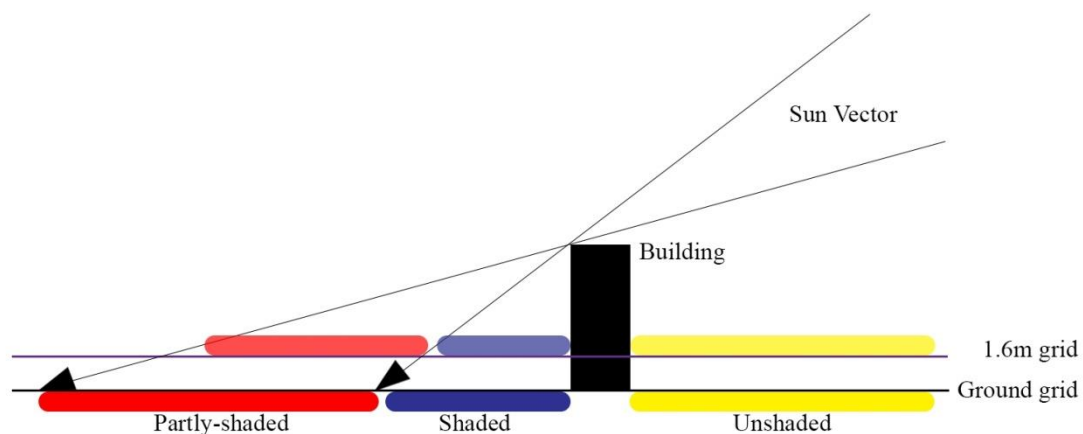
Figure 5.8 Analysis Sketch of APSH at Ground level and 1.6m Height

The combined histograms of APSH data on ground and 1.6m grid for site 05 and site 10 are exported as figure 5.9. The histogram shows the distribution of APSH data, with the value of accumulative sunlight hours on x-axis, and the frequency of the certain sunlight hour value on y-axis. The green line refers to the distribution of data on ground level while the blue line refers to data from the 1.6m grid. It is clear that at the section of lower APSH value (<250h), the blue line acts accordingly with the green. This echoes the fact that the year-round shaded area is barely affected by raising the calculation grid.

The section between 250h to 500h shows consistency in the changing trend between the two lines, yet with a small discrepancy as the blue line dives slightly deeper and has a tendency towards a right skew. This reflects the shrinkage of the shaded area when raising the calculation grid. The grid point at the same location of edges between the shaded and partly shaded areas will acquire larger sunlight hours on the

1.6m grid, which means the frequency of lower value (250-400h) decreases, and the frequency of relatively high value (400-600h) increases. This will show on the histogram as a right skew of the blue line to the green.
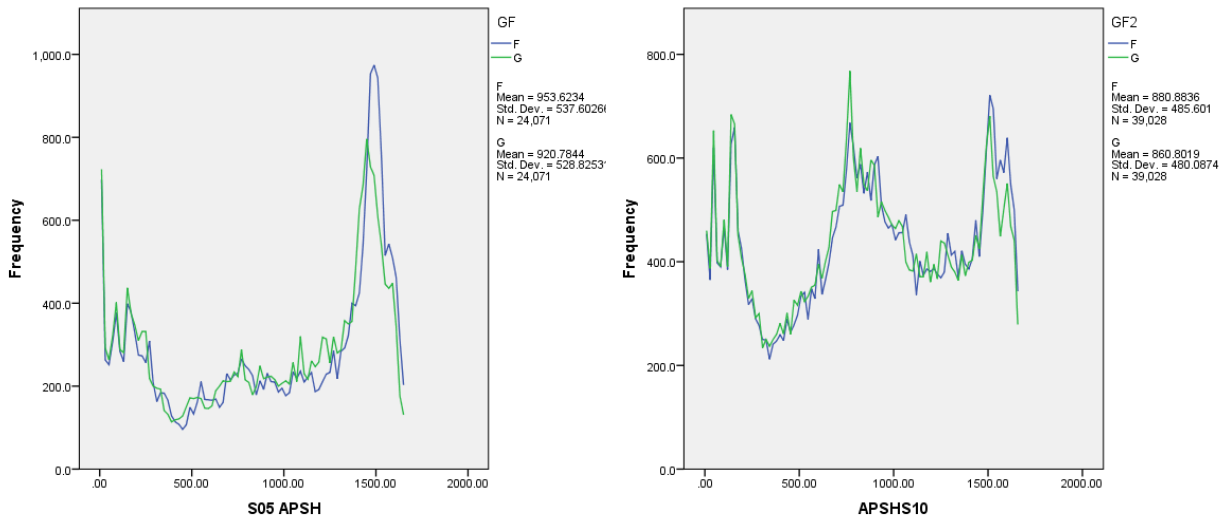
Figure 5.9 Curve Charts of APSHG and APSHF Histogram from S05 and S10

The partly shaded section (500-1400h) shows a similar trend on the whole but with fluctuations. This is in accordance with the complexity of APSH value changing in the partly shaded area. No straightforward changing trend could be seen on the partly shaded area because the changes are under the effect of multiple factors, i.e., sunlight incidence angle, building obstacles, etc., compared to the clearly seen shrinkage of the shadow area under the effect of grid height.

The unshaded area with a value larger than 1400h on the histogram shows obvious similarities in the pattern with a right skew of the blue line. On the 1.6m grid, many grid points on the edge of the partly shaded and unshaded areas tend to have higher value than those on the ground level grid. This contributes to the peak of the blue line which is close to the right of the green peak. The blue peak is both further to the right position and is higher than the green peak, which means the APSH value of blue peak is larger than the green. The number counted in the blue peak is also more than the green. This is because a considerable number of points at the edges of the partly

shaded and unshaded areas, upgrade into unshaded area via acquiring a higher sunlight hour and join to the group of higher value on the histogram.

In the last decreasing part of the histogram of the unshaded area, the blue is almost overlapping with the green line with a slightly higher frequency. This refers to the fact that the site area between the sunlight vector incidence direction and the building's south-facing façades are absolutely unshaded, and has received no influence from the changing of the calculation grid. Only on the southeast and southwest sides of the buildings on the front row facing in a south direction, a larger area totally unshaded will exist on a 1.6m grid compared with the ground. This results in the counts of maximum APSH on blue line becoming higher than the green.

A distribution difference comparison between APSH on the ground and a 1.6m grid is made as in figure 5.10. The nine percentiles (P10-P90) of APSHG and APSHF are calculated, as well as the minimum and maximum. The x-axis indicates the categories of statistic measures and the y-axis indicates the APSH value of each of the measures. The lower green line shows the value of APSHG, while the blue line above it shows that of APSHF. The difference area is shown in the figure. According to it, it could be said that the APSHG and APSHF generally share the same changing trend, while APSHF is a little higher than APSHG in all ranges. A larger difference is shown in categories over P50 and the peak difference is approximately at P60.

Through this comparison it could be said that APSHF is different from APSHG with slight variations in value, but the trend is identical. Because APSHG is a conventional measure of sunlight potential, to make the generalisation of the research result easier, APSHG is selected for further detailed discussion about sunlight hour and building distribution.
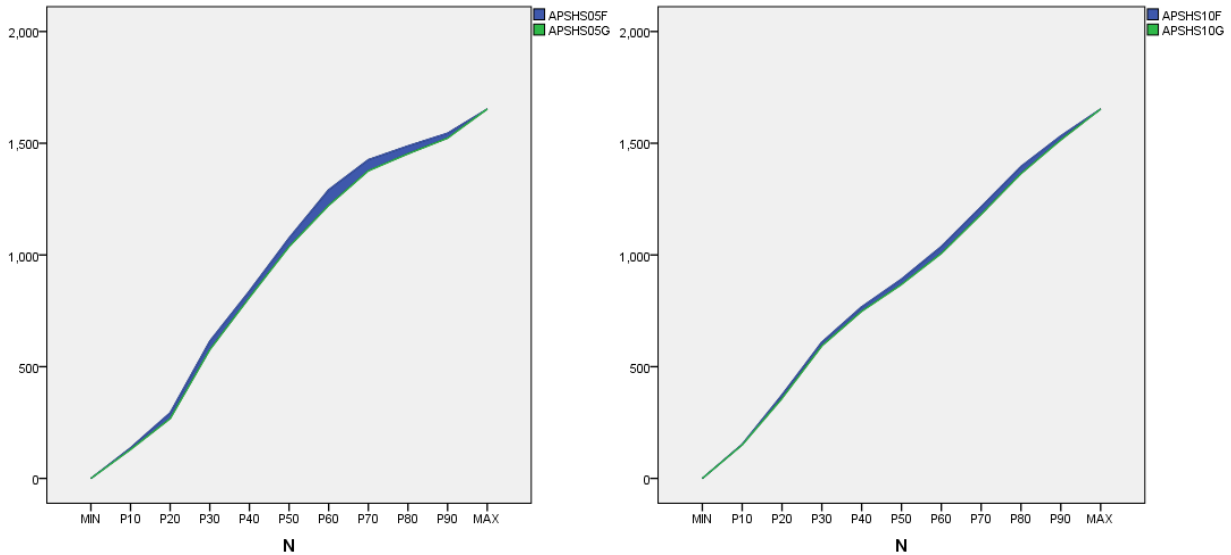
Figure 5.10 Area Difference of Curve by P10-P90 in S05 and S10

5.4.2 APSHG VS WPSHG

APSHG and WPSHG are both sunlight duration index simulated on a ground level grid, but are for year-round period and winter period, respectively. The comparison is aiming to define to what extent APSH and WPSH are different and whether they both should be selected as key indices for regression and optimisation.

Discussed from a time duration aspect, APSHG is simulated for one year, which is 4 times in time length of that for WPSHG. This can be seen in the maximum value of the histograms (Figure 5.11). Maximum value of APSHG is 1653h, while maximum of WPSHG is 348h.

From figure 5.11, a significant high frequency of value close to 0 for WPSHG can be seen. This obvious increase of the proportion of low value in winter sunlight hours, reflects the fact that a large area of a partly shaded section in annual simulation becomes heavily shaded in winter. The shadow becomes larger and longer due to solar azimuth and elevation angle change in winter (Figure 5.12).
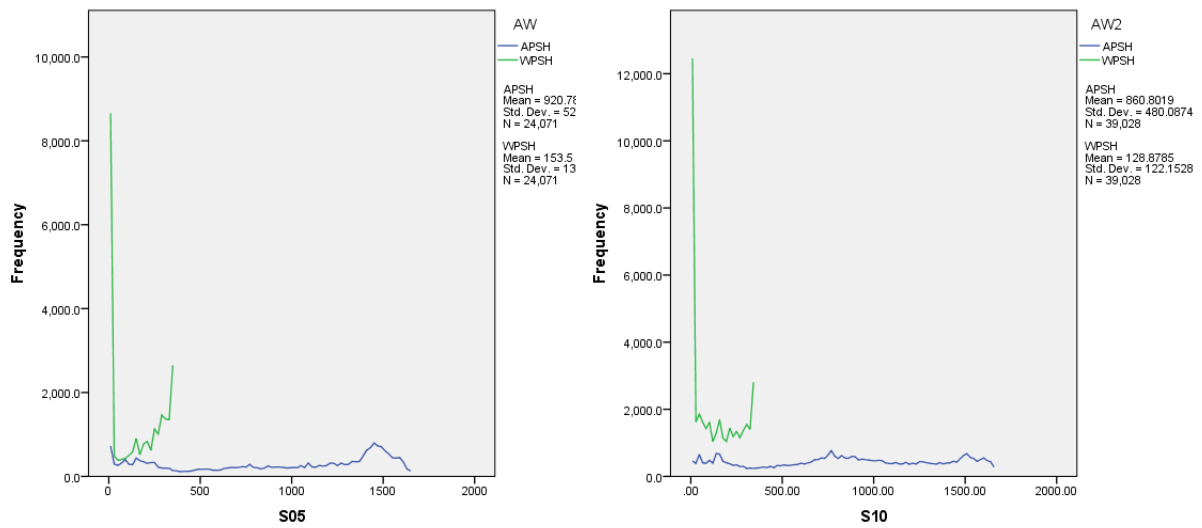
Figure 5.11 Combined Histogram Curve of APSHG and WPSHG in S05, S10
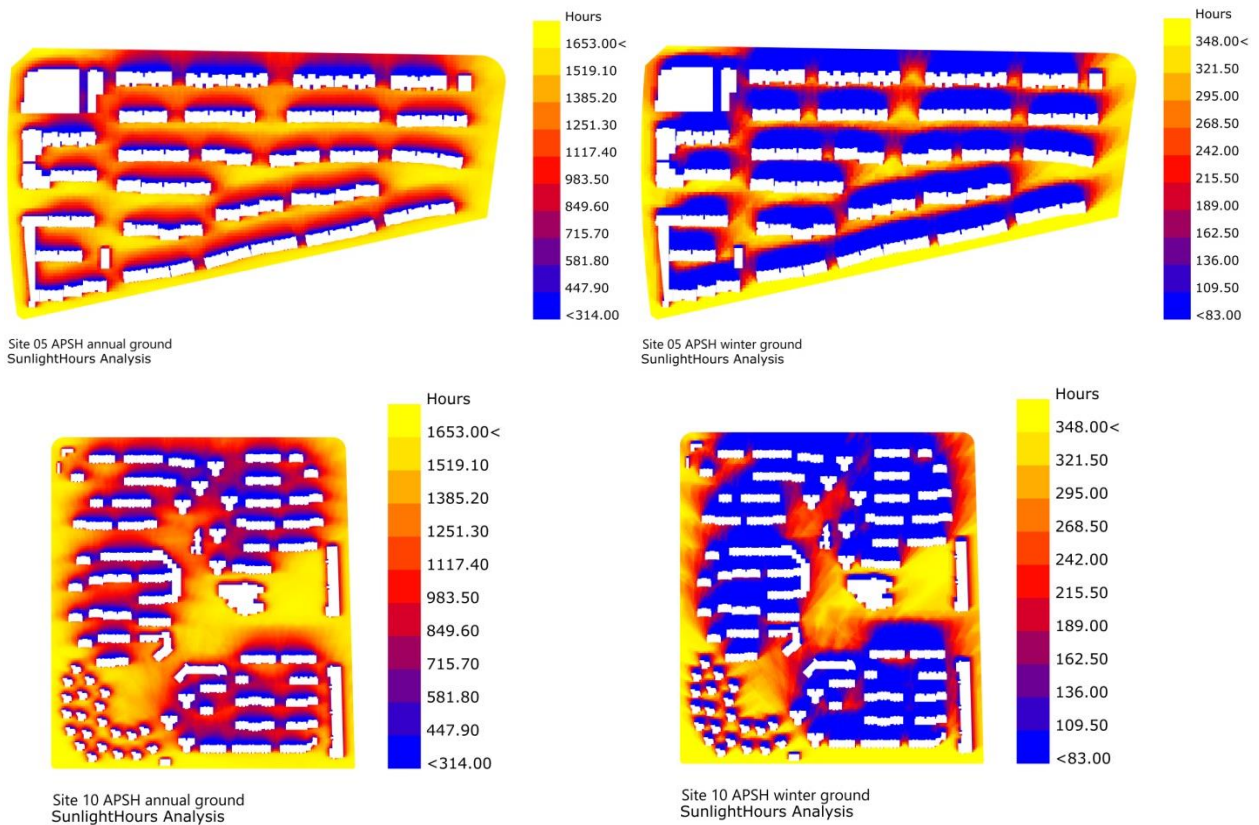
Figure 5.12 Maps of APSHG and WPSHG of S05, S10
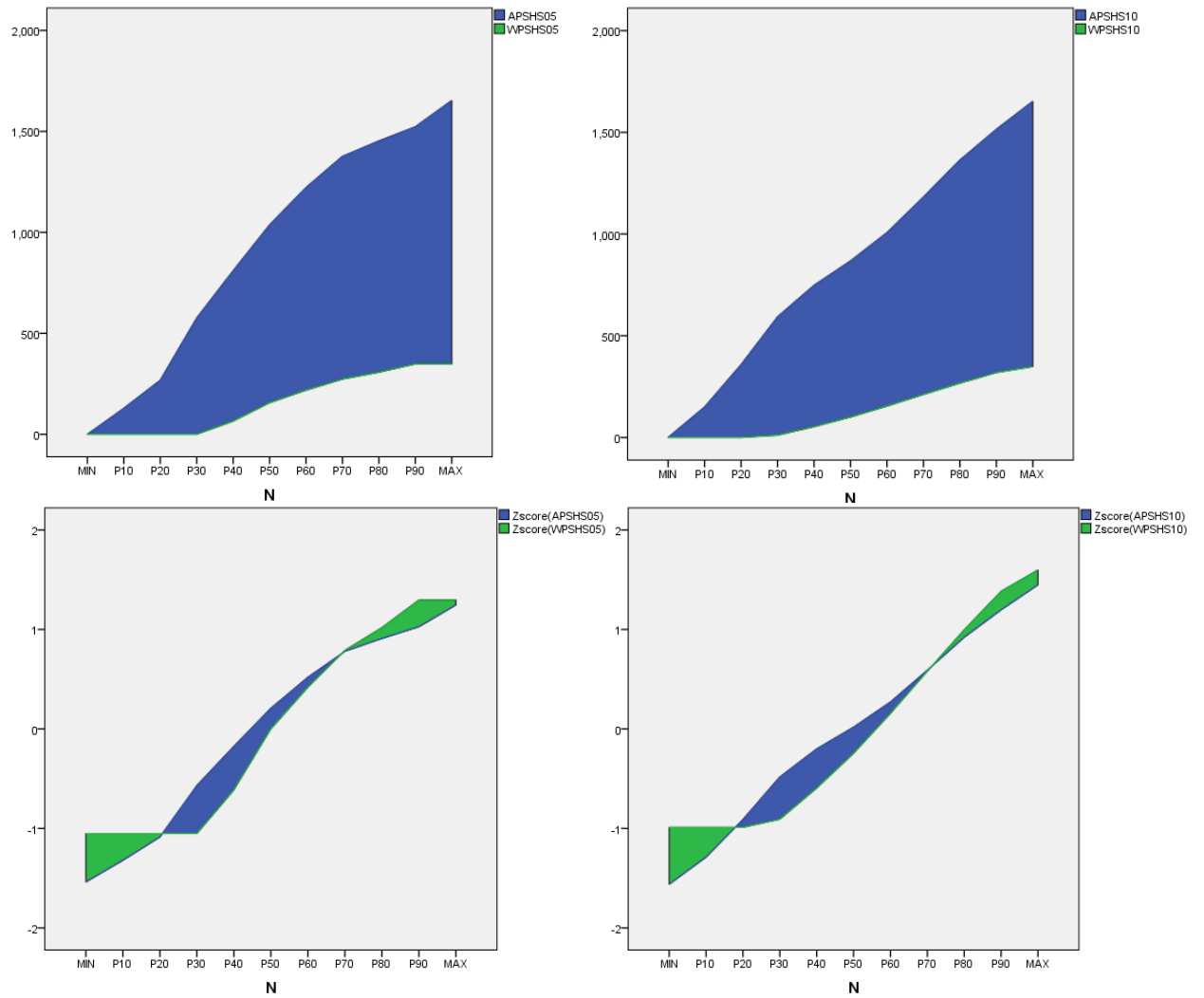
Figure 5.13 Difference Area Chart comparing APSH and WPSH and Normalised APSH and Normalised WPSH

Curves based on minimum P10-P90 and maximum of APSHG and WPSHG is drawn for sites 05 and 10. Due to the huge variation in the range of the two sets of data, a normalisation by z-score is calculated for the following comparison. The area difference between the curve of APSH and WPSH is shown in figure 5.13, as well as for the normalised version.

From the first group comparison, APSH and WPSH show huge differences from P10 to maximum. Obviously, they are very different indices that should both be in use in following regression and optimisation.

To make the conclusion more persuasive, the comparison between normalised APSH and WPSH also shows large discrepancies, with the critical value at P20 and P70. Normalised WPSH (Z-WPSH) has a smaller range, yet with a flat start and a steep changing rate after P30. However, normalised APSH (Z-APSH) has even, almost linear changing rate. This reflects that, for WPSH, the blocked area and un-blocked area have a clearer boundary; the area is either blocked or un-blocked. The partly blocked band does not make up a large ratio on the whole site. However, for APSH, there is a wide band of partly blocked area in between the blocked and un-blocked areas. This makes the APSH value change in a near-linear, gradual manner.

To acquire a quantitative impression of the difference between APSHG and WPSHG, the ratio of grid points with a value less than the critical threshold is calculated as seen in table 6.2. In WPSHG of site 05, the percentage of the heavily shaded area where violates minimum requirement of 83h in Hefei accounts for 41.51%, compared to 6.29% of that for APSHG of the same site. The same is shown in APSH and WPSH of site 10; the ratios of less than 413h year-round and 83h in winter are 5.88% and 46.65%, respectively.

Sunlight hours equal to 0 means thorough cover by shadow in the whole period. Although, as shown in table 5.2, the ratio of 0h sunlight of site 05 is 32% higher compared to site 10 which are similar in the winter period. A huge difference is shown in the ratio of 0h sunlight between site 05 and site 10 in annual simulation results. 0h-Ratio of S05 is 149% higher than that of S10.

Table 5.2 Comparison of Ratio below Threshold for APSHG and WPSHG in S05, S10

		S05		S10	
		APSH	WPSH	APSH	WPSH
COUNTS	ALL	24071	24071	39028	39028
	<83h	1665	9993	2293	18208
	=0h	379	7742	247	9540
RATIO%	<83h	6.92	41.51	5.88	46.65
	=0h	1.57	32.16	0.63	24.44

All of the above show the vital contribution of site distribution difference on the accumulative sunlight hour variation, indicated by APSHG and WPSHG in parallel and simultaneously. Therefore, it is necessary to keep both APSHG and WPSHG for next stage regression and optimisation.

5.4.3   WPSH VS SHW

The aim of this section is to examine to what extent SHW is different to WPSH and whether it is necessary to use both indices in the next stage of research.

As mentioned in the methodology, WPSH and SHW represent both sunlight hour duration simulated in the winter period, but under different criteria. The total available sunlight hours applied in criteria for WPSH and SHW are 348h and 925h, respectively. The reason for the large discrepancy between them is that direct solar radiation is considered as a filter for effective sunlight hours in defining total available sunlight for WPSH but not for SHW. The available sunlight hour and the required sunlight hour for WPSH are both total available daylight hours with the condition of direct solar radiation over 120W/m^2 which refers to effective sunlight hours. However, the available sunlight hour applied for SHW is actually the total available daylight hours, which is much longer than that applied in WPSH simulation. The criterion for SHW of 3h is limited to effective sunlight hours by limiting the simulation hours between 8:00-16:00.

The disadvantage of this rigid time-limit to acquire effective sunlight hour is that only the twilight and dawn period is deducted from the total available daylight hours. Weather variation could also have influence over the effectiveness of sunlight hours. The SHW simulation has not considered climate conditions in winter as in WPSH. They use the same climate-based sun path but WPSH allows the deduction of overcast and rainy hours, when effective sunlight hours could not be produced, from the counts

of total available sunlight hours. Limiting time to 8:00-16:00 cannot exclude weather impact.

Relatively speaking, using solar radiation as an exclusion method to acquire effective sunlight hours is more holistic compared to limiting the time period.
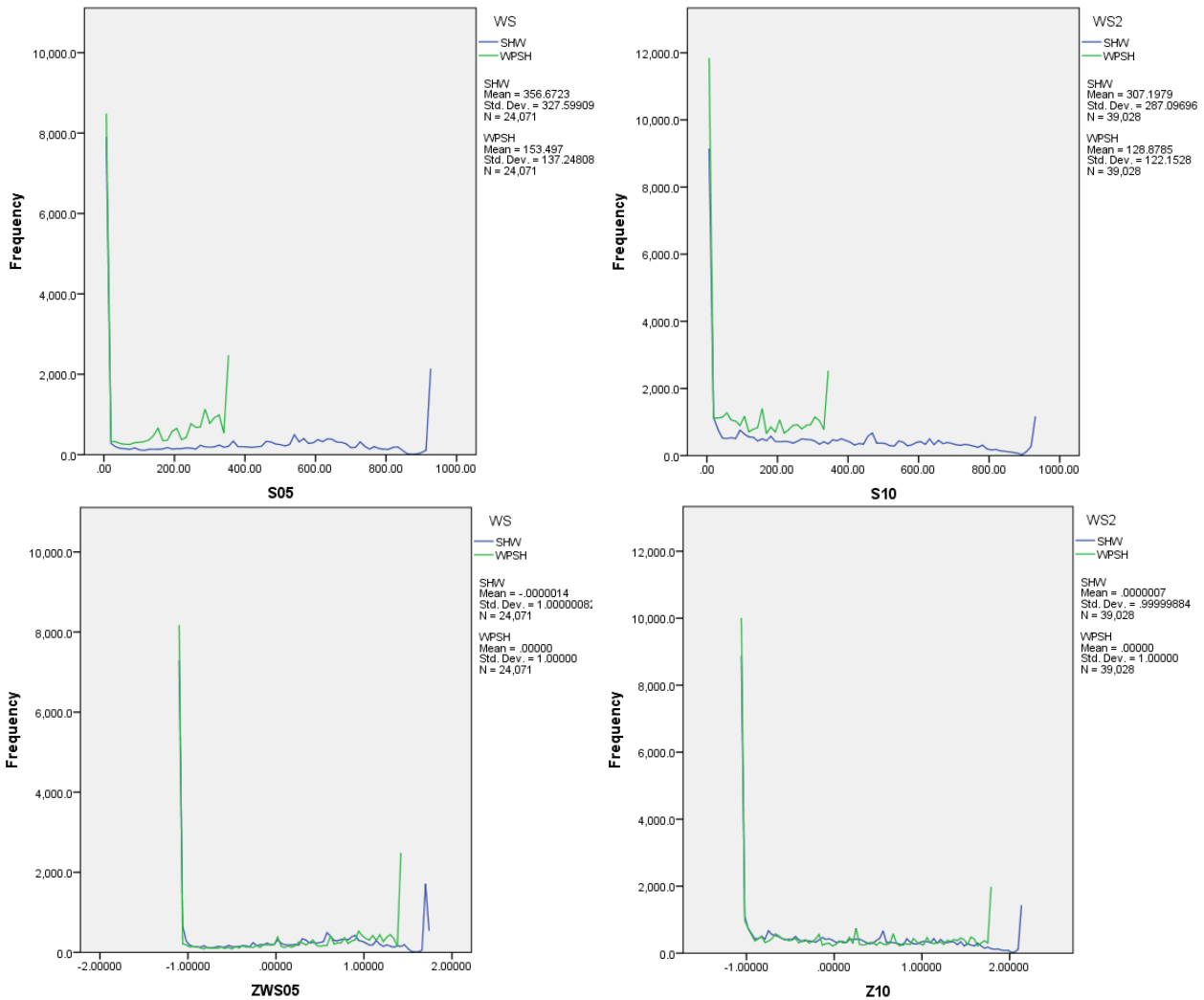
Figure 5.14 Comparison Histogram Curves of WPSH and SHW, and Z-WPSH and Z-SHW of S05, S10

From an aspect of easy application, SH on a single day is less time consuming than executing a sunshine duration test through the winter season, but testing on single day may miss out other sunlight blockage scenarios due to shadow movement of adjacent buildings through winter, which is very influential for sunlight conditions around high-rise residential buildings. It is suggested to improve the national standard by

amending the sunlight criteria from location based day assessment to climate cased period assessment in Chinese design regulation. Climate based period assessment of sunlight provides opportunities for the holistic consideration of impact from sun position, climate and shadow casting from adjacent constructions.

From the standardised WPSH and SHW distribution histogram in figure 5.14, the two calculation methods share a general pattern of distribution. High consistency shows in the lower section, namely in the heavily shaded area. In accordance with the aforementioned discussion, SHW tends to over-predict value in the higher section close to the maximum value than does WPSH.
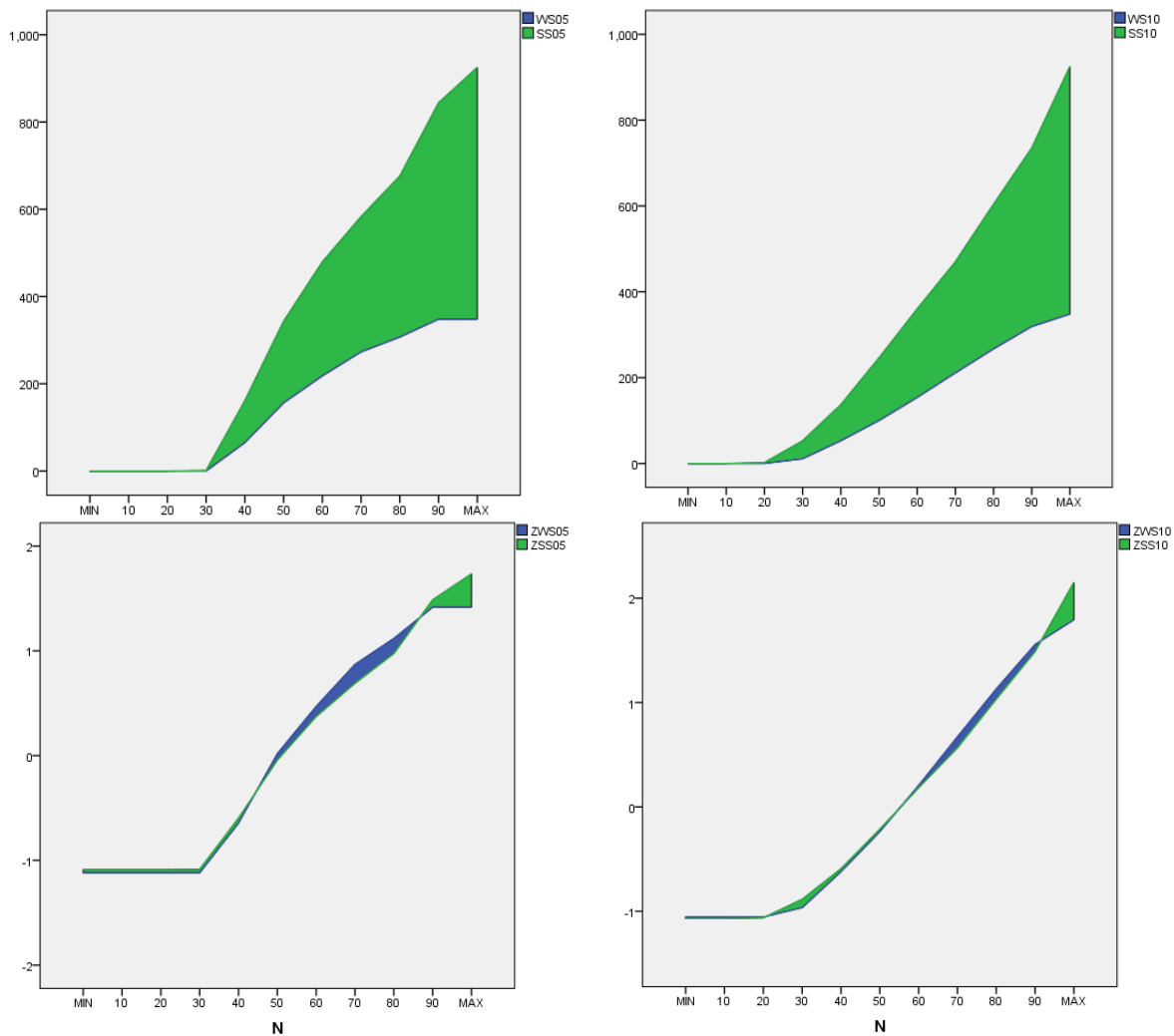
Figure 5.15 Comparison Histogram Curves of WPSH and SHW, and Z-WPSH and Z-SHW of S05

The difference area of WPSH and SHW in figure 5.15 shows that the two indices predict the same for sections below P30. SHW obviously predicts a higher value after P30, and the degree of over prediction becomes greater with the increase of data value.

The difference area of normalised WPSH and SHW shares a high level of consistency in the simulation result below P50, as shown in figure 6.14. Between P50 and P90, WPSH tends to predict higher than SHW, while SHW over-predicts in the section larger than P90. Based on this, it could be suggested that WPSH and SHW have very similar calculation methods and prediction ability. SHW is not as considerable as WPSH when considering maximum available sunlight hours, and WPSH is suggested for further regression and optimisation.

As shown in table 5.3, SHW offers tighter criteria than WPSH as a higher percentage of shadow area violates the SHW minimum threshold than that of WPSH in the same case, under the same conditions. For site 05, SHW is 3.68% tighter than WPSH; for site 10 SHW is 5.32% tighter than WPSH. This demonstrates that SHW criterion is expanded from an one day criteria in to a three month criterion, which is a little over constrained.

Table 5.3 Comparison of Ratio below Threshold for WPSH and SHW in S05, S10

	S05		S10	
	WPSH	SHW	WPSH	SHW
Threshold (h)	83	270	83	270
total counts	24072	24072	39029	39029
Below Threshold Counts	9993	10878	18208	20282
Below Threshold Ratio (%)	41.51	45.19	46.65	51.97

Overall, WPSH and SHW have no essential distinction, and WPSH enables more consideration. So WPSHG is applied for further regression and optimisation.

5.4.4   Summary

Through studies in this section, it can be noted that APSHG has no essential distinction to APSHF, but has a slight variation of degree in reflecting sunlight availability. As APSHG is one of the conventional indices for outdoor sunlight availability, APSHG is used for further study.

Similarly, the analysis result shows SHW has no essential distinction to WPSHG, and shares a very close trend in the curve chart. However, the definition of SHW does not allow as much consideration as WPSHG, so WPSHG is applied for further study. In addition, an update of the sunlight availability evaluation theory in Chinese design guidance is suggested: from location-based day assessment to climate cased period assessment.

APSHG and WPSHG are tested to be distinct. Both are necessary for assessing sunlight duration. They both indicate the vital influence of building distribution on the accumulative sunlight hour variation from difference aspects and show totally different changing trends. APSHG and WPSHG need to be simultaneously adapted for next stage clustering, regression and optimisation.

5.5 Qualitative Analysis of Clustering by APSH

5.5.1   Method and Discussion of Cluster Site Samples by APSH, WPSH and Their Statistic Measures

A clustering of sites based on sunlight availability indices and their statistic measures is carried out in SPSS. The aim of the clustering is: 1. to discover any possible common characteristics between samples regarding their sunlight conditions, based on excavation of their sunlight hour data; 2. to define any qualitative relationship

between the discovered characteristics and corresponding building distribution based on observation and comparison of the sunlight hour map between and within clusters.

The indices used for clustering are APSHG and WPSHG, namely annual possible sunlight hour on ground and winter possible sunlight hour on ground, respectively. For convenience purposes they will be termed as APSH and WPSH throughout the following section. Characteristics of the two indices are presented in the form of their values from all simulation grid points and the descriptive statistical values of the two indices. The statistical measures are 10 percentiles (P10-P90), 2 quartiles (P25 and P75), minimum and maximum. Therefore, the clustering is attempted 3 times for three combinations of statistical measures for each of APSH and WPSH. The combination of clustering measures is: all grid point values, P10-P90 (including P25 and P75) and P10-P90 with minimum and maximum (including P25 and P75). Hierarchical clustering is applied. Case cluster is applied for the condition of using data on all grid points, while variable cluster is applied for the condition of using descriptive statistical values.

However, the clustering result by all grid point values is highly correlated to the number of grid points, which is an indirect reflection of site size. For example, the clustering result by APSH grid value is three groups according to the number of grid values of over 2000, around 1000 and below 600. This makes the clustering results highly biased by factor of size, so that the effect of other characteristics is totally covered and is not identifiable in clustering results. Similarly, clustering results based on P10-P90 with minimum and maximum is also highly biased by the minimum and maximum. Comparing the clustering result by P10-P90 with minimum and maximum and by P10-P90, the latter combination presents more characteristics of sunlight condition, within and between clustering groups. Therefore, P10-P90 including P25 and P75 is selected as the basis of clustering calculation for both indices APSH and WPSH.

The output of the clustering from SPSS is dentrograms and tables of hierarchical clustering procedures. Groups are defined, referencing the outputs.

5.5.2 Analysis of Clustering of Sites and Discovery of Rules

In this section, cluster output based on statistical measure (P10-P90 including P25 and P75) of APSH is analysed. Characteristics of each cluster group is discussed and explained based on sunlight hour maps and sunlight hour data of sites in a group. After that, Qualitative tendencies and rules between cluster characteristics and corresponding building distribution is discovered and discussed.

Based on dendrogram of clustering by P10-P90 of APSH and physical meaning of each cluster by characteristics shown in sunlight maps, four clusters are ascertained: cluster A of very good sunlight possibility, B of good performance, C of fair performance and D of poor condition (figure 5.16).

The analysis of the clustering discovers the underlining reason of division of the clusters. Building distribution characteristics are extracted from each cluster, answering the question of what building distribution combination leads to the difference between clusters and within clusters.

The sites are clustered into 4 groups. The detail of clusters and derived conclusions are listed in table 5.4 as a summary.
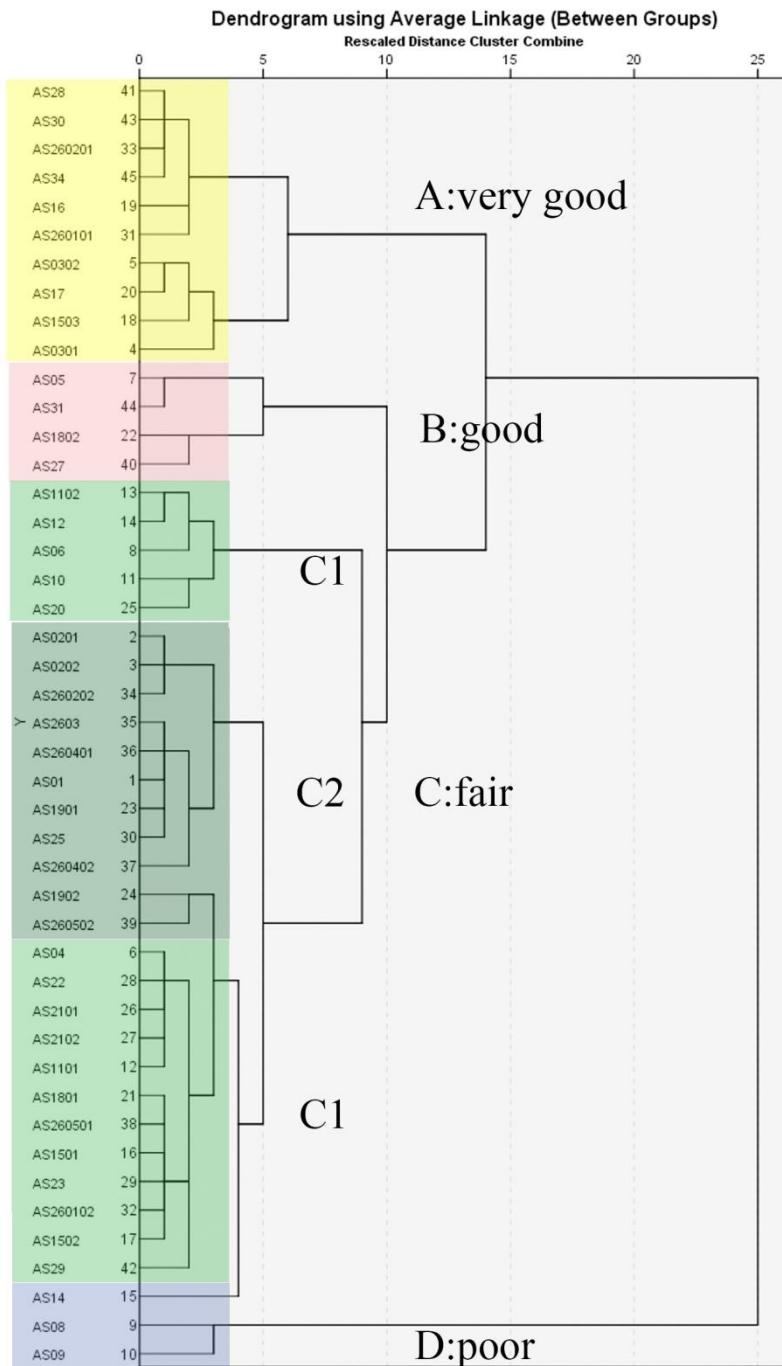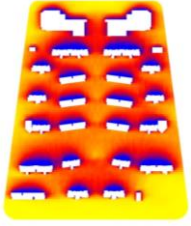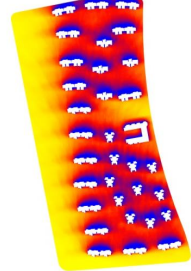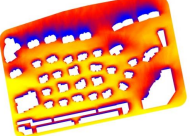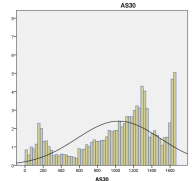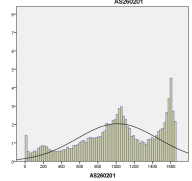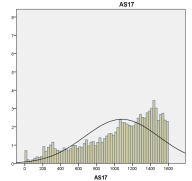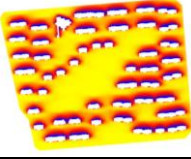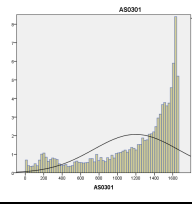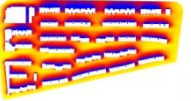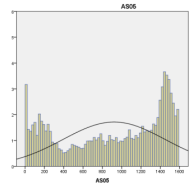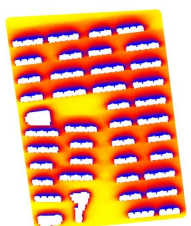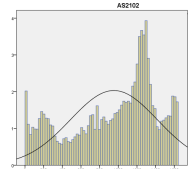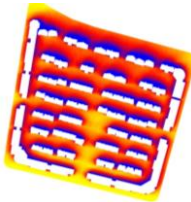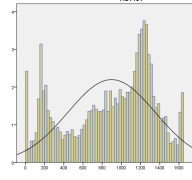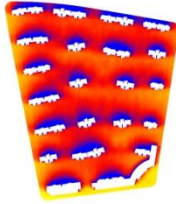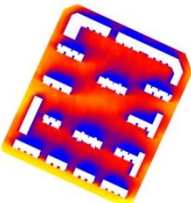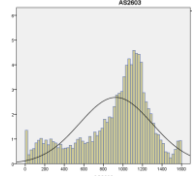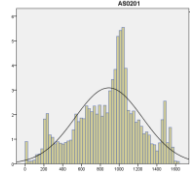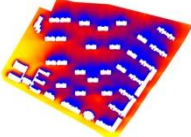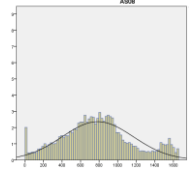
Figure 5.16 Dendrogram of Clustering by APSH P10-P90

Table 5.4 Summary of Clustering Based on P10-P90 of APSH Data

Cluster	Sub-groups	Site Description	Example	Wind map	Histogram	Characteristics	conclusion
A: Very good	A1	1.low/mid-rise,medium-density, on-edge open space; 2 high-rise, low-density, large on-edge open space; 3 mix-rise, low-density	S28 S30 S2602 01 S34 S16 S17 S2601 01	S30 S260201 S17		A1: 3 peaks, heavy right-skewed middle peak, taller maximum peak than minimum peak. A2: 2 peaks, very small minimum peak, very tall maximum peak.	1.low/mid-rise community, low-density and in-site open space, is the most ideal design combination for most extended APSH condition. 2. with open spaces on the edge of site, the low/mid-rise community provides higher overall APSH than high-rise community 3. mix-rise community in low-density with tall buildings on the north, is possible to generate very good sunlight environment
	A2	Low/mid-rise, low-density, in-site open space	S0302 S1503 S0301	S0301			
B: good	N/A	low/mid-rise, medium-density, no open space, long façade	S05 S31 S1802 S27	S05		Two peaks of similar size	medium-density, long façade and no open space leads to dramatic increase of shaded area, as a tall minimum peak with similar size to the maximum peak
C: fair	C1	Low/mid-rise, high-density, in-site open space	S1102 S12 S06 S10 S20 S04 S22	S2102		3 peaks, short minimum and maximum peak, tall right-skewed middle peak	1 for low/mid-rise high-density and high-rise medium-density sites, open space on the edge of site will boost maximum peak in

			S2101 S2102 S1101 S1801 S2605 01 S1501 S23 S2601 02 S1502 S29	S1101			histogram. 2 for low/mid-rise site in high-density, higher enclosure at boundary means higher lower tail of middle peak, between 600-1000h. 3 for high-rise medium-density site, shorter façade length causes smaller blocked area; higher enclosure at boundary, higher minimum peak and between 500-900h; higher building height causes higher minimum peak and less right skew of middle peak
	C2	High-rise, medium-density, no open space	S0201 S0202 S2602 02 S2603 S2604 01 S01 S1901 S25 S2604 02 S1902 S2605 02	S2603 S0201			
D: poor	N/A	High-rise, high-density	S14 S08 S09	S08		3 peaks, left-skewed middle peak	1 Higher enclosure causes further left skew on middle peak and boost of minimum peak. 2 on-edge open space boosts maximum peak

Qualitative relationships between building distribution combination and APSH performance are extracted from the analysis. These rules are consolidated into three levels: macro, meso and micro level. Macro level refers to rules related to macro strategy of site planning; meso level refers to the influence of indices and the priority of index application, and micro level refers to the detailed combination of indices and their effects.

Macro level:

1. To achieve longer overall APSH on a whole site, the prior sequence of building distribution combinations are low/mid-rise low-density, low/mid-rise medium-density or high-rise low-density, low/mid-rise high-density or high-rise medium-density, high-rise high-density.

2 The majority of practices have fair APSH conditions, and have three peaks in the histogram of APSH grid data. For good and fair conditions, the middle peak is right-skewed; for poor conditions, the middle peak is left-skewed.

Meso level:

1. With open spaces on the edge of the site, the low/mid-rise community provides higher overall sunlight hours than the high-rise community. This is especially obvious in the partly shaded area shown as a right-skewed middle peak.

2. A tall maximum peak should not be considered as a sign of good APSH performance. In new designs, a highly right-skewed middle peak and middle peak merges with maximum peak is the suggested pattern of the histogram. In the case of sites that need an improvement in APSH performance, boosting maximum peak by adding on-edge open space is still one of the possible approaches.

Micro level:

1. Best combination: low development intensity including low/mid-rise community, low-density and in-site open space, is the most ideal design combination for most extended APSH level.

2. Good combination: compared to common low/mid-rise and high-rise community, extreme combination of mix-rise community in a low-density with tall buildings on the north is still possible to generate a very good sunlight environment; similar to but not as long APSH as low development intensity combination.

3. Worst combination: high-rise high-density site with no open space leads to most blocked and partly blocked area; shown as tall minimum peak and left-skewed middle peak.

4. Open space on-edge: for low/mid-rise high-density, high-rise medium-density and high-rise high-density site, open space on the edge of the site will boost maximum peak in the histogram. Very large open space in mix-rise sites will have a similar effect.

5. Enclosure: for low/mid-rise site in high-density, an increase of the enclosure at the boundary will cause an increase in the frequency of the lower tail of middle peak, in the range of 600-1000h.

6. Enclosure: for high-rise medium-density site, an increase of enclosure at the boundary will cause an increase in frequency of the minimum peak and the lower tail of middle peak between 500-900h.

7. Enclosure: for high-rise high-density sites, increasing the enclosure at boundary will cause further left skew on the middle peak and boost the height of the minimum peak in the histogram.

8. Façade length and open space: for low/mid-rise medium-density sites, a long façade and no open space in site causes a dramatic increase of the shaded area, expressed as a tall minimum peak with a similar size to the maximum peak.

9. Façade length: high-rise medium-density site, shorter façade length leads to smaller ratio of the blocked area, shown as small minimum peak in histogram.

10. Building height: for high-rise site of medium-density increased building height will cause an increase in the minimum peak and less right skew of the middle peak in the histogram. This corresponds to decreasing of sunlight hours in low and middle level ranges.

Combination of building distribution and their effects mentioned in micro level rules are summarised in table 5.5.

Further discussion about influential indices and less influential indices is expanded. From table 5.5, the morphology parameter of building orientation, homogeneity and site size are not as allied as cluster basis parameters. From the comparison of sample sites between and within clusters, it could be suggested that homogeneity and site size are not influential to APSH performance. Building orientation shows some impact on the APSH behind the building, especially when façade length is long, and boundary is closed, i.e., S08 and S1102. However, building orientation is not significantly influential to the overall APSH condition on the site, but has more of a local impact. The reason for this could be because APSH is an accumulative parameter considering the sun path year-round, which means solar vectors within 180 degrees of due south on both sides are all adapted. The variation of the adapted solar vector buries angle differences of building orientation in the changing trend of APSH. Based on this analysis, building orientation should be more influential in WPSH clustering, as well as wind speed clustering in thermal studies.

Table 5.5 Summary of micro level rules from APSH clustering

Distribution Characteristics		High overall APSH			Histogram appearance		
		Low APSH range	Middle APSH range	High APSH range	Two peaks	Three peaks	
						right-skewed middle peak	left-skewed middle peak
Size	Small site						
	Large site						
Homogeneity	Even distribution						
	Clustered distribution						
Density	Dense separation	7	5, 7,	4, 4"		4, 4", 5,	3, 7
	Medium separation	8, 9,10	6, 10	4'	8,	4', 6, 8, 9, 10	
	Loose separation	1, 2,	1, 2,	1, 2,	1, 2,		
Open space existence	No open space						3
	Open space inside site	1,	1,	1,	1,		
	Open space on dominant wind direction						
	Open space on edge			4		4	
Boundary enclosure	Open boundary						
	Less enclosed	7	5, 6, 7,			5, 6,	7,
	Highly enclosed						
Façade length	Long façade						
	Short façade	8, 9			8,	8, 9,	
Building orientation	Perpendicular to dominant wind						
	Parallel to dominant wind						
Building height	Low/mid-rise	1, 8,	1, 5,	1, 4	1, 8,	4, 5, 8,	
	mix-rise	2,	2,	2,	2,		
	High-rise	7, 9,10	6, 7, 10	4',4",		4', 4", 6, 9, 10	3, 7,
	Less building height	10	10			10	
Key building location	High-rise on the north	2,	2,	2,	2,		

5.6 Qualitative Analysis of Clustering by WPSH
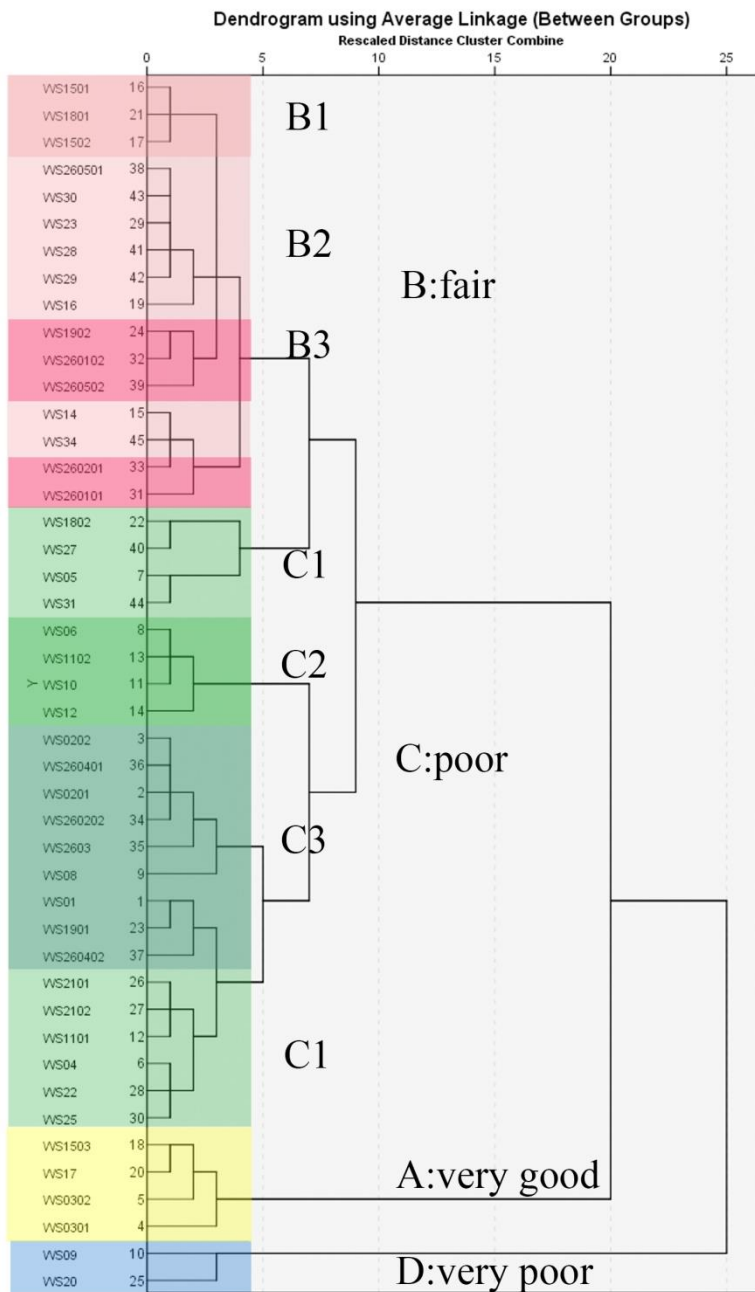

Figure 5.17 Clustering Result by P10-P90 of WPSH
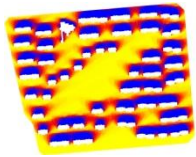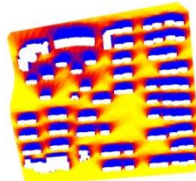
Using a similar research method and procedure, clustering of sample sites based on WPSH statistical measures is carried out. The statistical measures of WPSH are nine percentiles of P10-P90 and two quartiles of P25 and P75. The group results are shown in Figure 5.17. Four Clusters are generated, A for very good WPSH conditions, B for

fair conditions, C for poor conditions and D for very poor conditions. Sub-groups of each cluster are explained in the following sections.
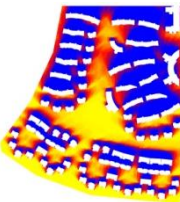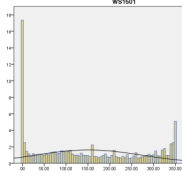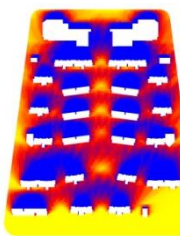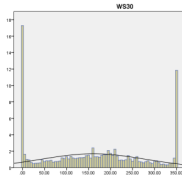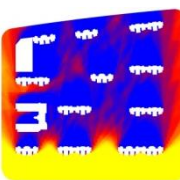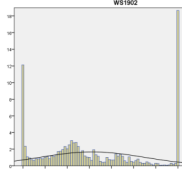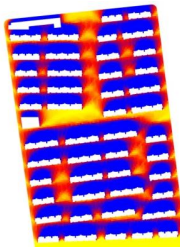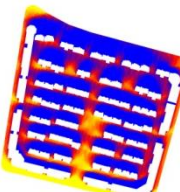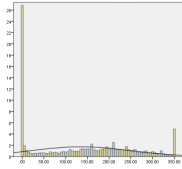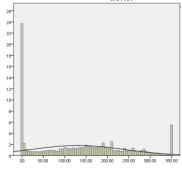
In this section, sites are clustered based on statistical measures of WPSH data. The statistical measures are ten percentiles and four quartiles. Four clusters are generated for very good, fair, poor and very poor WPSH performance. Details of each cluster are summarised in table 5.6, together with the conclusions of analyses.

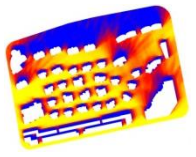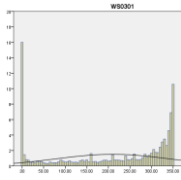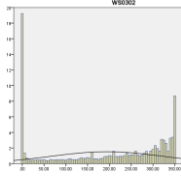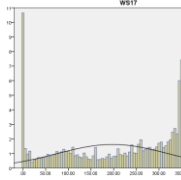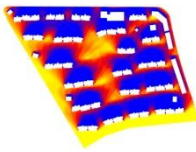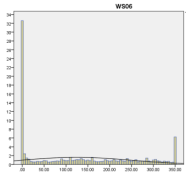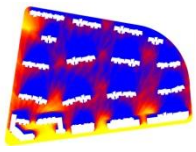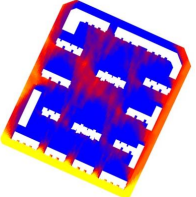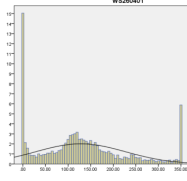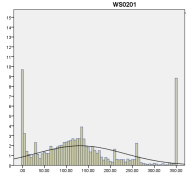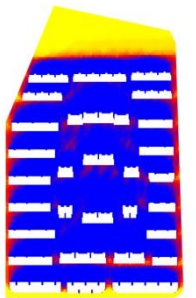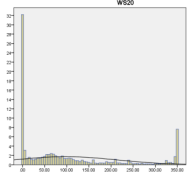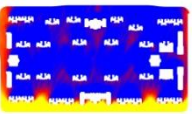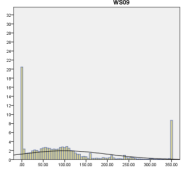The rules discovered from the analysis between and within clusters are categorised into three levels: macro – referring to the overall design strategy; meso - comparison between building distribution parameters; and micro - referring to detailed variation trends between WPSH performance and distribution parameters.

Table 5.6 summary of Clustering based on P10-P90 of WPSH

Cluster	Sub-groups	Site Description	Example	Wind map	Histogram	Characteristics	conclusion
A: Very good	N/A	Low/mid-rise, Low-density, Uneven, In-site Open Space	S1503 S17 S0302 S0301	S0301 S0302 S17		1. Two peaks 2. Minimum peak height 10-15% 3. Maximum peak height 10% 4. middle peak axis over 300h	in low/mid-rise site of low-density 1. increasing building height increases minimum peak height. 2. increasing in-site open space increases maximum peak height. 3. mix high-rise buildings in causes decreasing in height of minimum peak, but generate a left-skewed middle peak. 4. low-rise buildings, low-density, in-site open space, uneven is best combination

B: Fair	B1	Low/mid-rise, Mix Density, Uneven, In-site Open Space	S1501 S1801 S1502	S1501		1. Minimum peak height 15-20%, 2. maximum peak height 5%, 3. middle peak location 100h	1. Evenly distributed, medium-density site could provide higher middle peak location than uneven, mix density sites. 2. For low/mid-rise site and high-rise site in medium-density, larger on-edge open space on south leading to taller maximum peak. 3. for low/mid-rise site in medium-density, the enclosure at boundary leading to further left skew of middle peak. 4. High-rise building in medium-density leads to taller and more left-skewed middle peak, but shorter minimum peak, compared to low/mid-rise buildings.
	B2	Low/mid-rise, Medium-density, Even, South On-edge Open Space	S2605 01 S30 S23 S28 S29 S16 S14 S34	S30		1. Minimum peak height 15-20%, 2. maximum peak height 10%, 3. middle peak location 150-200h	
	B3	High-rise, Medium-density, Uneven, South On-edge Open Space	S1902 S2601 02 S2605 02 S2602 01 S2601 01	S1902		1. Minimum peak height 10-15%, 2. maximum peak height 20%, 3. middle peak location 100h	
C: Poor	C1	Low/mid-rise, high-density, in-site open space	S1802 S27 S05 S31 S2101 S2102 S1101 S04 S22 S25	S2101 S1101		1. Minimum peak height 5-10%, 2. maximum peak height 20-25%, 3. middle peak location 150-200h	1. For low/mid-rise high-density enclosed site, it tends to have tall minimum peak and very left-skewed middle peak 2. For low/mid-rise high-density site, increasing enclosure leads to further left skew in middle peak. 3. For low/mid-rise high-density site, increasing building height leads to

	C2	mid-rise buildings, high-density, enclosure in-site open space	S06 S1102 S10 S12	S06		1. Minimum peak height 5-10%, 2. maximum peak height 30-40%, 3. middle peak location 150h	increase in minimum peak. 4. In high-rise high-density site, on-edge open space on the south boosts the height of maximum peak. 5. Under medium-density condition, compared to low/mid-rise buildings, high-rise sites leads to shorter minimum peak, but lower location of middle peak axis. Namely, high-rise sites tend to have less ratio of totally blocked area, but higher ratio of partly shaded area.
	C3	High-rise Medium-density, Even, Enclosed On-edge Open Space on South	S0202 S260401 S0201 S260202 S2603 S08 S01 S1901 S260402	S260401 S0201		1. Minimum peak height 10-15%, 2. maximum peak height 5-15%, 3. middle peak location 100-150h	
D: Very poor	D1	Low/mid-rise, high-density, Even, Enclosed, On-edge Open Space	S20	S20		1. Minimum peak height 20-30%, 2. maximum peak height 10%, 3. middle peak location below 100h	1. High-rise or low-rise buildings combined with high-density, enclosure and evenness leads to worst WPSH condition. 2. For high-density site, open space in other direction does not lead to tall maximum peak as space on the south do
	D2	High-rise, high-density, Even, Enclosed, On-edge Open Space	S09	S09			

Macro level:

1. To achieve longer overall WPSH on the whole site, the prior sequence of building distribution combinations are low/mid-rise low-density, low/mid-rise medium-density or high-rise low-density, low/mid-rise high-density or high-rise medium-density, high-rise high-density.

2 The majority of practices have poor WPSH conditions, and have three peaks on the histogram of WPSH grid data. Only in the case of the very good cluster, the middle peak is right-skewed or merged into a maximum peak. All other clusters have a left-skewed middle peak.

Meso level:

1. Evenly distributed, medium-density sites could provide higher middle peak location than uneven, mixed density sites.

2. The sequence of contribution to the height of maximum peak from the various types of open spaces is (high to low): on-edge on south, on-edge on other direction, large in-site open space, small in-site open space.

3. Under medium-density conditions, compared to low/mid-rise buildings, high-rise sites leads to a shorter minimum peak but a lower location of the middle peak axis. Namely, high-rise sites tend to have less ratio of a totally blocked area, but a higher ratio of a partly shaded area.

Micro level:

1. Best combination: low-rise buildings, low-density, unevenness, and in-site open space, are the best combination of distribution attributes for extended overall WPSH.

2. Worst combination: no matter whether high-rise or low-rise buildings when

combined with attributes of high-density, enclosed evenness leads to worst WPSH condition.

3. Building height: in low/mid-rise site of low-density, decreasing building height will decreasing minimum peak height.

4. Building height: for low/mid-rise high-density sites, decreasing building height leads to significant decrease in the minimum peak.

5. High-rise buildings: in low/mid-rise sites of low-density, integrated high-rise buildings will cause significant decreases in minimum peak, but will generate a left-skewed middle peak.

6. Enclosure: for low/mid-rise sites in medium-density and high-density, the enclosure at the boundary leads to further left skew of the middle peak.

7. Enclosure: for low/mid-rise high-density enclosed sites, there tends to be a tall minimum peak and a very left-skewed middle peak

8. In-site open space: in low/mid-rise sites of low-density, increasing in-site open space will increase maximum peak height.

9. On-edge open space: for low/mid-rise and high-rise sites in medium-density, and high-rise sites in high-density, larger on-edge open space in the south leads to taller maximum peak.

Combinations of distribution attributes mentioned in micro level rules are listed in table 5.7.

Table 5.7 Summary of Micro Rule from WPSH Clustering

		High overall WPSH			Histogram appearance		
		Low WPSH range	Middle WPSH range	High WPSH range	Two peaks	Three peaks	
						right-skewed middle peak	left-skewed middle peak
Size	Small site						
	Large site						
Homogeneity	Even distribution						
	Clustered distribution	1	1	1			
Density	Dense separation	4,		9			6, 7
	Medium separation			9			6
	Loose separation	1, 3	1	1, 8			
Open space existence	No open space						
	Open space inside site	1	1	1, 8			
	Open space on dominant wind direction						
	Open space on edge			9			
Boundary enclosure	Open boundary						
	Less enclosed						
	Highly enclosed						6, 7
Façade length	Long façade						
	Short façade						
Building orientation	Perpendicular to dominant wind						
	Parallel to dominant wind						
Building height	Low/mid-rise	1, 3, 4,	1	1, 8, 9			6, 7
	mix-rise						
	High-rise			9			6
	Less building height	3, 4					
Key building location	High-rise on the north						

5.7 Comparison and Discussion on Grade Differences of Clustering Result Based on APSH and WPSH

Based on the APSH and WPSH clustering results, similarities and discrepancies are shown between the clusters. The APSH sites are clustered into four performance groups of very good, good, fair and poor performance. WPSH also generates four performance groups of very good, fair, poor and very poor performance. Comparing the content of each group, some bundles sites remain in the same level cluster, while others are graded down to a lower-level cluster. Few cases upgrade to a higher-level cluster. The majority of sites are assessed in the 'fair' group by APSH, and downgraded into the poor group when assessed by WPSH. The overall difference from APSH to WPSH clusters results in sites downgrading from a higher level to a lower level. This is especially obvious in sites with high-rise buildings. Sites with high-rise buildings may perform as 'good' in APSH assessment, but perform as 'very poor' in WPSH assessment. This shows that a high-rise building has an influential factor on WPSH performance.

As shown in figure 5.18, four sub-groups in APSH clustering downgrade in WPSH clustering, two remain in the same level clusters, and one case upgrades to a higher-level cluster.

In APSH clustering, part of cluster A (very good) remains in 'very good' in WPSH, while the other part downgrades to 'fair' in WPSH. Subgroup A2 (40%) from A has distribution characteristics of low/mid-rise, low-density and in-site open space, and it remains in the A 'very good' WPSH cluster.

Sub-group A1 (60%) from A, has distribution characteristics of mid-rise, medium-density or high-rise, low-density, with large on-edge open space, and it downgrades to the B 'fair' WPSH group. This indicates that building height and

density may be the reason for degraded performance in winter compared to the annual level.
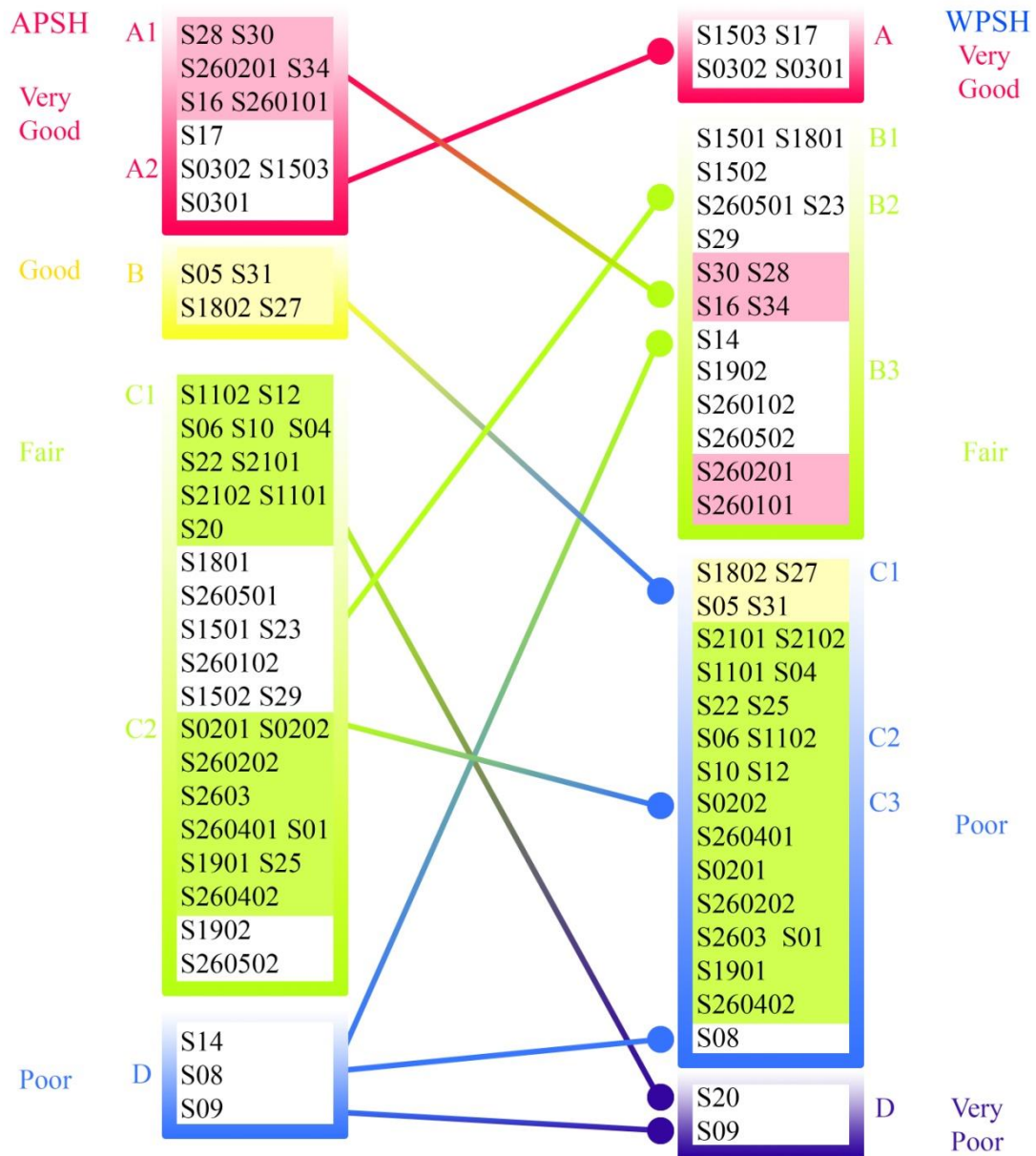
Figure 5.18 Comparison of Clustering Result (Position Change of Sites)

All cases in cluster B (good) degrade two levels to 'poor' in the WPSH cluster. These cases have distribution characteristics of low/mid-rise, medium-density, long façades evenness, enclosure and no open space. This phenomenon and the comparison with

other cluster changes shows even and dense low/mid-rise sites could have heavy shadow coverage in winter.

Seven out of seventeen cases (41.2%) in cluster C1 (fair) in APSH clustering remain in the 'fair' level, cluster B in WPSH clustering. These cases share characteristics of low/mid-rise, mix density, unevenness, in-site open space or low/mid-rise, medium-density, even, and south on-edge open space.

Nine out of seventeen cases (52.9%) in cluster C1 (fair) degrade to poor level WPSH cluster C. These cases share characteristics of low/mid-rise, medium-density, evenness, enclosure, small in-site space or low-rise, high-density, evenness, and enclosure. This shows that dense and even low/mid-rise sites would tend to have poorer sunlight conditions in winter. Site 20 in C1 (5.9%) downgrades two levels to the 'very poor' WPSH cluster D. It consists of long façade 6 floor buildings in high-density with large on-edge open space to the North. The dramatic degrade may have two reasons: 1. The north open space provides good sunlight accumulation in APSH however is partly heavily shaded in winter. Therefore, it no longer contributes as a large on-edge open space in the same way as the others located in the south in other sites. 2. The dense, even array of mid-rise buildings will cause heavy shadows between buildings, joining into a large area with very low WPSH value.

Two out of eleven cases (18.2%) in cluster C2 (fair) of the APSH cluster remain in 'fair' level WPSH, cluster B. These are sites of high-rise, low/medium-density, unevenness, and large south on-edge space. This result could be explained as the negative influence of high-rise buildings in winter compensated by large south open space with abundant sunlight even in winter.

Nine out of eleven cases (81.8%) in cluster C2 (fair) in APSH cluster degrade to the 'poor' level WPSH, cluster C. They are sites of high-rise buildings of medium-density, and evenness without open space. This indicates that evenness in high-rise

communities also causes heavier shadow in winter as well as in low/mid-rise communities.

Comparing changes in sites from C1 and C2, low/mid-rise sites have a higher ratio to remain at the same level in WPSH assessment (41%) than do high-rise sites (18.2%). Only three cases exist in the APSH cluster D (poor). Of these, one stays the same, one, downgrades, and one upgrades. Site 08 remains in the 'poor' level in the WPSH cluster. It is a high-rise site in high-density with enclosed boundary, yet half of its buildings have an orientation of 40° NW. The rotation may cause the building to cast a smaller shadow on the testing time section 14:00, and the footprint overlaps its own shadow. Therefore, the negative impact from high-rise buildings in winter is balanced.

Site 09 from cluster D degrades to a very poor level in WPSH cluster D. It is a site of high-rise buildings of high-density and long façades.

Site 14 from cluster D upgrades two levels to 'fair' in WPSH cluster B. It consists of buildings of 6 floors and 11 floors with characteristics of enclosure, long façade, and south on-edge open space. The buildings of S14 are formed in a courtyard style, therefore half of the building orientation is due east. These buildings will produce minimum shadow in winter on the ground, as their footprint overlaps their shadow to the maximum state. It is apparent that rotating building orientation from due south to due east could improve sunlight availability in winter to the max, but it will have a severe impact on indoor daylight and sunlight accessibility which is of a higher priority than outdoor sunlight. Because the Chinese residential building market is especially sensitive to the orientation if not due south, the rotation of the building is usually limited to within 10°. The improvement of the outdoor sunlight level by changing building orientation still needs further discussion.

Due to the difference in nature of APSH and WPSH, APSH design requires long accumulative hours on the whole site, while WPSH design requires the least amount of negative influenced area. Therefore, it could be said that the rules exacted from

APSH design is more suggestive of better performance, yet rules from WPSH design are instead about avoidable combinations. Therefore, the cases with an acceptable APSH level which have not degraded in WPSH assessment are of interest to the study. The cases which remained or upgraded are listed in Table 5.8 with their distribution characteristics.

Table 5.8 Rules Comparison from Analysis of APSH and WPSH Clustering

APSH Cluster	Cluster Change	Building Height	Density	Open Space	Evenness	Enclosure	Orientation
A	remain	low/mid-rise	low-density	in-site open space			
C1	remain	low/mid-rise	low/medium-density	in-site open space	uneven		
		low/mid-rise	low-density	south on-edge open space	even		
C2	remain	high-rise	low/medium-density	large south on-edge open space	uneven		
D	remain	high-rise	High-density	no		enclosed	half buildings of 40° NW orientation
D	upgrade	mid-rise	Medium- density	south on-edge open space		enclosed	half buildings of 90° E orientation

As seen in table 6.7.1, low/medium-density with the proper amount of open space is the key to maintain an acceptable WPSH level. Any adverse factor on WPSH performance in the distribution will need to be balanced by other favourable factors. For example, high-rise buildings in C2, is balanced by large on-edge open space to the south.

For medium/high-rise communities in medium/high-density and even enclosed, namely under rather adverse conditions, changing building orientation from due south in the appropriate amount could help to mitigate negative WPSH performance.

Due to the fact that the required distribution for favourable APSH is not fully in accordance with that of WPSH, the distribution characteristics which fulfil both is, in practise, highly recommended. Low/medium-density with the proper amount of open space is the key to maintain an acceptable WPSH level while boosting APSH performance, in which any adverse factor on WPSH performance in the distribution will need to be compensated by other favourable factors (i.e., high-rise clusters is balanced by large on-edge open space to the south). For medium/high-rise communities in medium/high-density and even enclosed, namely under rather adverse conditions, changing building orientation from due south in the appropriate amount could help out the case of negative WPSH performance.

5.8 Quantitative Analysis of APSH Statistical Measure and Building Morphology Parameter

To define the relationship between sunlight performance indices and building morphology parameters, multiple linear regressions are conducted. Because a series of dependent variables, namely performance indices is applied, and these have inner relations between each other, the ideal regression for studying their relationships is via the multiple linear regression model. However, for the purposes of depicting the impact of building distribution on a single performance index, the multiple linear regression model is used. The mathematical relationships will be used for multi-domain multi-objective optimisation in the following work.

5.8.1 Regression Parameter Selection

To execute multiple linear regressions, dependent parameters and independent parameters are needed. A series of statistic measures describing APSH performance is calculated as possible predictive parameters. Meanwhile, groups of statistic measures of building morphology parameter are calculated as possible independent parameters. In the regression, building morphology parameters are under control on their levels, and various scores of APSH performance variables are attributed to the level change of the building morphology parameter. Therefore, it is necessary to select the most representative indices from statistical measures of APSH performance and building distributions.

As discussed in Chapter 4, building distribution parameters are grouped and representatives are selected from each group for regression and optimisation. The selected indices are listed in table 5.9.

Table 5.9 Representative Building Morphology parameter for Regression

Group	Representatives	Abbreviation
Site scale	Total land area	TLA
Built amount	Residential building area	RBA
Existence of high-rises	TriangleSD	TSD
	Plot ratio	PR
Density	Building density	BD
	Average façade ratio	aFR
Set back	Average distance to road	aD
Site shape	Shape factor	SF
Low and long façade amount	Average low façade ratio	aLFR
Short façade amount	Average min façade length	aFLmin
Large interval amount	Average interval area	aIAmean
Small interval amount	Average lower corner area	aCAL
Interval depth amount	Average higher interval depth	aIDH

Statistical measures of APSH are selected based on the explanation ability shown in analysis of simulation result of APSH. As discussed in section 6.5, shown in curve

chart of P10-P90 value of the sample site from the 'good' performance group, P20 of APSH is the first turning point of the curve, where the gradient of the curve starts to rise. Similarly, P70 of APSH is another turning point where the up-protruding curve connects to a flatter gradient curve. Therefore, P20 and P70 are selected as representative of critical measures. P50 is also adapted for regression as it represents the overall condition of APSH. H413Ratio and H0Ratio are selected as the lower limit in assessing APSH performance, because their nature of proportions of grid points are below the required critical value as in regulation and guidance. Interquartile range is also used as a representative of the factor of data range, yet eliminates the influence of max and min values. The range could represent the spread of APSH grid data. The summary of independent indices selection is shown in table 5.10.

Table 5.10 Summary of Statistical Measure Selection for APSH Data

Parameter	Abbreviation	unit	range		Definition	Reason
P20	APSH-P20	H	0	1653	20% percentile of APSH grid value of a site	first turning point of histogram curve
P50	APSH-P50	H	0	1653	50% percentile of APSH grid value of a site	Representing average and overall level of data
P70	APSH-P70	H	0	1653	70% percentile of APSH grid value of a site	second turning point of histogram curve
H413Ratio	APSH-THR(413)	/	0	1	Ratio of grid points with sunlight duration below 413 hours	Violation Ratio below critical threshold value mentioned in regulation and guidance
H0Ratio	APSH-THR(0)	/	0	1	Ratio of grid points with sunlight duration equal to 0 hours	Violation Ratio below critical threshold value mentioned in regulation and guidance
Quatile Range	APSH-IQR	H	0	1653	The difference between two quartiles of APSH grid value of a site	Indicating the spread of APSH data

5.8.2 Regression Result of P20, P50, P70 of APSH Data

The three APSH performance indicators (P20, P50 and P70) are all focusing on APSH value at key grades. By controlling the key grade threshold value, it is possible to construct the idea APSH configuration at each level band.

P20 is the second percentile of all APSH grid data. It refers to the APSH value below which grid point values account for 20% of all data. By P20's definition, it is clear that it represents the lower end level of APSH data distribution, namely the area with relatively poor APSH performance in a site. To achieve overall good APSH performance, it is expected to have a higher P20 value of APSH. Similarly, P50 and P70 refers to the APSH value below which the grid point values account for 50% or 70% of all data; namely the areas with medium level performance or relatively good performance in a site, respectively. P50 and P70 are also expected to be high for a site with good overall APSH performance.

The regression of P20 adopts two approaches. Firstly, the regression of P20 with selected representative indices is operated. However, the power of explanation is not satisfactory. Secondly, a regression of P20 with all possible morphology parameters is supplemented. The two-model comparison could be seen in table 5.11.

Table 5.11 Comparison of Two Regression Approach of P20

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
P20 VS representatives[a]	.666	.444	.416	107.299923377	16.339	.000
P20 VS all indices[b]	.726	.528	.492	100.083191699	14.895	.000

a predictors: (constant), TSD, PR.
b predictors: (constant), TSD, AR, RBFA

For the current sample, F is 16.339 and 14.895 for the two regressions, respectively, which are significant at P<0.01, which means both regression equations are valid.

In the regression of P20 with representatives, the R square equals to 0.666, therefore the variance of the regression counts for 66.6% of all variances of the data. The adjusted R square is 0.444, which means the regression has a poor level of cross validation in a different sample from the same population. The shrinkage of prediction

power of the outcome in another sample is 0.222, namely 33.3% of the current level of explanation. This means the generalisation in another sample in the same population of the P20 regression equation with representatives is not satisfactory.

Comparatively, for regression of P20 with all indices, the R square is 0.726 and adjusted R square is 0.528. The shrinkage is 0.198 or 27.3% of current explanation power. The generalisation is still not satisfactory; however, the power of explanation increases. Sample size expansion is required for future work to eliminate the generalisation shrinkage.

The indices for regression of P20 seems not sufficient to cover all influential factors, therefore further study on building the distribution index regarding low level APSH is still needed. The current regression equations may not be able to guide future design but are still integral to understanding the relationship of low level APSH and building distribution are attributed.

The regression of P20 with representatives includes indices of TSD and plot ratio (table 6.9.5), while the example with all indices includes indices of TSD, aspect ratio and residential building footprint area (table 6.9.6).

As seen in table 5.12, Higher P20 requires higher TSD and a lower plot ratio, namely clustered distribution and less high-rise buildings. Shown in table 5.13, Higher P20 requires higher TSD, lower aspect ratio and higher RBFA. In other words, it refers to clustered distribution, the lower ratio of residential superficial area by an un-built space (related to number of high-rise buildings), and a higher residential building footprint area (related to larger site, more low/medium rise).

Comparing two equations, the tendency of distribution attributes pointed out by the indices are the same: clustered, less high-rises and more low/medium-rises. Because regression of P20 with all indices shows a higher power of explanation and it is not over fitting with too much independent variables, the equation including TSD, AR,

and RBFA is adopted for further discussion.

Table 5.12 Details of Regression of P20 with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	456.765	50.471		9.050	.000
TriangleSD	.372	.065	.789	5.691	.000
Plot Ratio	-85.292	32.395	-.365	-2.633	.012

Table 5.13 Details of Regression of P20 with All Possible Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	360.374	53.850		6.692	.000
TriangleSD	.425	.064	.902	6.669	.000
Aspect Ratio	-83.349	27.159	-.396	-3.069	.004
Residential Building Footprint Area(㎡)	34.418	14.809	.271	2.324	.025

Regression of P50 is operated in two attempts: regression with selected representative indices and with enhanced representatives. The comparison of the two models is shown in table 5.14. Because the second model has a higher power of explanation and enhanced the independent variables by one over the regression result with selected indices, the regression of P50 with enhanced representatives is used for further discussion.

Table 5.14 Comparison of Two Regression Approach of P50

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
P50 VS Selected[a]	.813	.661	.644	69.940108474	39.913	.000
P50 VS enhances representatives[b]	.858	.736	.708	63.304732246	27.121	.000

a predictors: (constant), PR, TSD,

b predictors: (constant), PR, TSD, AR, aIDL

For regression of P50 with enhanced representatives, the R square is 0.736 with an adjusted R square of 0.708. The shrinkage of explanation power is 0.028 or 3.8% of degree of explanation for the sample in use, namely the generalisation of P50 regression with enhanced indices is good. Its regression details, including coefficients, are listed in table 5.15.

As seen in table 5.15, a higher P50 level means a lower plot ratio, higher TSD, lower aspect ratio, and a higher average lower interval depth. This equals to the combination of attributes including fewer high-rise buildings, clustered distribution, low aspect ratio and increased minimum interval depth (or larger smallest interval area), all of which could lead to higher overall APSH level in a site, as represented by its median value P50.

Table 5.15 Details of Regression of P50 with Enhanced Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1133.693	48.709		23.275	.000
Plot Ratio	-116.716	35.967	-.598	-3.245	.002
TriangleSD	.252	.039	.640	6.423	.000
Aspect Ratio	-72.826	32.440	-.415	-2.245	.031
Avginterval Depthlow	3.261	1.468	.188	2.222	.032

Regression of P70 is operated with representative indices, and tested by regression with all building morphology parameters. The two regression results are identical. For regression of P70 with representatives, the R square is 0.747 and adjusted R square is 0.735 with shrinkage of 0.012 or 1.6% (table 5.16). It could be said that P70 regression with representatives could generalise in the 'very good' level.

Table 5.16 Summary of Regression of P70 with Representative Indices

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
P70 VS Representatives[a]	.864	.747	.735	66.944593915	60.530	.000

a predictors: (constant), PR, aD

Details of the regressions of P70 shows in table 5.17. To achieve a higher P70 level, lower plot ratio and higher average distance to road are required. In other words, higher level P70 prefers fewer high-rise buildings and larger set back distance at the boundary or large on-edge open space.

Table 5.17 Details of Regression of P70 with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1476.281	39.436		37.435	.000
Plot Ratio	-183.398	17.054	-.848	-10.754	.000
AvgDistance	2.805	.861	.257	3.257	.002

As seen in table 5.18, three regression equations of P20, P50 and P70 are listed. P70 equation is the most explained variable by its indices with R square of 0.747. By the application of a singular equation, it is possible to put special interest on certain grades of APSH with its impact factors, while by holistic consideration of three equations, it is possible to study the tied up effects among APSH grades and among morphology parameters.

Table 5.18 Regression Equation Comparison Between P20, P50 and P70 Regression

Model	Regression Equation	R^2	Sig
P20 With All	P20=0.425×TSD-83.349×AR+34.418×RBFA+360.374	0.528	.000
P50 With Enhanced	P50=0.252×TSD-72.826×AR-116.716×PR+3.261×aIDL+1133.693	0.736	.000
P70 With Selected	P70=2.805×aD-183.398×PR+1476.281	0.747	.000

5.8.3　Regression Result of Interquartile Range, H413Ratio and H0Ratio of APSH data

The three APSH performance indicators (interquartile range, H413Ratio and H0Ratio) are all, to some extent, reflecting the evenness of the APSH level in a site. Compared to P20, P50 and P70, these three indicators are not concerned with the absolute value of APSH, but rather to what degree the variation of APSH is at all grid points.

Interquartile range is defined as the difference between two quartiles of the APSH grid value of a site. It is an alternative index to the range of APSH with the advantage

of eliminating impact from minimum and maximum value which is fixed to 0 and 1653 in this research. Interquartile range indicates the spread of overall APSH data. The expectation of good overall APSH performance is to have as long sunlight hours as possible, evenly distributed over the whole site under certain design limitations; namely all data concentrated in a small range at a high level. Therefore, it is expected to have a smaller interquartile range for a good APSH performance. When applied together with P20 or P70 as lower limit of design constraints, these indices will point to a site with even and relatively high values at all grid points.

H413Ratio and H0Ratio are two ratios of area violating the critical standard of outdoor sunlight accessibility under local conditions, as suggested in UK guidance. Definitions of H413Ratio and H0Ratio are ratios of grid points with sunlight duration below 413 and equal to 0 hours to the total number of grid points, respectively. It is clear that, to achieve better APSH performance, as low a level as possible of H413Ratio and H0Ratio is required.

For regression of interquartile range, two approaches have been made: with selected and with enhanced selected indices. The results show that regression with enhanced representative indices has included the indices in regression with selected representatives. Furthermore, the enhanced version regression has higher explanation ability of the variance with meaningful supplemented indices. Therefore, the enhanced version is adopted for further discussion. The summary of the two regressions is shown in table 5.19.

Table 5.19 Comparison of Two Regression Approaches of Interquartile Range

Model	R	R Square	Adjusted R Square	Std. Error Of The Estimate	F	Sig.
Interquartile VS Representatives	.739[b]	.547	.524	112.708111084	24.710	.000
Interquartile VS Enhanced Representatives[b]	.829[e]	.687	.646	97.254039970	16.688	.000

a predictors: (constant), TSD, PR.

b predictors: (constant), TSD, PR, LMRBA, HRBR, aD

For regression of the interquartile range with enhanced representative indices, the R square is 0.687 and adjusted R square of 0.646, namely with shrinkage of 0.041 or 6.0% of current explanation power. The equation would have good ability in generalisation.

The details of the equation of interquartile range are shown in table 5.20. Small interquartile range is expected for good overall APSH performance. It is required to have higher TSD, higher PR, higher LMRBA, higher HRBR and lower aD. In other words, this means a combination of attributes of clustered distribution, high amount and high ratio of high-rises, high amount of low/medium-rises (or larger site) and less set back at boundary. It could be concluded that a large-scale site with a high ratio of high-rises and in-site open space, but without on-edge open space tends to have an even level of APSH.

Table 5.20 Details of Regression of Interquartile Range with Enhanced Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1020.092	67.710		15.066	.000
TriangleSD	-.326	.075	-.595	-4.331	.000
PlotRatio	-62.061	34.599	-.228	-1.794	.081
Low/Medium-Rise Residential	-10.238	2.974	-.428	-3.443	.001
HighriseRatio	-175.525	68.163	-.410	-2.575	.014
AvgDistance	3.043	1.366	.222	2.229	.032

Comparing to regression of P20, P50 and P70, the only requirement on distribution shared in regression of interquartile range is the tendency to clustered distribution and large site scales according to regression of P20. A large amount and a high ratio of high-rises required in an equation of interquartile range is against which of P20, P50 and P70. Less on-edge open space is against regression of P70.

For regression of H413R and H0R, two approaches are applied for each: with selected indices and enhanced representatives. However, the power of explanation for all four equations is barely satisfactory. Therefore, the regressions of H413R and H0R with enhanced representative indices are listed in table 5.21 as comparison. The regression results could only be used for observation on impacts of building morphology parameters on H413R and H0R performances.

Table 5.21 Summary of Regressions of H413R and H0R

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
H413R with enhanced representatives[a]	.699[c]	.488	.450	3.502341640	12.712	.000[d]
H0R with enhanced representatives[b]	.477[b]	.227	.189	1.262554954	6.022	.005[c]

a predictors: (constant), TSD, aILmax, aCAH

b predictors: (constant), aFLR, AS

The details of regression of H413R with enhanced representatives are shown in table 5.22. As seen in the table, to achieve low H413R, there is a requirement for higher TSD, higher aILmax and lower aCAH. In other words, the attributes of the site with a low violation ratio below 413 hours are clustered distributed, open on boundary and closed at corner. This image echoes with the requirement for high P20, P50 and P70. Namely, clustered and open boundary patterns could holistically help the site in reaching average standard. With further adjustment on the clustered open-boundary site, a special boost in value in certain APSH grades could be made.

Table 5.22 Details of Regression of H413R with Enhanced Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	27.386	1.947		14.065	.000
TriangleSD	-.007	.002	-.428	-3.356	.002
AvgMaxIntervalLength	-.210	.058	-.594	-3.653	.001
AvgCornerHigh	.002	.001	.456	2.990	.005

The details of regression of H0R with enhanced representatives are shown in table 5.23. For lower H0R, it is necessary to have lower aFR and lower AR, namely open boundary and less high-rises compared to the unbuilt area. This result refers to a low plot ratio site with open boundary which could have the lowest proportion of area with zero-hour APSH.

Table 5.23 Details of Regression of H0R with Enhanced Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-2.483	1.171		-2.120	.040
AvgFaçadeRatio	.045	.017	.372	2.706	.010
aspect ratio	.671	.289	.320	2.324	.025

Although regression of H0Ratio is not sufficiently explained by these two morphology parameters, the tendencies are still uncovered which suggest that to depress the ratio of all-year-round completely shaded points, a high level of special openness (derived by low AR) and of boundary openness are suggested.

To summarise, the equations are listed in table 5.24.

Table 5.24 Regression Equation Comparison Between P20, P50 And P70 Regression

Model	Regression Equation	R^2	Sig
Interquartile with Enhanced	APSH-IQR= -0.326×TSD-62.061×PR-175.525×HRBR -10.238×LMRBA+3.043×aD+1020.092	0.687	.000
H413R with enhanced	APSH-THR(413)= -.007×TSD-.210×aILmax+0.002×aCAH +27.386	0.488	.000
H0R with enhanced	APSH-THR(0)=0.045×aFR+0.671×AR-2.483	0.227	.000

5.8.4   Regression of SVF

SVF is an index describing the sky openness of a site. It is an assessing parameter of daylight availability on building façades and heat island effect in urban environments. SVF influences heat island effect by impacting the ability of an urban environment to emit long wave radiation to the sky. Therefore, SVF is also adapted as a dependent

variable to assess residential ward daylight and thermal performances. SVF plays an important role in bridging sunlight, daylight performance assessment and thermal comfort assessment.

SVF is not a conventional index applied at an early stage of planning design. This makes it hard for designers to form an image at an early stage of a sample building distribution pattern, when SVF is limited to a certain value. But this cannot become the reason that SVF is not suitable to be applied as an early-stage parameter. With help of optimisation tools and visualisation tools, the combination of indices could be mapped at an early design stage as a sample pattern.

Two attempts of regression for SVF are operated: one with selected representative indices of building distribution, one with additional enhancement from the remaining indices. The summary of two regression models are shown in table 5.25. Compared to the regression model with representative indices, the enhanced model includes one more independent of aspect ratio. Because the enhanced model has a higher power of explanation by higher R^2, and the add-in index of aspect ratio has perfect physical meaning fitting the scenario of the regression, the enhanced model is used for further discussion.

Table 5.25 Comparison of Model Summary of SVF Regressions
with Representatives and Enhanced Representatives

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
SVF VS representatives[a]	.895	.801	.787	2.536247343	53.811	.000
SVF VS enhanced representatives[b]	.908	.824	.806	2.414929441	45.795	.000

a predictors: (constant), PR, TSD, aD
b predictors: (constant), PR, TSD, aD, AS

The regression between SVF and enhanced representative indices has R square of 0.824 and adjusted R square of 0.806, with shrinkage on explanation power in

generalisation of 0.018. In other words, the generalisation of this regression equation is good. Coefficients of the regression can be seen in table 5.26.

Table 5.26 SVF Regression Model Details

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	71.885	1.437		50.024	.000
plot ratio	-6.437	1.371	-.705	-4.696	.000
TriangleSD	.007	.002	.358	4.077	.000
AvgDistance	.093	.034	.201	2.740	.009
aspect ratio	-2.803	1.239	-.341	-2.263	.029

The model indicates that to achieve higher SVF, lower plot ratio, higher TSD, higher aD, and lower aspect ratio are required. In other words, to achieve higher SVF over site, design strategies suggested are: not too many high-rises, clustered distributed, large set back at boundary or large on-edge open space, and less aspect ratio (less ratio of superficial area of building against unbuilt area in-site).

The regression equation of SVF could be consolidated as:

SVF=0.007×TSD-6.437×PR+0.093×aD-2.803×AR+71.885

5.8.5 Comparison and Application of Regression Equations

Comparison of Regression Equations

Regression coefficients of all seven equations are listed and compared in table 6.8.21. Comparing standardised coefficients of regressions for P20, P50 and P70, shared and discrepant indices are discussed (table 5.27). Lower end and overall level of APSH, represented by P20 and P50 performance, are both sensitive to evenness of distribution (positive) and aspect ratio (negative). TSD has a 29.1% higher influence on P20 than P50, while AR has a 4.6% higher influence on P50 than P20. Higher end and overall level of APSH, represented by P70 and P50, are both negatively

correlated to PR. The impact of PR is 29.5% higher on P70 than P50. Lower end, median level, and high end of APSH are respectively, positively correlated to RBFA, aIDL and aD.

Table 5.27 Standardised Coefficient Comparison between P20, P50, P70, QuartileRange, H413Ratio, H0Ratio and SVF Regression

Index Group		High-rises				Site scale		Large interval		Interval depth	Set back	Density
Models	Expect	TSD	AR	PR	HRBR	RBFA	LMRBA	aILmax	aCAH	aIDL	aD	aFR
P20 with All	Large	0.902	-0.396			0.271						
P50 with Enhanced	Large	0.640	-0.415	-0.598						0.188		
P70 with Selected	Large			-0.848							0.257	
Quartile with Enhanced	Small	-0.595		-0.228	-0.410		-0.428				0.222	
H413R with enhanced	Small	-0.428						-0.594	0.456			
H0R with enhanced	Small		0.320									0.372
SVF with enhanced	Large	0.358	-0.341	-0.705							0.201	

Unlike P20-P70 regression which is concerned with APSH value at a certain grade, quartile range and two threshold ratio are concerned with APSH evenness. Quartile range and H413Ratio share TSD as an indicator and TSD has a 28.1% higher impact on quartile range than to H413Ratio. Except impact from indices of high-rise cluster, H413Ratio also includes aILmax and aCAH from a cluster of a large interval amount in its equation. The equation of H0Ratio has not enough power of explanation and includes only two indices: AR from the high-rise cluster and aFR from the density cluster.

SVF regression shares an identical tendency of distribution characteristics as an overall condition of APSH value, including lower range P20, average level P50 and higher level P70. The impact of TSD on SVF is higher than that of P70 but lower than those of P20 and P50. The impact of AR on SVF is between those of P20 and P50. The impact of PR on SVF is between P50 and P70. The impact of aD is slightly lower than that of P70.

Generally speaking, the regression of the group which cares about APSH value (P20, P50, P70 and SVF) shares similar requirements of distribution attributes to achieve good APSH performance: clustered distributed, less high-rise buildings, more low/medium-rise buildings, certain amount of openness on the boundary, and larger on-edge open space. In other words, these descriptions refer to an image of low-density, clustered distributed, low/medium-rise communities with a relative open boundary. This is the ideal low-density community design pattern which could be compromised with other factors in design.

However, the regression of the group which is concerned with APSH evenness (quartile range) shows disparate requirements on distribution attributes to achieve good APSH performance when comparing to that of P20-P70 and SVF. The disparate requirements refer to clustered distribution, more high-rise buildings, more low/medium-rise buildings (larger site), and less on-edge open space, namely, a large clustered distributed site of a considerate number of high-rises with few on-edge open spaces. This pattern leads to homogeneous APSH level over the site. To have higher and evener APSH level, a combination of quartile range with interested APSH grades' indices is necessary.

The regression group which is concerned with the proportional configuration of APSH (H413Ratio and H0Ratio) shows opener constraint on distribution patterns which is easier to achieve. The requirements of the satisfactory threshold (H413Ratio) are evenly distributed, relatively enclosed and relatively open at corner. The

requirements for the worst threshold (H0Ratio) are more high-rises and more enclosed at boundary. This could suggest that as an evenly distributed community with relatively enclosed boundaries but relatively open corner is the strategy to allow as much area in site as possible to meet the lowest outdoor sunlight standard; however, a high-rise community with a relatively enclosed boundary is the strategy to dislodge the proportion of zero sunlight area. This means we only ensure the lowest and therefore worst standard is met to the greatest extent, but no good APSH performance is ensured.

These reverse tendencies in the distribution requirement for good overall APSH by variously assessing variables provide opportunities to balance and optimise design by adjusting building distribution patterns. However, achieving a higher overall APSH level still has higher priority over APSH evenness in a site. Therefore, in design optimisation procedures, it is necessary to fulfil P20, P50, P70 and SVF requirements before achieving sunlight evenness. H413Ratio and H0Ratio requirements are also vital, even compulsory. Because the constraints are loose, they are more likely to align with requirements from those of the group concerned with APSH value.

To bring the regression results to a conclusion, they suggest that even distribution helps with reaching an average standard, while clustered distribution helps to boost APSH value at the area of interest. Low/medium-rises help to boost the APSH value at various levels, while high-rises help to reduce the worst sunlit areas. The detailed rules are listed below:

1. High-rises with a relatively enclosed boundary leads to the dislodging of zero sunlight areas.

2. Even for communities with enclosed boundaries, having open corners can lead to the maximum proportion of area of meeting the lowest standard.

3. Clustered low/medium-rises with open boundaries leads to a boost in APSH value in various levels.

4. Large clustered sites with high-rises and no on-edge open space lead to maximum evenness of APSH.

Application of Regression Equations

The unstandardised regression coefficient b is the change in the outcome associated with a unit of change in the predictor variable. For a smoother connection to the practice and easier reference for designers, the amount of change in the building distribution parameter associated with a unit improvement in the performance variable is calculated, converted from the regression coefficients.

The regression coefficient of the abovementioned seven regressions in the previous section are transformed in table 5.28. The required compromises for each influential parameter in design are calculated according to per unit improvement of corresponding performance variables. For P20, P50, P70 and quartile range, the unit of change is one hour increase of effective sunlight. For H0Ratio, H413Ratio, the unit is a 10% decrease in proportion, and for SVF, is a 10% increase in proportion.

Table 5.28 Equivalent Change of Building Distribution to
One Unit Improvement of APSH Performance Variables

Independent Variables			P20 +1H	P50 +1H	P70 +1H	quartile range-1H	H413Ratio -10%	H0Ratio -10%	SVF +10%
Cluster	Name	Range							
High-rise	TSD	60-1200	2.353	3.968		3.067	1428.571		1428.571
	AR	0.5-4	-0.012	-0.014				-14.903	-3.568
	PR	0.6-3.6		-0.009	-0.005	0.016			-1.554
	HRBR	0-1				0.006			
Site scale	RFPA	0.3-7	0.029						
	LMRBA	0-35				0.098			
Large interval	aILmax	18-86					47.619		
	aCAH	0-5837					-5000.000		
Interval depth	aIDL	10-36		0.307					
Setback	aD	13-61			0.357	-0.329			107.527
Density	aFR	42-91						-222.222	

Seen from the table, for an improved APSH performance, the general trend of the morphology parameter predicted by a series of performance indices is the same, except for the quartile range which shows reverse requirements on distribution.

Some step sizes of independent parameters for one unit change of performance variable are too large, violating practical and economical principles. As shown in table 5.28, a few expected changes in distribution for one unit change of performance indices violate the range of the corresponding morphology parameter. To improve APSH performance by changing these index values is impractical. This phenomenon is consistent with the regression coefficient of the relevant index in the function, which is relatively small indicating less sensitivity over the objective variable. With expansion of sample size and the enhancement of the predictive accuracy of the regression function, the sensitivity of DVs would be improved to attenuate this phenomenon. These are: to increase TSD by 1428.571 for a 10% decrease of H413Ratio, to decrease AR by 14.903 and aFR by 222.222 for a 10% decrease of H0Ratio, and to increase TSD by 1428.571 and aD by 107.527 for a 10% increase of SVF. If we only expect a 1% improvement in H413Ratio, H0Raio and SVF, the change step is still applicable as it no longer violates the distribution index range. Both the indices adopted in the H0Ratio could not provide valid equivalent compromises for improved H0Ratio, therefore, this equation needs further alternation after data sample expansion.

The remaining equivalent change of morphology parameters are practical in design and very effective. For example, only a 0.009 decrease of PR could rise one hour of effective sunlight assessed by P50. These morphology parameters are more possibly applied in the early design stage as control parameters following the suggestions from APSH performance equations. To apply, refer the corresponding change of distribution index based on the required performance change. For instance, for one hour improvement in low level APSH assessed by P20, TriangleSD needs to be

increased by 2.353; RFPA needs to be increased by 290 m^2; and AR needs to be decreased by -0.012.

5.9 Quantitative Analysis of WPSH Statistical Measure and Building morphology parameter

5.9.1   Regression Parameter Selection

WPSH is used for assessing sunlight availability accumulated in winter. A series of statistical measures of WPSH are calculated for analysis. As discussed in section 6.6, the curve chart of WPSH P10-P90 value shows two turning points on the curve: at P30 and P70. P30 is the turning point between the flat section of zero value and the up-protruding section on the curve in the middle range. P70 is the turning point of the middle range curve and the steep high range curve. Also, for cases of high-rise buildings in medium-density and low/mid-rise buildings in medium-density, the largest vertical distance occurs at P70 between the two curves. Namely, it is the indicator which could reflect the largest difference between two categories of WPSH conditions. Additionally, P70 is the intersection point of the low/mid-rise and the high-rise curve from the worst WPSH condition group. P50 is used because it is representative of the overall level of WPSH. For a good WPSH performance, P30, P50 and P70 are all expected high for extended sunlight hours in low, average and high level of WPSH, respectively.

H83Ratio is the ratio of area violating the lower limit of WPSH. The threshold value is calculated according to proportional guidance of UK sunlight requirements in residential areas. H0Ratio is the ratio of the area with zero-hour sunlight, namely the ratio of area permanently shadowed area in the winter season. It is clear that for better WPSH performance, the H83Ratio and H0Ratio are both expected to be as low as possible. Interquartile range is an alternative index to range, with advantage of eliminating the impact from minimum and maximum. Interquartile range is

representative of evenness in the value distribution of WPSH. Similar to APSH, interquartile range is expected low for a value evenness of WPSH in the site.

WPSH performance indices are consolidated in table 5.29.

Table 5.29 WPSH Performance Indices Adopted for Regression

Parameter	Abbreviation	unit	range		Definition	Reason
P30	WPSH-P30	H	0	348	30% percentile of WPSH grid value of a site	turning point between zero section and up-protruding section in middle range of WPSH percentile curve
P50	WPSH-P50	H	0	348	50% percentile of WPSH grid value of a site	Representing average and overall level of data
P70	WPSH-P70	H	0	348	70% percentile of WPSH grid value of a site	turning point of middle range section and steep high range section on percentile curve; point with largest vertical difference between curves of various WPSH catogories.
H83Ratio	WPSH-THR(83)	/	0	1	Ratio of grid points with sunlight duration below 83 hours in winter	Violation Ratio below critical threshold value mentioned in regulation and guidance
H0Ratio	WPSH-THR(0)	/	0	1	Ratio of grid points with sunlight duration equal to 0 hours in winter	Violation Ratio below critical threshold value mentioned in regulation and guidance
Interquartile Range	WPSH-IQR	H	0	348	The difference between two quartiles of WPSH grid value of a site	Indicating the spread of WPSH data

Building distribution parameters applied in WPSH regressions are identical to those for APSH regression. For details of building morphology parameter, refer to table 6.9.1.

5.9.2 Regression Result of P30, P50, P70 of WPSH Data

Similar to the analysis of regression of APSH statistical measures, the analysis of the regression of WPSH measures is simply summarised.

Two or three attempts are operated for P30, P50 and P70 regression, the models with selected morphology parameter are adopted, instead of enhanced selected models, for higher level of explanation power on variations of the equations. Details of the models refer to table 5.30.

Table 5.30 Model Summaries of Regression of P30, P50 and P70

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
P30 with selected	0.723[a]	0.522	0.487	24.892678422	14.583	.000
P50 with selected	0.764[b]	0.583	0.552	25.869837827	18.640	.000
P70 with selected	0.843[c]	0.710	0.680	27.776397677	23.880	.000

a. Predictors: (Constant), TriangleSD, plot ratio, AvgLowFaçadeRatio

b. Predictors: (Constant), plot ratio, TriangleSD, AvgLowFaçadeRatio

c. Predictors: (Constant), plot ratio, AvgDistance, AvgLowFaçadeRatio, TriangleSD

As seen in the table, F-tests for all three equations are significant, and their R squares are 0.522, 0.583 and 0.710, respectively. The power of explanation is acceptable for this sample size. The adjusted R squares are 0.487, 0.552 and 0.680, respective for P30, P50 and P70, namely shrinkages in generalisation are 6.7%, 5.3% and 4.2% of the current power of explanation. This means the three equations are valid in this sample and could be generalised in other samples from the same population.

Details of standardised coefficients of P30, P50 and P70 refer to table 5.30 in the model comparison. Regression equations of P30, P50 and P70 are shown in table 5.31.

Table 5.31 Regression equations of P30, P50 and P70

Model	Regression Equation
P30 With Selected	$P30=0.110×TSD-27.139×PR+0.348×aLFR+29.125$
P50 With Selected	$P50=0.093×TSD-62.485×PR +0.468×aLFR+174.978$
P70 With Selected	$P70=0.043×TSD-76.169×PR +0.584×aLFR+1.446×aD+257.102$

Seen from the equations, to achieve a higher level P30 or P50 value, higher TSD and aLFR, lower PR are required. In accordance with the cluster analysis of building morphology parameters, this requirements of achieving a higher level P30 or P50 value mean to achieve higher value at low level and overall level of WPSH the site should be clustered distributed, through a low plot ratio (less high-rises), and with a larger amount of low and long façades at boundaries. This image refers to a clustered low/medium-rise community with some degree of enclosure at the boundary. A higher value of P70 shows identical requirements on morphology parameters as P30 and P50, except for asking for higher aD, namely larger on-edge open space on the edges of the site.

The differences between coefficients of P30, P50 and P70 show that TSD has the highest impact on P30 compared with the other two, while PR and aLFR are most powerful in influencing P70 among the three. Only P70 is under influence of aD.

5.9.3   Regression Result of Interquartile Range, H83Ratio and H0Ratio of WPSH data

Multiple attempts have been operated for regression of Interquartile Range, H83Ratio and H0Ratio. The summaries of models are listed in table 5.32. Regression of the interquartile range with selected representative indices is adopted while regression of H83Ratio with enhanced representatives is used because, except for the representatives, one more index is added by the stepwise regression method. However, due to the un-satisfactory result of regression of H0Ratio when selected, the version with all possible indices is adopted, even though the R square of H83Ratio and H0Ratio model is only satisfactory at 0.515 and 0.591, respectively.

Table 5.32 Model Summaries of Regression of P30, P50 and P70

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
Interquartile Range with selected	0.859[a]	0.737	0.718	28.050160679	37.455	.000

| H83Ratio with enhanced | 0.718[b] | 0.515 | 0.465 | 6.123386113 | 10.355 | .000 |
| H0Ratio with all indices | 0.769[c] | 0.591 | 0.560 | 4.766470817 | 19.247 | .000 |

a. Predictors: (Constant), plot ratio, AvgDistance, building density

b. Predictors: (Constant), TriangleSD, plot ratio, AvgLowFaçadeLength, highriseRatio

c. Predictors: (Constant), TriangleSD, AvgLowFaçadeLength, building density

Details of the three regression models are shown in table 5.33. As seen from the equation of WPSH-IQR, for lower IQR, it is required to have higher PR, lower aD and BD. Namely a site of low-density,a high plot ratio site with rare on-edge open space tends to have a high level of WPSH evenness over the site. For a low H83Ratio, high TSD, aLFL and HRBR, and low PR are required. Here, the changing trends of HRBR and PR are contradictory, and PR still has high predicting power, which means although low PR is needed, the number of high-rises is still restrained by the certain level of high-rise ratios among all residential buildings. To summarise, clustered distributed sites of relatively low plot ratio with some level of enclosure at the boundary has a small ratio of violation of 83h effective sunlight hours in winter. Similar to H83Ratio, low H0Ratio requires high TSD and aLFL, and low BD. In other words, clustered distributed sites in low-density with some level of enclosure at the boundary has a small ratio of a permanent shadowed area. From the difference between the equation of H83Ratio and H0Ratio, the prediction power of the restrictive relationship between PR and HRBR is exchangeable for BD.

Table 5.33 regression equations of WPSH-IQR, WPSH-THR(83) and WPSH-THR(0)

Model	Regression Equation
Interquartile Range With Selected	WPSH-IQR =354.726×BD-63.713×PR+1.337×aD+211.399
H83Ratio with enhanced	WPSH-THR(83)= -0.020×TSD+12.051×PR-0.044×aLFL-8.107×HRBR +35.249
H0Ratio with all indices	WPSH-THR(0)= 57.806×BD-0.012×TSD-0.053×aLFL+19.146

5.9.4 Comparison and Application of Regression Equations

Standardised coefficients of the six regression equations of WPSH performance are listed for comparison in table 5.34. Regressions of P30, P50 and P70 show identical requirements and changing tendencies on the morphology parameter. TSD shows the highest impact on P30 than P50 and P70, while PR and aLFR shows the highest impact on P50. P70 is also influenced by aD.

Just as APSH performance regressions, interquartile range shows reverse requirements on distribution compared to that of percentiles of WPSH. The requirements from H83Ratio and H0Ratio regressions accord with those of P30-P70 in a high degree.

To summarise, for a higher value at various levels of WPSH, clustered, low plot ratio sites with certain enclosures at the boundary are referred. For higher standard reaching ratio, except for the above requirements, low building density is needed. However, the value evenness of WPSH refers to sites of high plot ratio, low building density, and less on edge open space.

Table 5.34 Standardised Coefficients of Regression Equations of WPSH Performance

Index Group		High-rises			Set back	Density	Low/long façade amount	
Models	Expect	TSD	PR	HRBR	aD	BD	aLFR	aLFL
P30 with selected	Large	0.940	-0.469				0.250	
P50 with selected	Large	0.719	-0.972				0.302	
P70 with Selected	Large	0.262	-0.932		0.350		0.297	
interquartile with selected	Small		-0.725		0.301	0.340		
H83R with enhanced	Small	-0.724	0.865	-0.369				-0.306
H0R with all	Small	-0.511				0.407		-0.437

Compared the standard coefficients from regressions of WPSH performance to which of APSH performance,

Site scale factor, large interval factor and interval depth mentioned in APSH regression are not involved in WPSH regression. Meanwhile, low and long façade amount adopted in WPSH regression is not mentioned in APSH.

WPSH-P30, P50 and P70 show identical requirements of clustered distribution and low plot ratio as shown in APSH-P20, P50 and P70 regression. Furthermore, WPSH performance has tighter restraints on distribution than APSH performance, i.e., the standardised coefficient of TSD and PR are 0.719 and -0.972, respectively, for WPSH-P50, compared to 0.640 and -0.598 for APSH-P50. The standardised coefficient of aD is 0.350 for WPSH-P70, compared to 0.257 for APSH-P70.

Regressions of WPSH-IQR, THR(83) and THR(0) involve disparate distribution factors compared to those of APSH-IQR, THR(413) and THR(0). As they are not comparable, they need to be discussed on a case basis.

For the smooth bridging of the regression result in practice, the amount of compromises of each distribution index for one unit improvement on WPSH performance variables are listed in table 6.10.7. The calculation is based on WPSH regression equations using unstandardised coefficients.

As shown in table 5.35, achieving a small H83Ratio and H0Ratio increase of HRBR by 1.234 and aLFL by 227.273 and 188.697, respectively, is impractical (shown as red in table). Excepting these, the remaining equivalent change of the morphology parameter is valid and effective in improving WPSH performance.

Table 5.35 Equivalent Change of Building Distribution to
One Unit Improvement on WPSH Performance Variables

Independent Variables			P30 +1H	P50 +1H	P70 +1H	Interquartile range-1H	H83Ratio -10%	H0Ratio -10%
Cluster	Name	Range						
High-rise	TSD	60-1200	9.091	10.753	23.256		500.000	833.333
	PR	0.6-3.6	-0.037	-0.016	-0.013	0.016	-0.830	

	HRBR	0-1					1.234	
Setback	aD	13-61			0.692	-0.748		
Density	BD	0.07-0.4				-0.003		-0.173
Low/ Long	aLFR	0-93	2.874	2.137	1.712			
Façade	aLFL	0-219					227.273	188.679

5.10   Summary

In this chapter, assessment parameters of sunlight availability all-year-round and in winter are discussed in parametric studies and are selected to be APSH and WPSH for further analysis and optimisation. Based on simulations of APSH and WPSH operated on Grasshopper platform, clustering of sites and regression of sunlight performance variables are carried out. Clustering of sites based on APSH and WPSH characteristics provides qualitative design suggestions of distribution attribute combinations for improved sunlight performance. Regressions of a series of APSH and WPSH performance variables could provide quantitative relationships between performance variables and building morphology parameters which could predict performance from building distribution pattern.

The achievements of this chapter are multiple. Firstly, it defines the key representative statistical measures for APSH and WPSH to describe the sunlight atmosphere of a site from low level to medium then high level. Secondly, the achievement is the matching of sunlight availability histogram pattern of a site to the building morphology characteristics.

Thirdly, the qualitative rules for building distribution are extracted from the analysis of sunlight hour data. For longer overall APSH and WPSH over the site, the preferred distribution combinations are low/mid-rise low-density, low/mid-rise medium-density or high-rise low-density.

Based on clustering by APSH, all three level (macro, meso and micro) rules are

expounded. Here they are summarised as:

1. The majority of practices have fair APSH conditions and have three peaks in the histogram of APSH grid data. For good and fair conditions, the middle peak is right-skewed; for poor condition, the middle peak is left-skewed.

2. With open spaces on the edge of site, the low/mid-rise community provides higher overall sunlight hours than high-rise community. This is especially obvious in the partly shaded area, shown as the right-skewed middle peak.

3. The best attribute combinations for APSH performance are low development intensity including low/mid-rise community, low-density, and in-site open space. Good combinations are mix-rise community in a low-density with tall buildings on the north. The worst combination is a high-rise high-density site with no open space.

Based on the site clustering by WPSH, design rules could be summarised as:

1. The majority of practices have poor WPSH conditions and have three peaks in the histogram of WPSH grid data. Only for the very good cluster, the middle peak is right-skewed or merged into a maximum peak. All other clusters have a left-skewed middle peak.

2. Evenly distributed, medium-density sites could provide higher middle peak location than uneven, mixed density sites.

3. High-rise sites tend to have a lower ratio of totally blocked areas, but a higher ratio of partly shaded areas.

4. The best attribute combination: low-rise buildings, low-density, unevenness and in-site open space, Worst combination: no matter high-rise or low-rise buildings combined with attributes of high-density, enclosed, evenness.

The fourth achievement of this chapter is 13 regression equations of APSH and WPSH performance variables. The equivalent compromises of morphology parameter for one unit improvement of sunlight performance are calculated based on the equations. This is for the convenience of reference in the design procedure. The equations are consolidated as below in table 5.36. These functions will be applied in the following multi-domain multi-objective optimisation for meta-model training.

This chapter is the parallel performance simulation and data analysis of sunlight availability domain, the same as for acoustic and thermal domain.

Table 5.36 Consolidation of Regression Equations of Sunlight Availability

Model	Regression Equation
APSH-P20 With Selected	APSH-P20=0.110×TSD-27.139×PR+0.348×aLFR+29.125
APSH-P50 With Selected	APSH-P50=0.093×TSD-62.485×PR +0.468×aLFR+174.978
APSH-P70 With Selected	APSH-P70=0.043×TSD-76.169×PR +0.584×aLFR+1.446×aD+257.102
Interquartile with Enhanced	APSH-IQR= -0.326×TSD-62.061×PR-175.525×HRBR -10.238×LMRBA+3.043×aD+1020.092
H413R with enhanced	APSH-THR(413)= -.007×TSD-.210×aILmax+0.002×aCAH +27.386
H0R with enhanced	APSH-THR(0)=0.045×aFR+0.671×AR-2.483
SVF with enhanced	SVF=0.007×TSD-6.437×PR+0.093×aD-2.803×AR+71.885
WPSH-P30 With Selected	WPSH-P30=0.110×TSD-27.139×PR+0.348×aLFR+29.125
WPSH-P50 With Selected	WPSH-P50=0.093×TSD-62.485×PR +0.468×aLFR+174.978
WPSH-P70 With Selected	WPSH-P70=0.043×TSD-76.169×PR +0.584×aLFR+1.446×aD+257.102
Interquartile Range With Selected	WPSH-IQR =354.726×BD-63.713×PR+1.337×aD+211.399
H83Ratio with enhanced	WPSH-THR(83)= -0.020×TSD+12.051×PR-0.044×aLFL-8.107×HRBR +35.249
H0Ratio with all indices	WPSH-THR(0)= 57.806×BD-0.012×TSD-0.053×aLFL+19.146

Chapter 6 Sample Simulation in Outdoor Thermal Comfort Domain and Performance Data Analysis

As indoor thermal comfort becomes more recognised, increasing studies are focusing on outdoor thermal comfort. For the consideration of controllable performance indices in the early design stage, only wind speed, mean radiant temperature and longwave radiation is adopted under the affect of building distribution.

The aim of this chapter is to prepare simulation data and relationships between thermal comfort indices and building morphology parameters for following multi-domain multi-objective optimisation. The objective of the chapter is first to decide the proper thermal comfort performance indices and their corresponding statistic measurements for data analysis; second it is to discover the reasoning behind the performance data distribution pattern and building distribution form; the third objective is to discover the qualitative guidance for building distribution form to achieve good thermal comfort; fourth is to achieve mathematical relationships of thermal comfort and building distribution for the meta-model training in MD-MOO.

Thermal performance simulation is operated in Envi-met. Statistical tools of hierarchical clustering and multiple linear regression are used for qualitative and quantitative data analysis.

The structure of this chapter is 6.1 background of thermal simulation; 6.2 thermal simulation software comparison; 6.3 simulation model setup in Envi-met; 6.4 parametric study of MRT and WS; 6.5 qualitative analysis for MRT; 6.6 quantitative analysis for MRT; 6.7 qualitative analysis for WS; 6.8 quantitative analysis for WS; 6.9 qualitative and quantitative analysis for LRE; 6.10 integration and of regression results and 6.11 summary. For the workflow of this chapter, see figure 6.1.
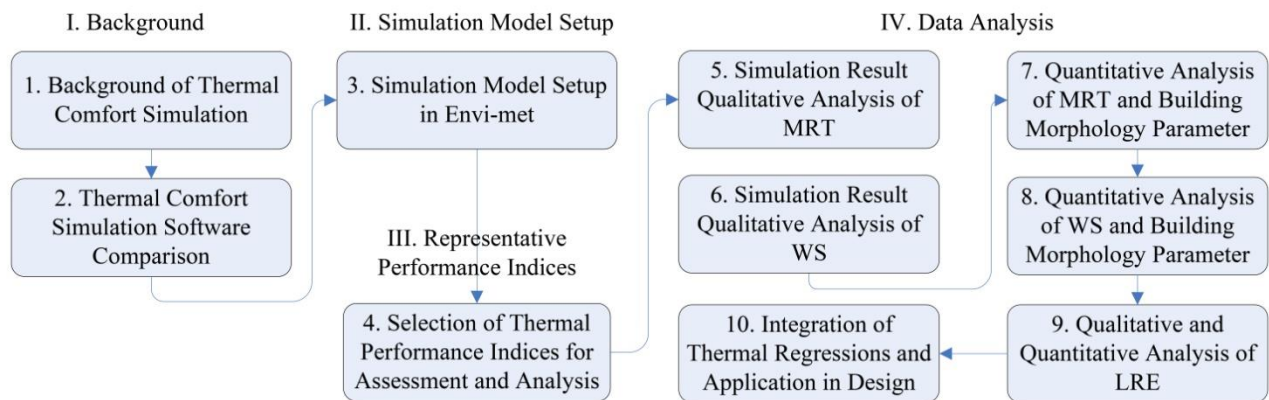
Figure 6.1 Sample Simulation in Thermal Comfort Domain and Data Analysis,
the expansion of box 6 in Overall Content Structure

Acronyms for Chapter 6

aCAL/AvgCornerLow	average corner area low value
aD/AvgDistance/Dmean	average distance to road of a site
aFLmin/AvgMinFaçadeLength	average min façade length
aFR/AverageFaçadeRatio	average façade ratio
aIAmean/AvgIntervalArea	average interval area
aIDH/AvgIntervalDepthHigh	average interval length high value
aIDL/AvgIntervalDepthLow	average interval depth low value
aLFR/AvgLowFaçadeRatio	average low-rise façade ratio
APSH	annual possible sunlight hours
BD/BuildingDensity	building density
CFD	computational fluid dynamics
GUI	graphical user interface
LRE	longwave radiation from environment
LRE- P70	70% percentile of LRE grid value of a site
MD-MOO	multi-domain multi-objective optimisation
MRT	mean radiant temperature
MRT-P25	25% quatile of MRT grid value of a site
P10, P20, P30, P40, P50, P60, P70, P80, P90	ten percentiles of all collected simulation data over one grid
P25, P75	two quatiles of all collected simulation data over one grid
PR/PlotRatio	plot ratio
RBA/ResiBuildingArea	residential building area
SF/ShapeFactor	site shape factor
SVF	sky view factor
TLA/TotalLandArea	total land area

TSD/TriangleSD	standard deviation of triangle area
WPSH	winter possible sunlight hours
WS	wind speed
WS-Max	Maximum of WS grid value of a site
WS-P20	20% percentile of WS grid value of a site
WS-P50	50% percentile of WS grid value of a site

6.1 Background of Outdoor Thermal Comfort Simulation

In residential context, thermal research is a mature field fully developed in both indoor and outdoor environment. For the indoor context, are related to heating and cooling load, energy consumption, life-cycle consumption, indoor objective thermal comfort and perceptive thermal comfort assessing with different indices. For the outdoor context, thermal researches concentrated on solar radiation studies, in related to productive efficiency of photovoltaic solar panels, heat gain at building surfaces, solar requirement of plant, etc. There grows popularity in the outdoor perceptive thermal comfort assessed with various metric systems. Thermal researches are quiet often combined with daylight and sunlight studies.

The outdoor thermal simulation in residential ward is usually focusing on heat gain on building surfaces or main recreational area. For high-rise building clusters, mainly of commercial or offices buildings, a few for residential building clusters, the ground level wind environment are considered.

As the research interest of the building distribution in residential wards, the focus would be the space between building clusters and open space-potential activity zone-with in the ward. Hence, wind environment, solar availability and environmental temperature would be the proper aspects of consideration from the objective thermal evaluation viewpoint.

With consideration of the complexity of CFD model and calculation cost, wind speed, mean radiant temperature, longwave radiation from environment are used as evaluation metrics. Wind speed between buildings is closely related to building arrangement. The wind speed map between buildings and in open space at specific time of the interested date is used. The mean radiant temperature and longwave radiation are used, as they directly express the affects of the building mass location.

6.2 Thermal Comfort Simulation Software Comparison and Selection

A great number of tools are applied for thermal related simulation and evaluations with different emphasis. A few tools related to the intended simulations in this research are discussed.

VENTSIM

It is a tool aiming for underground mine ventilation simulation. It is capable of various types of ventilation data, i.e., simulate airflows, pressure and heats, from a modelled network of airways. It has a 3d GUI for graphical design.

AIRPAK

It is a tool for computational fluid dynamics. It is capable of calculation of airflow modelling, contaminant transport, room air distribution, temperature and humidity distribution, and thermal comfort.

FLOVENT

It uses the computational fluid dynamic method. The simulation capability include airflow, heat transfer, and contamination distribution for built environments.

FLUENT

It is a modelling tool of natural ventilation in buildings. It models airflow under specified conditions in the CFD program. Additional analysis is required to estimate annual energy savings.

STAR-CD

It is a limited function software solely for detailed simulation of heating, ventilation, smoke, pollutant dispersal and fire hazard using CFD method. Limited deliberate natural ventilation is also acceptable in this software, including calculation of natural air infiltration regarding to temperature difference, wind speed and effective leakage area or infiltration rates.

URBAWIND

It models the wind in urban area. The natural airflow in the buildings then could be calculated according to the affect of the surrounding buildings and local climatology.

Energy-10

It is a programme aiming to define cost-effective and energy efficient design for small commercial and residential buildings. A good combination of design strategies could be made according to its simulations on daylighting, passive solar heating, and high-efficiency mechanical system.

DOE-2

It is a comprehensive hour-by-hour simulation tool focusing on the daylighting and glare calculation and hourly energy simulation.

ENERGY PLUS

It is a complex building energy simulation program, capable of modelling buildings with simulations associated heating, cooling, lighting, ventilating, and other energy flows.

IDA RTV

It is the software for simulations in road tunnel ventilation and fire with consideration of real traffic. Air pressure, flow rates, temperature and pollution concentrations are available in simulation.

Selected Software

ENVI-met

It is a free modelling and simulation tool for urban environment for 3D surface-plant-air interactions. It is widely applied in urban climatology, architecture, building design and environment planning, etc. The simulation resolution is 0.5 to 10m in space and 10 sec in time. It is capable of simulation including flow between

and around buildings, heat and vapour exchanges at ground surface and walls, turbulence, exchange at vegetations, bioclimatology and particle dispersion. The prognostic model applied is based on fluid and thermo dynamic laws.

Because the availability of the software, and the matching of flow and radiation focused simulation to the research interest, Envi-met is selected as the simulation tool for the sample cases regarding to outdoor thermal comfort in summer time.

6.3 Simulation Model Setup in Envi-Met for Mean Radiant Temperature, Wind Speed and Longwave Radiation from Environment

In this section, the configuration and method of simulation of mean radiant temperature (MRT), wind speed (WS) and longwave radiation from environment (LRE) in Envi-met for thermal comfort assessment is introduced, as well as the assessment criteria for the three aspects and discussions.

6.3.1　Model Configuration and Input

To operate an atmosphere model for thermal assessment in Envi-met, information of building geometry and meteorology data are required as input. Building geometry is saved in a thermal model file generated in the Space model tool integrated in Envi-met, while meteorology data and simulation configurations are noted in a configuration text file.

The configuration file records the simulation name, file base name for input and output and all basic simulation settings. The basic settings of the simulations are: start day of simulation (23.06.2014), start time of simulation (06:00), total simulation time span (16h, namely to 23:00), time step of saving model (every 60min), initial wind speed at 10m above ground (3m/s), wind direction (130 degree), roughness length at reference point on ground (0.1), initial temperature of atmosphere (293K or 19.85

Celsius degree), specific humidity at 2500m (7 g water/kg air), and relative humidity at 2m (50%). The humidity and initial wind speed and its direction are referred to the climate database of Hefei, where it is selected as a representative location of sites from SU-ZHE-WAN region.

Although the interested time slot of the simulation is between 14:00 to 16:00 long time span simulation is suggested from sunrise over the specific time point. This is due to the accuracy requirement of the initialisation of simulation to start with a smaller gradient of increasing rate or even zero of solar radiation, namely early morning or twilight of the simulation day, respectively. This is the requirement of the simulation software in order to ensure the stability of the simulation results. The stabilisation of thermal inertia of materials and air movement due to solar radiation, which decides accuracy of thermal simulation, also require a sufficient time length for simulation.

The geometry model is built on the Envi-met editor of a 2.5D modelling tool. The model is a 2D raster image of the footprint area of buildings and ground with elevation and material specifications (Figure 6.2).

Firstly, the resolution of grid needs to be defined as 6m×6m for the lateral grid and 3m for the vertical grid, and the size of the model space need to be calculated and defined based on the real size of the case and the grid resolution to accommodate the case model. Three more layers of grids will be added to the grid number. This is the requirement of the boundary condition setting for thermal simulation, which is set as open boundary condition in this research. This means the value of the next grid point which is close to the border is copied to the border at each time step.

Secondly, the building footprints are specified on the raster pixels with elevation. Because the vertical grid size is 3m, the same as the building storey height, the vertical grid number of a building is the same value as its storey number.

Thirdly, the location of the ground and road are specified as being comprised of loamy soil and asphalt, respectively. The confirmation of the location of buildings and roads are referencing background bitmaps of the site general plan inserted in the Envi-met editor.

The simulation is run in Envi-met 31, which has three sizes of simulation space for different grid sizes of 100×100, 180×180 and 250×250. To balance the atmosphere model accuracy and simulation time cost, a proper size of simulation space needs to be selected to accommodate the whole case model with three layers of grid at the boundary.
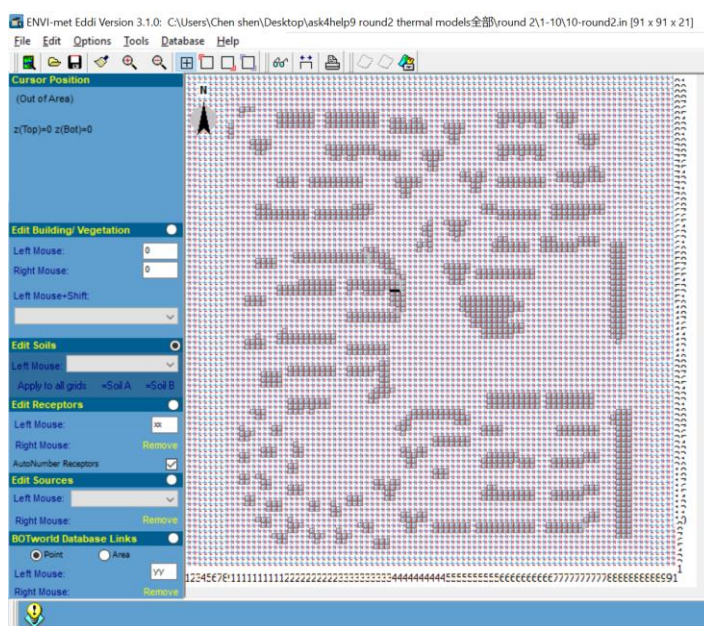
Figure 6.2 Model Space in Envi-met

6.3.2 Simulation Output

All the simulation results are recorded into three groups of data from three calculation models, the atmosphere data, surface data and soil data. Only atmosphere data is used. The first step is the data extraction from the database into an ASIII file for further analysis, using the Xtract program in the Envi-met package. The second step is to export thermal maps of MRT, WS and LRE through the Leonardo mapping tool in the

Envi-met package. The database is recorded at a 60 min time step of a 16h time span. Only data at temporal sections of 14:00 and 16:00 at height of 3m above ground which are of research interest are extracted into tables and maps (Figure 6.3). Example maps are shown in figure 6.3. For the purpose of parallel comparison between sites, the legend scale of WS and MRT maps are kept identical according to the as 0-3m/s, 293-347 K (19.85-73.85°C) and 3.5-343.5 W/m^2, respectively.
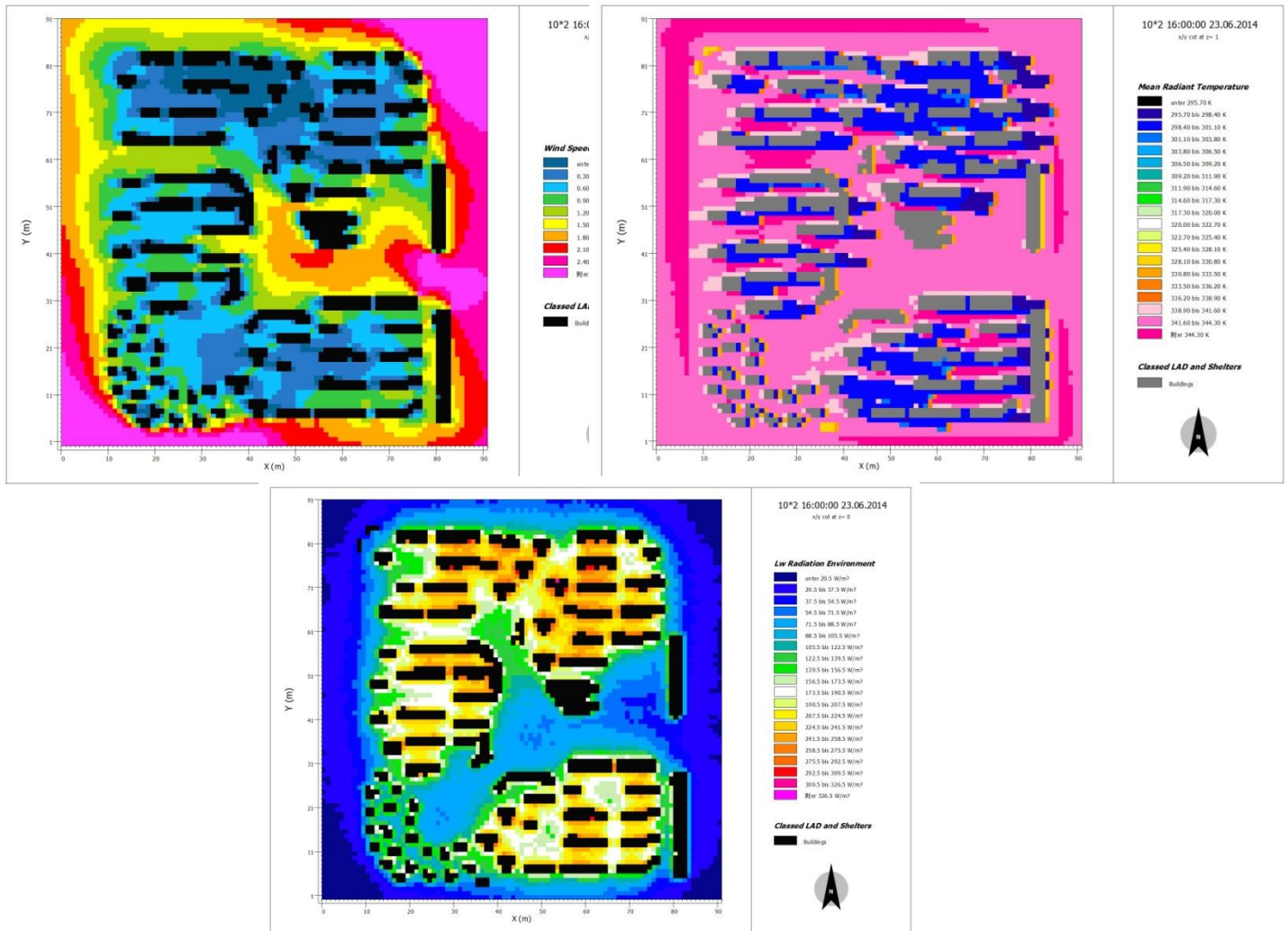
Figure 6.3 WS, MRT and LRE Maps of Site 10

Based on reviews of outdoor thermal comfort and consideration of local climate, the criteria of assessment MRT and WS are 1. As higher wind speed as possible; 2. As lower mean radiant temperature and longwave radiation are possible. Because of low average wind levels in this region and wind blockage due to dense urban structure, it is barely able to form wind turbulence at the foot of high-rise buildings, especially in

residential wards. An early summer day afternoon is the research interest, when low heat stress is expected, namely as low air temperature, mean radiant temperature and longwave radiation is expected.

It worths to mention that the mean radiant temperature (MRT) and longwave radiation from the environment (LRE) are correlated objective variables. By definition, LRE is included in MRT. However considering the research interest, LRE is still studied to provide information from the energy and radiation aspects. Since the interest is only in 14:00-16:00, the shadow cast by residential building blocks would overlay the building footprint. LRE is capable of presenting this information.

6.4 Selection of Thermal Comfort Performance Indices for Assessment and Analysis

6.4.1   MRT Data Comparison at 14:00 and 16:00

Mean radiant temperature (MRT) is a variable that indicates human perception of radiation from their surrounding environment. It strongly affects human perception of temperature along with air temperature. It is originally applied in indoor conditions, so that the surroundings are walls, ceiling, floor and installation of these components. If applied in outdoor conditions, the factors influencing MRT include sky, ground, vegetation, and surrounding buildings. In this research, MRT calculation only includes surrounding buildings, ground and sky in the Envi-met simulation.

Mean radiant temperature is a momentary variable directly under the influence of the shadow cast by surrounding buildings at a certain moment. Because in early summer, air temperature reaches peak value at 14:00 at the specific site location, and outdoor activities start to flourish from 16:00 according to local social habits, momentary values of MRT at 14:00 and 16:00 are adopted in further analysis.

However, compared to MRT, annual possible sunlight hours (APSH) and winter possible sunlight hours (WPSH) are accumulated variably representing the changes year-round. Accumulated variables are more holistic when reflecting the overall performance of certain assessing variables. Therefore, the momentary MRT in early summer is currently applied as a representative assessment in the primitive study of the relationship between thermal comfort and building distribution. In future studies, accumulated MRT measure is preferred instead of momentary MRT, as it considers building shadow sweep in the outdoor activity slot in possibility aspect. The aim of this research is to define and expand the usable area of optimised integrated performances, namely the overlapped area of optimised performance in acoustic, thermal and sunlight, by adjusting building distribution. Therefore, performance assessment indicators in three environmental aspects would prefer accumulated indices representing the possibility of achieving each performance criterion. Therefore, the optimised integrated result would be an overlapped area with the highest overall possibility of good performance in all three aspects.

The building shadow casting at 14:00 and 16:00 tends to have a strong effect on MRT value distribution as the facade and ground materials are fixed presenting no effects on MRT difference. For outdoor activity, occupants in a residential site are capable of selecting areas with high or low MRT according to metabolic rate of the activity. Sites not largely covered by building shadows in activity slots would be less problematic than fully covered by shadows. As excessive building shadow disables an occupant's initiative of selection staying position by MRT, not sufficient building shadows over the ground of residential ward is still improvable by the optimised design of vegetation distribution and sun shading constructions.

The thermal simulation is carried out at 14:00 and 16:00 on 23rd June because, in early summer, peak thermal stress will occur at 14:00 and the main outdoor activities of residential ward habitants will happen around and after 16:00. Therefore, observation of mean radiant temperature and wind speed at 14:00 and 16:00 is most

supportive for the study of the impact from thermal comfort regarding outdoor activities in residential wards.

MRT maps and curves of corresponding sites at 14:00 and 16:00 are compared. The result shows that the map at 14:00 has a shorter shadow from the buildings, namely a higher average MRT due to higher ground surface temperature. The building shadow shown on 16:00 MRT map is cast on an easterly (slightly to south) direction of the building footprint, namely causing a reduction of the average MRT level. It is observable that from 14:00 to 16:00 the building shadow area increases to the east and largely overlaps the building's footprint. Therefore, building shape and distribution resulting in a large shadow area increase rather than highly overlapping their own footprints have a higher potential of achieving a lower average MRT level at 16:00. In other words, there is a better thermal comfort potential at a peak outdoor activity slot.

For easier analysis, statistical measures of MRT from all grid points are adopted: minimum, maximum, percentiles from P10 to P90, and quartiles P25, P75. Clustering of sites based on 14:00 and 16:00 MRT statistical measures are compared. As displayed in two dendrograms of clustering analyses, a similarity shows in the general structure of the four groups. However, at 16:00 all sites have a lower average MRT compared to those at 14:00 due to the reduction of solar radiation. All four groups from MRT 16:00 clustering are upgraded in a parallel sense when compared to the four groups from MRT 14:00 clustering. Despite this, two small groups (group 1 and group 2) of sites appear to upgrade by a larger step than other sites. The two update level groups show a reduction of MRT not only due to less solar radiation but also more importantly due to distribution characteristics.

Curves of MRT statistical measures at 14:00 and 16:00 are compared (Figure 6.4). The curves show that 1) for group 1 and 2 at 14:00, the P10s have reached a high level, namely low level MRT accounts for less than 10% of all data from grid points;

2) for group 1 and 2 at 16:00, the MRT value leaps are located between P20-P30 and P10-P20, respectively. Namely, low level MRT accounts for at least 20%.
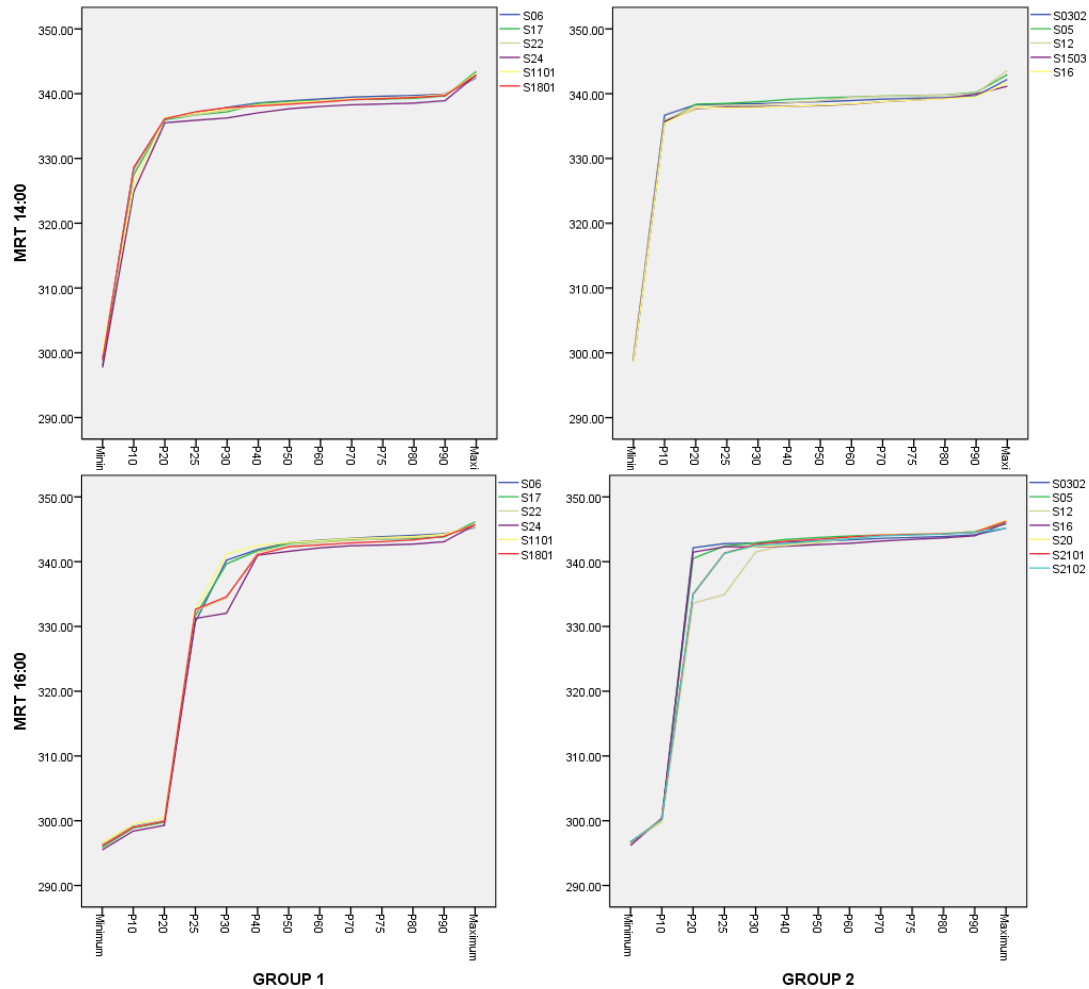
Figure 6.4 Curve Comparison of Upgraded Groups from MRT 14:00 Clustering
to MRT 16:00 Clustering

By observation through all MRT maps and building height maps of group 1 and 2, it can be summarised that group 1 shares distribution characteristics of the low/mid-rise community with few high-rise buildings or with high degree enclosure at the boundary: group 2 shares characteristics of low/mid-rise community with long facade and high density (Figure 6.5). These two groups include cases with especially low MRT level.
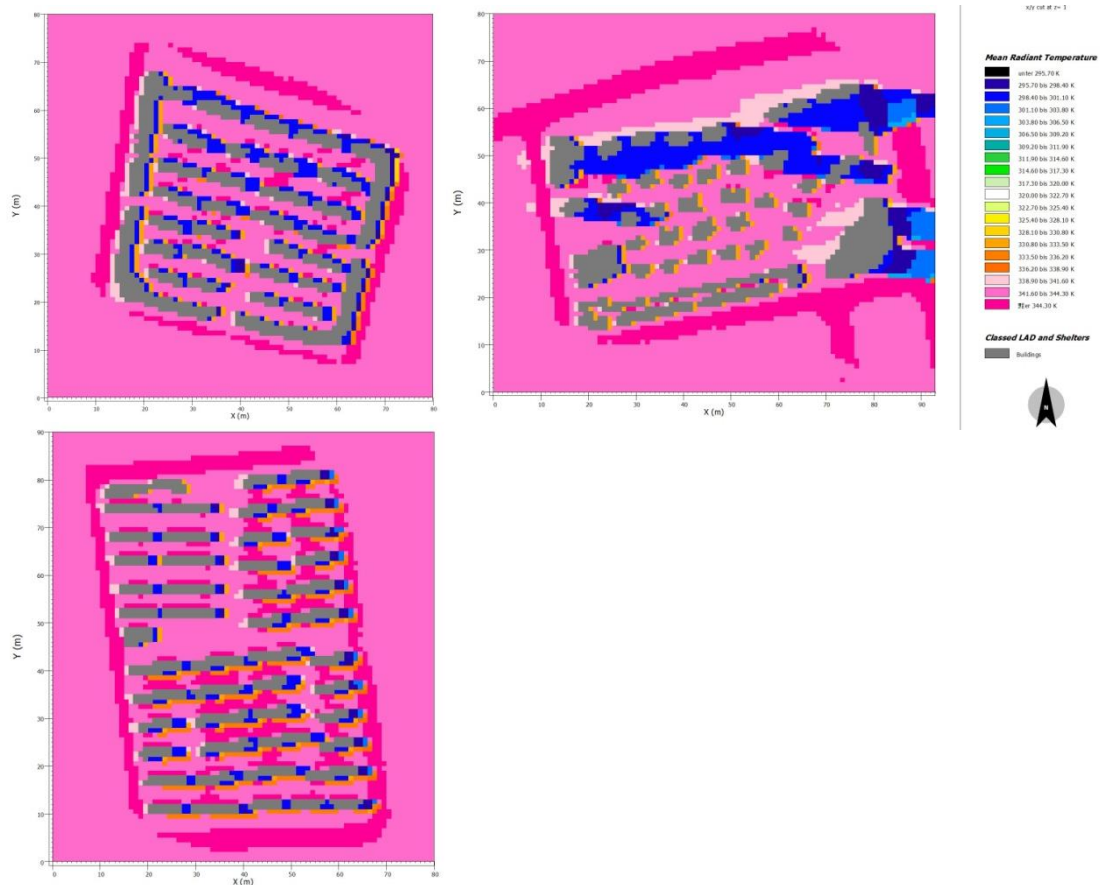
Figure 6.5 MRT 16:00 Map Showing Distribution Characteristics
from Group 1 (S1101 Up-Left, S17 Up-Right) and from Group 2 (S2101 Below)

The MRT value on the grid point over ground at a 3m height will definitely be influenced by surface temperature. Therefore, a building shadow cast on ground leading to a lower surface temperature will also cause a lower MRT value. However, the sunlight availability in the residential outdoor environment prefers as little shadow casting as possible. In summary, when optimising MRT level, sunlight availability needs to be considered in an integrated sense. As shown in MRT maps, MRT is expected to be high on the map of 14:00, but is expected to be low on the map of 16:00.

To summarise, because the objective of this section is to discover the impact of building distribution on mean radiant temperature at 3m at a peak outdoor activity slot, MRT at 16:00 is more suitable to be adopted for further analysis. During the

comparison of 14:00 and 16:00 data, it is discovered that the building distribution pattern resulted in the largest MRT reduction from 14:00 to 16:00: low/mid-rise community with few high-rise buildings or with high degree enclosure at boundary or with long facade and high density.

6.4.2 Statistical Measure Selection of MRT at 16:00

Since MRT data at 16:00 is selected, for simpler data structures, instead of all MRT value from grid points, statistical measures of minimum, maximum, percentiles of P10-P90 and quartiles of P25 and P75 are used to understand the overall data distribution. However, for easier handling regression and optimisation, a representative statistical measure needs to be selected.

Curves of all statistical measures of MRT from all sites appear as a polyline with two flat sections at beginning and ending and a steep leap section in the middle. The lower flat section indicates the area with covered in building shadowing resulting in low MRT value, while the higher flat section indicates the area exposed in direct sunlight resulting in high MRT value. The steep leap indicates the sudden transformation of MRT value crossing the boundary of building shadow. It is apparent that the two percentile at the two ends of the leap in the curve is the key variable worth notice.

As shown in clustering analysis of MRT 16:00 statistical value, the locations of the leaps of various clusters are between P10-P20, P20-P30 and P30-P40. The majority of sites have leaps between P20 and P30. In other words, P20 and P30 are either the beginning point of a leap or the ending of a leap. To enhance the distinguish resolution of the representative variable, P25 is selected as it is capable in distinguish sites with leap just between P20 and P30 from other sites which has leap ending at P20 or starting at P30.

This is also supported by variable clustering of all statistical measures of MRT 16:00. The clusters are 1) Min and P10; 2) P20, P25 and P30; 3) P40-P90, P75 and max.

With the support from clustering analysis of MRT 16:00 and the consideration of distinguish resolution, P25 of MRT 16:00 is selected to be applied in further regression and optimisation as a representative statistical measure. The selection of statistical measure is data dependent, for MRT 14:00 P25 is not applicable as a representative.

6.4.3   Selection of Statistical Measure for Wind Speed

A series of statistical measures are calculated for wind speed data: minimum, maximum, ten percentiles (P10-P90) and two quartiles (P25 and P75). Seen from the curve of all statistical measures, the curve of wind speed data is gradually increased polyline without significant transaction point. Therefore, several representative measures need to be selected to describe the overall trend of the curve. The approaches applied are variable clustering of WS statistical measures, and curve characteristic analysis.

As the result of variable clustering, WS statistical measures are clustered into three groups: 1) P10, P20, P25, P30, P40, P50 and P60; 2) P70, P75 and P80; 3) minimum, 4) maximum. The representatives of measures could be selected from each of the clusters. However, the minimum of each case is almost the same which is close to zero. Therefore, the representatives would be selected from the remaining three clusters.

Referencing the characteristics of curves (figure 6.6), key statistical measures are observed. P20 is the first turning point on the curve from a steeply increased polyline transforming to a slow polyline. P70 is the start point where the second clear transformation of the polyline gradient occurs.
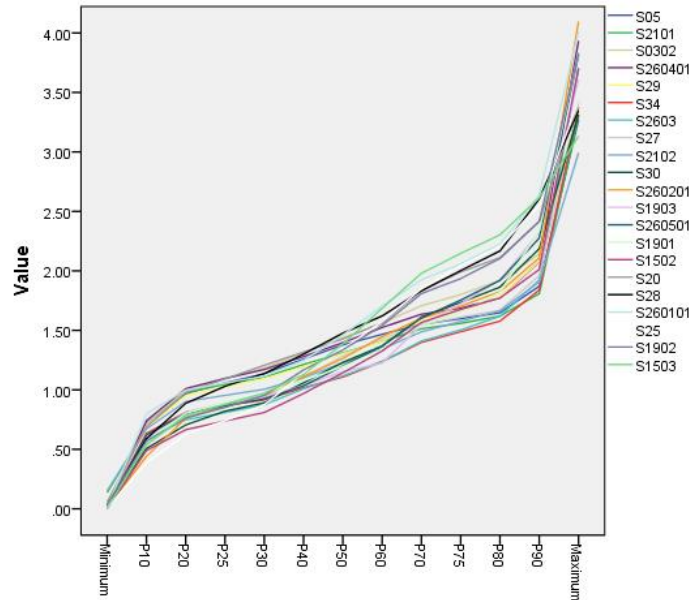
Figure 6.6 Curves of Statistical Measures of Wind Speed by Clustering Results

The maximum of WS indicates the degree of wind speed increase due to possible vortex and turbulence and it should be limited within a limit of health hazard. However, by observation of wind speed data, in this researched urban environment no excessively high wind speed exists. The criterion for maximum wind speed is expected to be as high as possible for the purpose of air-based contamination removal in a residential ward.

To summarise, P20, P70 and maximum of wind speed data are selected to describe the overall data distribution of each case.

6.4.4   Summary

Based on the discussion of MRT data at different time slots, and the comparison of various statistical measures, MRT at 16:00 is selected for further study, and P25 of MRT 16:00 is used as a representative measure.

Similarly for wind speed data, P20, P70 and maximum value is selected as representative measures to describe the overall data distribution of wind speed.

It is also noticed during discussion that the building distribution pattern which could result in the largest MRT reduction from 14:00 to 16:00 is low/mid-rise community with few high-rise buildings or with a high degree of enclosure at boundary or long facade and high density.

6.5 Qualitative Analysis of MRT Simulation Result

6.5.1   Method and Discussion of Clustering Site Samples by Statistic Measures of Individual Thermal Comfort Indices

Because the residential wards are affected by a mixture of variables, the outdoor environmental performances also show great discrepancies together with similarities. To make the analysis on thermal performance and its relationship with distribution more accurate and more powerful in future predictions of different kinds of site, it is necessary to categorise sites based on a clustering calculation through SPSS by key variables.

Clustering of sites is calculated based on WS, MRT and LRE characteristics. These characteristics are presented in the form of descriptive statistical values of percentiles from 10 to 90, quartiles, minimum and maximum, as well as all directly available grid point values. For each of the three variables, three attempts of clustering are executed using data on all grid points, P10-P90 values (including P25 and P75), or P10-P90 with minimum and maximum values. Hierarchical clustering is applied for site

categorising: case cluster method is applied for the condition of using data on all grid points, while variable cluster method is applied for the condition of using descriptive statistical values. Between-groups linkage calculation method is applied for the clustering with calculation measure of squared Euclidean distance. Output of the clustering is a dentrogram and a table of hierarchical clustering procedure.

Based on the analysis of the dendrogram from the sample cluster by grid values of three variables, cluster by value is invalid and abandoned for further analysis. The reason is that its dendrogram shows a larger combining distance between groups for the first step than the later steps which is against the methodology of hierarchical clustering.

Through between-group comparison of the several clusters calculated by each of the three variables, their thermal performance differences could be identified on maps, and different distribution characteristics could be specified on the histogram of certain variables. Discussion is also made about the discordance between clusters by different thermal variables: WS, MRT and LRE. An integrated cluster based on P10-P90 of the normalised data of all three variables is also calculated to discover the difference in influential power between the three variables. However, the integrated clustering result is almost identical to that of MRT on its own. This is as a result of the data distribution of MRT, which is dramatically different from the data distribution of WS and LRE. So the influence of WS and LRE on clustering is heavily covered by MRT data. As a result, the integrated clustering by WS, MRT and LRE is not further discussed here.

Clustering of site by building distribution indices are also calculated as a comparison to clustering by thermal performance variables. The discrepancies between clusters by building distribution and thermal performance variables are discussed. Based on this discussion, qualitative design guidance is suggested in reference to the strength of

correlation connection between building distribution and thermal performance variables.

6.5.2 Consolidation of Clusters by MRT 16：00 and Summary

Clustering of sites by MRT 16:00 data applies a series of statistical measures: percentiles of P10-P90, quartiles of P25 and P75, minimum and maximum. Clustering is executed in two approaches: with minimum and maximum or without. The results are identical. The group condition is shown on the dendrogram of clustering (figure 6.7). Three types of sites are divided by MRT characteristics.

The clustering result highly accords with the curve pattern of the statistical values of MRT data. The curve appears in a polyline with three sections. The first and the last sections are generally flat, indicating the MRT in the shadow area and exposed area, respectively. While the steep leap section between the two flat section refers to the sudden change of MRT when crossing the boundary of shadow. The more right the leap locates, the higher proportion the lower MRT value accounts for; namely a higher potential of outdoor thermal comfort in a residential ward.

Cluster A of good performance is the group with a leap on the curve of statistical measures located between P30-P40 or P40-P50. Cluster B of fair performance and cluster C of poor performance are groups of sites with leap located at P20-P30 and P10-P20, respectively.

The objective of this section are 1) To understand MRT data distribution characteristics and key turning points on the curve through clustering; 2) to understand how building distribution affects MRT level; 3) to summarise qualitative design rules of MRT performance and building distribution.
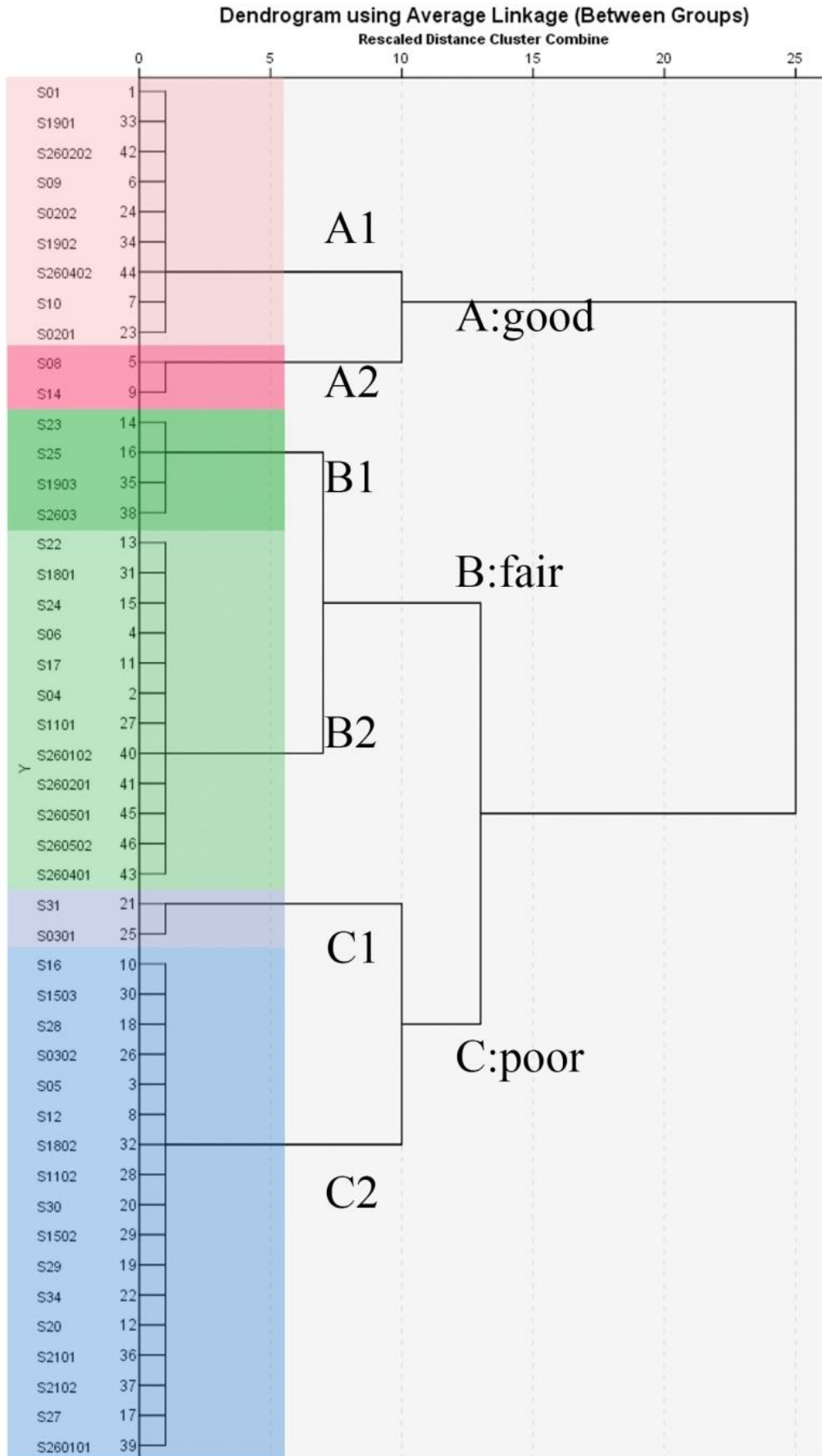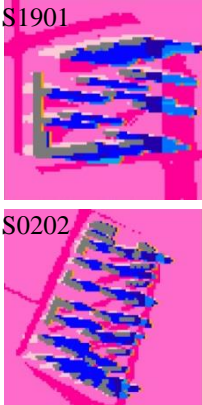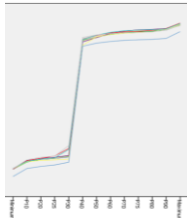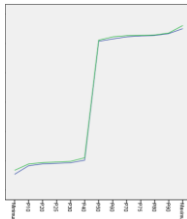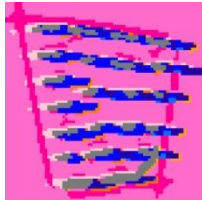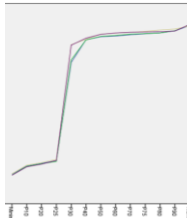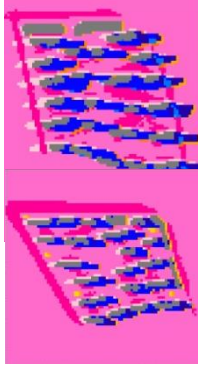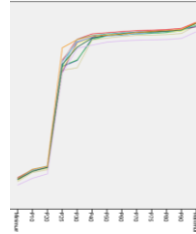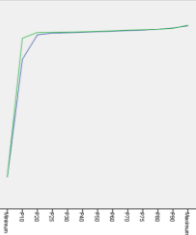
Figure 6.7 dendrogram of clustering of sites by MRT 16:00 statistical measures

The details of the clustering by mean radiant temperature are summarised in table 6.1.

Table 6.1 Summary of MRT 16:00 Clustering Results and Derived Rules of Design

Cluster	Sub-groups	Site Description	Example	MRT map	Curve and Characteristics	conclusion
A: Good	A1	High plot ratio, 100% high-rise ratio, Clustered, Open boundary or relative enclosed	S01, S1901, S260202, S09, S0202, S1902, S260402, S10, S0201	S1901 S0202	MRT level leaps at P30-P40	1) high plot ratio, clustered, open boundary or relative enclosed, will lead to high proportion (30%) of small MRT value 3) Higher plot ratio is most dominant in reducing MRT 4) Orientation of buildings manifest a stronger impact on MRT than on accumulated sunlight hours, due to the momentary characteristic of MRT
	A2	Highly enclosed with extreme plot ratio or with extreme density, evenness and medium plot ratio	S08, S14	S08 S14	MRT level leaps at P40-P50	2) Highly enclosed site with extreme plot ratio and extreme density and medium plot ratio will lead to very high proportion (40%) of small MRT value, but may cause sunlight availability problem. 5) win-win on MRT and sunlight availability is achievable with proper arrangement of distribution: high PR, low density, south on-edge open space and east-west orientation.
B: Fair	B1	medium-high plot ratio, medium density, medium enclosure, medium SVF medium-high high-rise ratio	S23, S25, S1903, S2603	S2603	Leap at P25-P30	1) Medium level distribution results in satisfactory performance in MRT and sunlight, namely no violation but also no distinction.

	B2	low plot ratio, low-medium distance to road, medium-high max facade length, and attribute pairs (high density with low high-rise ratio or low-medium density with high high-rise ratio)	S22, S1801, S24, S06, S17, S04, S1101, S260102, S260201, S260501, S260502, S260401	 S260502 S06	 Leap at P20-P25	2) Rows of buildings lying along direction from west-by-north to east-by-south will reduce the shadow area and proportion of small MRT value relatively due to shadow overlapping 3) low density high-rise community with large on-edge open space could balance good sunlight performance and fair MRT performance
C: Poor	C1	low plot ratio, low building density and zero high-rise ratio	S31, S0301	S0301 	 Leap at minimum-P10, P10 reaches to relatively high value of 330-340K (56.85-66.85℃)	1) low-rise sites in low density, even some in high-density, tend to have poor MRT performance due to short shadow from building. 2) high-rise buildings may have receded contribution to MRT performance within a certain site, due to the site shape being narrow in east-west direction.
	C2	low plot ratio, low-medium density, low facade ratio, low max facade length	S16, S1503, S28, S0302, S05, S12, S1802, S1102, S30, S1502, S29, S34, S20, S2101, S2012, S27, S260101	S05 	Leap at P10-P20	

Category	Sub	Description	Sites	Plan image	MRT curve	Notes
	B2	low plot ratio, low-medium distance to road, medium-high max facade length, and attribute pairs (high density with low high-rise ratio or low-medium density with high high-rise ratio)	S22, S1801, S24, S06, S17, S04, S1101, S260102, S260201, S260501, S260502, S260401	S260502 S06	Leap at P20-P25	2) Rows of buildings lying along direction from west-by-north to east-by-south will reduce the shadow area and proportion of small MRT value relatively due to shadow overlapping 3) low density high-rise community with large on-edge open space could balance good sunlight performance and fair MRT performance
C: Poor	C1	low plot ratio, low building density and zero high-rise ratio	S31, S0301	S0301	Leap at minimum-P10, P10 reaches to relatively high value of 330-340K (56.85-66.85℃)	1) low-rise sites in low density, even some in high-density, tend to have poor MRT performance due to short shadow from building. 2) high-rise buildings may have receded contribution to MRT performance within a certain site, due to the site shape being narrow in east-west direction.
	C2	low plot ratio, low-medium density, low facade ratio, low max facade length	S16, S1503, S28, S0302, S05, S12, S1802, S1102, S30, S1502, S29, S34, S20, S2101, S2012, S27, S260101	S05	Leap at P10-P20	

Based on the rules of design, it is worth noting that the building distribution resulting in balanced MRT and sunlight performance is encouraged and results in good MRT and poor MRT performance.

The rule discovered from the clustering analysis of the mean radiant temperature at 16:00 is stated in three levels: macro, meso and micro.

Macro level:

1. Higher plot ratio is most dominant in reducing mean radiant temperature.

2. The requirement on building distribution by low MRT and high APSH and WPSH are opposite but not in opposition, so that a win-win of good sunlight and MRT performances is achievable with the proper arrangement of distribution characteristics.

3. Medium level distribution results in satisfactory performance in MRT and sunlight; namely no violation but also no distinction.

4. A novel index of accumulated MRT in a time slot verified by outdoor activity study is suggested to be created for future work, to enhance its summary power over thermal comfort in the activity period.

Meso level:

1. Balanced combination of building distribution characteristics leading to good sunlight performance and good-fair MRT performance includes: High-rise community in low density, surrounded by large on-edge open space (especially on the south), and the proper amount of east-west orientated buildings of considerable height.

2. Orientation of buildings manifests a stronger impact on MRT than on accumulated

sunlight hours, due to the momentary characteristic of MRT.

3. Building orientation optimisation regarding outdoor thermal and sunlight comfort, surface temperature on facade, mean radiant temperature from facade on vertical grid are suggested to be supplemented in future work.

Micro level:

1. Best combination: if integrated, high plot ratio, clustered distribution, open boundary or a relatively enclosed, large residential building area will lead to high proportion (30%) of small MRT value, namely a cooler outdoor environment.

2. Extreme combination: Highly enclosed site with 1) extreme plot ratio or 2) extreme density, evenness and medium plot ratio will lead to very high proportion (40%) of small MRT value, but may cause sunlight availability problems.

3. Combination causing the largest MRT reduction: low/mid-rise community with 1) few high-rise buildings or 2) with high degree enclosure at boundary or 3) with long facade and high density could result in largest reduction of MRT from 14:00-16:00, namely the dramatic transformation from sufficient sunlight potential to strong shadow protection during the activity slot.

4. Site shape: high-rise buildings may have receded contribution to MRT performance within a certain site, due to the site shape being narrow in an east-west direction which caused building shadow at 16:00 exceeding site boundary.

5. Direction of building rows: rows of buildings lying along the direction from north-west to south-east, although the building orientation is north-south, will relatively reduce shadow area due to shadow overlapping; therefore reducing the proportion of small MRT value.

6. Worst combination: low-rise community in low density, even some in high-density, tends to have poor MRT performance due to short shadow from buildings.

6.6 Qualitative Analysis of Wind Speed Simulation Result

6.6.1   Consolidation of Clusters by Wind Speed

Clustering by wind speed is executed in two approaches using P10-P90 and using P10-P90 with minimum and maximum included. The cluster result shows a tiny difference between the two approaches: only 4 sites out of 46 are located to different cluster when considering minimum and maximum, compared to only using P10-P90. Based on the observation of wind speed maps at 16:00, minimum wind speed is sensitive to detailed building shape, and maximum wind speed is sensitive to the empty space outside of the road-facing buildings where air initially circulates. To keep the focus on the influence from building distribution on wind speed, cluster by P10-P90 is used, by which impact of extremities of wind speed on clustering sites is screened out.

Base on P10-P90 of wind speed, 4 clusters are generated (Figure 6.8). Cluster D has very good wind speed conditions; cluster A is good; cluster B is fair and, finally, cluster C is poor.
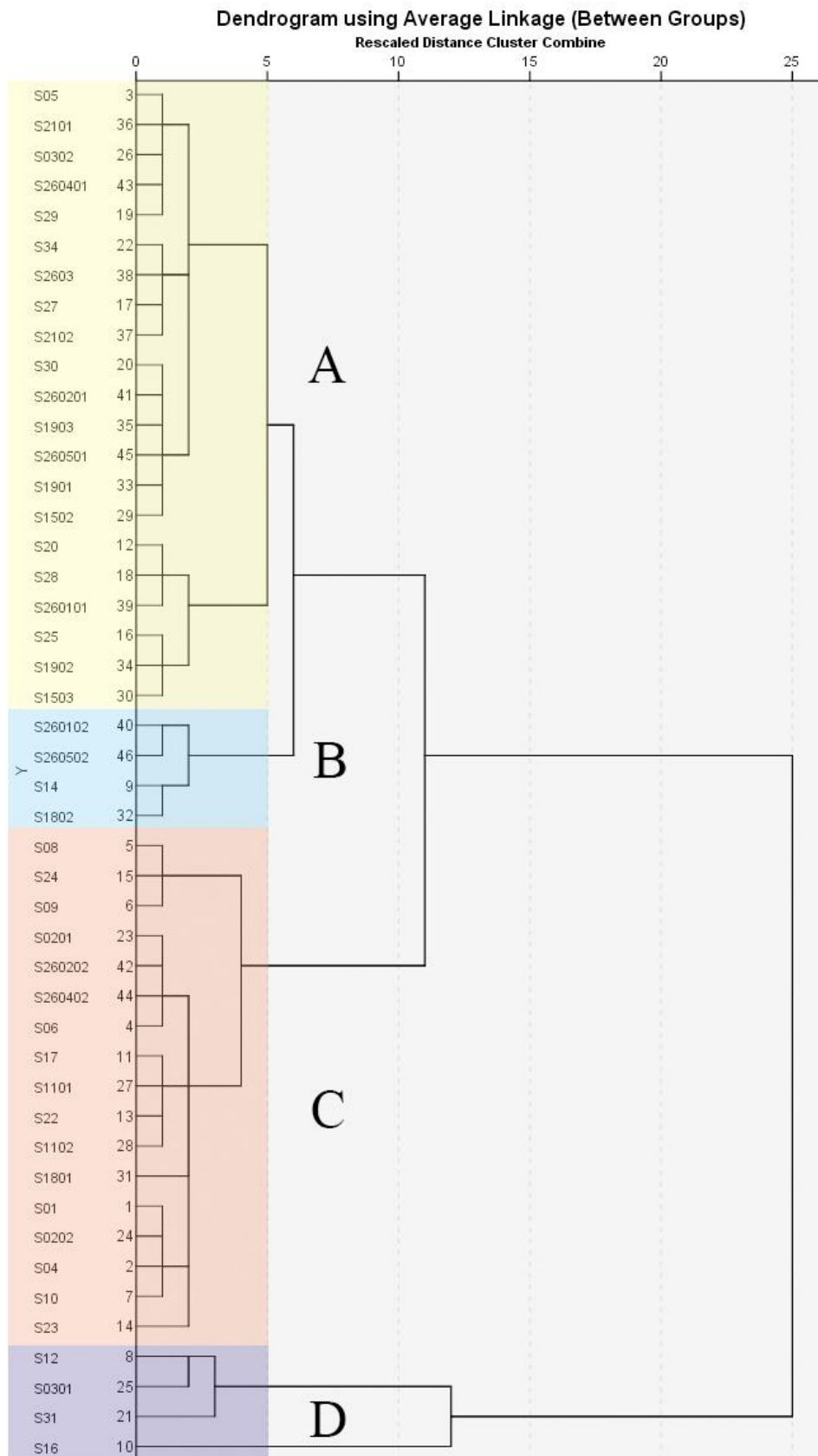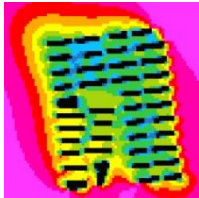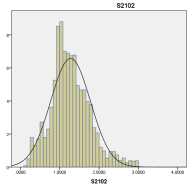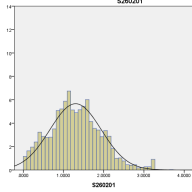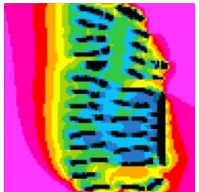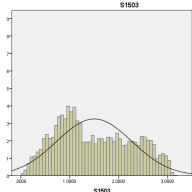
Figure 6.8 Dengrogram of Site Clusters by WS P10-P90
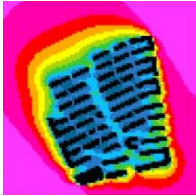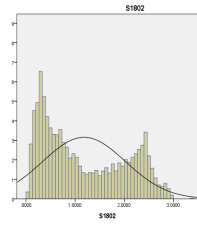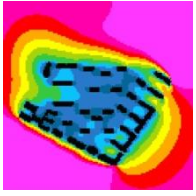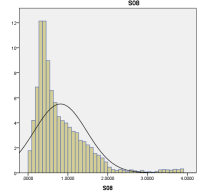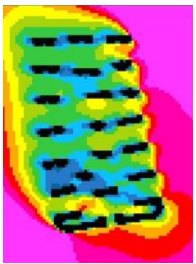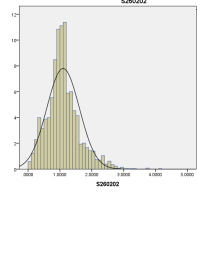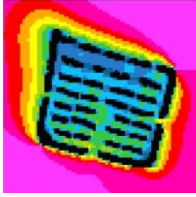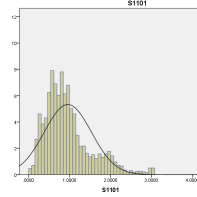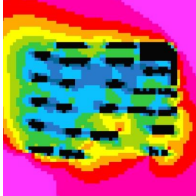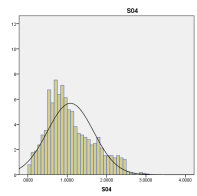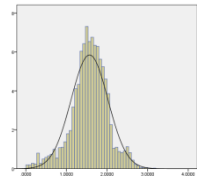
6.6.2 Summary

The objective of this section is to discover 1. The key character and category of wind speed distribution of different sites with various building distribution based on the study on wind map and data statistics; 2. To explore how building distribution characteristics will influence wind speed distribution; 3. Provide qualitative design guidance based on the defined relationship.

In this section, analysis about characteristics of clusters by wind speed is expounded together with possible qualitative design strategies. An overview of all clusters is shown in table 6.2.

Table 6.2 Overview of Clustering by Wind Speed

Cluster	Sub-groups	Site Description	Example	Wind map	Histogram	Characteristics	conclusion
A: good	A1	Less dense, low/mid-rise	S2101, S2102, S29	S2102		Single peak, Peak axis around 1-1.3m/s, Thick right tail	1 In low/mid-rise, large building separation, in-site open space and small size help with better wind speed; 2 In high-rise, small size helps on low and high wind speed level, but not on middle wind level
	A2	High-rise, Open boundary	S1903, S260201	S260201			
	A3	Segmented open space wrapping part of site	S1503, S20	S1503		Similar to cluster B, Two peaks merge into trapezoidal	1 open space at south and north leads to bipolar 2 open space at west and east or small space at north lead to

B: Fair and Bipolar	N/A	Open space at south edge, Or large open space at north edge	S26050 2, S1802	S1802			Two peaks, Left peak for poor wind area, Right peak for high wind open space	trapezoidal histogram, 3 less-left-skewed wide single peak is good sign for new design; bipolar histogram is good sign for site need improvement in wind performance
C: Poor wind speed	C1	High-rise, enclosed	S08	S08			Left skewed single peak; Peak axis close to 0; Medium tail.	1 When enclosed, low/mid-rise is better; When less enclosed, high-rise is better; 2 Decreasing enclosure works better for high-rise than low-rise; 3 decreasing enclosure helps for low wind level in high-rise, but helps for high wind level in low/mid-rise site; 4 Decreasing enclosure is more effective than decreasing building height.
	C2	High-rise, less enclosed	S26020 2	S260202				
	C3	Mid-rise, enclosed	S1101	S1101				
	C4	Mid-rise, less enclosed	S04	S04				
D: Very good	N/A	Very small, or very open/or large open space on the south/east	S12, S31, S0301, S16	S0301			Peak axis 1.5-2.5 mi/s; Concentrated single peak.	Small, even, open site has very good wind performance if no large building mass at upper wind direction

Both very good and really poor wind speed performances are possible for low and mid-rise community and high-rise community.

To avoid worse wind performance, a good balance of multiple building distribution attributes is required. Qualitative rules discovered from analysis are extracted from the conclusions. Rules are provided in three levels: macro level about holistic guidelines, meso level about influential index selection and micro level about attribute pairs combination.

Discovered rules:

Marco level:

1. The bipolar histogram could not be applied as a sign or criterion of good overall wind speed performance.

2. For new built residential wards, building distribution of a histogram with less-left-skewed wide single peak is more encouraged.

3. For existing sites in need of improvement in wind speed and new design limited by high development intensity building distribution of non-segmented open space with bipolar histogram is encouraged.

Meso level:

1. High-rise community is considered better than low-rise community, when the site is less enclosed, due to higher wind speed in low wind speed level. The critical statistical value of wind speed is P60 when comparing low/mid-rise and high-rise site under less enclosure.

2. Low/mid-rise community is better than high-rise community when the site boundary is enclosed, due to higher overall wind speed.

3. High-rise community increases more overall wind speed than low/mid-rise community when there is a decrease in the site boundary enclosure.

4. Decreasing enclosure at the boundary is more powerful than decreasing building height to raise overall wind speed.

5. Enlarging open space inside a site increases overall wind speed more than enlarging building separation does in low and mid-rise sites.

6. Smaller and larger site sizes will lead to different wind speed distribution in designing a high-rise community. Selection will be balanced according to what is emphasised in the design, e.g., breezy playground or peaceful and fresh neighbourhood atmosphere. The critical statistical value is P30 and P75 when comparing small and large sites of high-rise buildings.

Micro scale strategies:

1. Best combination: Smaller, even, open distributed, open boundary and no large building mass blocking the dominant wind direction. A combination of these attributes leads to very good overall wind speed performance.

2. Decrease enclosure: This causes an increase in wind speed in the low wind speed level for high-rise sites and causes an increase in wind speed in high wind speed level for low and mid-rise sites.

3. Smaller site size: This increases overall wind speed in low and mid-rise sites.

4. Larger building separation and open space inside site: Increases overall wind speed in low and mid-rise sites.

5. Open space location: Open space on the south and east side of the sites generates a clearer bipolar histogram of wind speed for a site, than when located on the north and west side.

6. Building mass and orientation: Long facade and large building mass within 90

degrees (no matter facing or on the side) of upwind direction causes a fast wind speed decrease behind and in front of building.

7. Worst combination: High density, low-rise, long facade array perpendicular to wind direction, combination of these attributes leads to the worst overall wind condition. E.g. 2101 2102(A) 1802(B) 1801(C)

A combination of attributes and their effects on wind condition and on the wind map histogram are also listed in table 6.3. The number in the table indicates the order of the pair.

Table 6.3 List of Attribute Pairs and Their Effect

		High overall wind speed			Histogram appearance		
		Low wind speed level	Middle wind speed level	High wind speed level	Left skewed single peak	bipolar	Right skewed single peak
Size	Small site	1,7	1,7	1,7			
	Large site						
Homogeneity	Even distribution	1	1	1			
	Clustered distribution						
Density	Dense separation						
	Loose separation	1,6	1,6	1,6			
Open space existence	Open space inside site	6	6	6			
	Open space on dominant wind direction					9	
	Open space wrapped around site						
Boundary enclosure	Open boundary	1	1	1			
	Less enclosed	4		5			
	Highly enclosed	3	3	3			
Facade length	Long facade						

	Short facade	2	2	2			
Building orientation	Perpendicular to dominant wind						
	Parallel to dominant wind	2	2	2			
Building height	Low/mid-rise	1,3,6,7	1,3,6,7	1,3,5,6,7			
	mix-rise						
	High-rise	4					

6.7 Quantitative Analysis of Statistical Measure of Mean Radiant Temperature and Building Distribution Indices

6.7.1 Regression Parameter Selection

The selection of regression indices includes two aspects: performance assessment and building distribution aspect. To represent the key characteristics of mean radiant temperature data, P25 is selected based on the discussion in parametric study of MRT in 7.3.2. P20 and P30 are the key positions where the dramatic leaps on the curves of MRT data occur. P25 locates in the critical position and is capable of distinguishing the leap ends at P25, started at P25 (end at P30) and started at P30. Therefore, in the regression of mean radiant temperature P25 is applied.

P25 of MRT is the first quartile or 25% of all MRT grid value of a certain site. Its unit is K and in a range of 297.59-342.90K (24.44-69.75 °C). It will be noted as MRT-P25 in following analysis.

To describe the distribution of buildings, a series of indices are selected according to the parametric study of distribution parameters in chapter 4 (table 6.4). In the case of unsatisfactory regression results, all distribution indices may still be applied in regression attempts as a cross validation approach to compare to the regression with representative distribution indices.

Table 6.4 Representative Building Distribution Indices for Regression

Group	Representatives	Abbreviation
Site scale	Total land area	TLA
Built amount	Residential building area	RBA
Existence of high-rises	TriangleSD	TSD
	Plot ratio	PR
Density	Building density	BD
	Average facade ratio	aFR
Set back	Average distance to road	aD
Site shape	Shape factor	SF
Low and long facade amount	Average low facade ratio	aLFR
Short facade amount	Average min facade length	aFLmin
Large interval amount	Average interval area	aIAmean
Small interval amount	Average lower corner area	aCAL
Interval depth amount	Average higher interval depth	aIDH

6.7.2   Regression Result of P25 of Mean Radiant Temperature at 16:00

P25 value of mean radiant temperature refers to the value of one grid point, of which 25% of all points have a value lower than this particular value. As understood from the parametric study of mean radiant temperature, P25 locates at the critical position where the leap on the curve of data happens. The lower section indicates the area of site ground with low MRT under impact of building shadow casting at 16:00 on the simulation day. Meanwhile the higher section indicates the area of site ground without thermal protection from building shadows. Considered from a thermal comfort aspect in early summer, which is the simulated time, the higher proportion of the lower section on the curve and smaller the value of lower section, the better thermal comfort potential it represents. Therefore, for better thermal performance potential, P25 is expected low.

The regression of MRT-P25 has been attempted twice with distribution representative indices and all distribution indices. The regression with representatives shows insufficient power of explanation to the variation of MRT, therefore another attempt with all indices is tested for an alternative result. The comparison shows regression

with all distribution indices involving plot ratio and average lower interval depth at the front row of buildings, namely one index more than regression with representative indices (only involve plot ratio). Therefore the regression with all distribution indices is applied for further analysis.

The regression model of MRT-P25 has R square of 0.582 and adjusted R square of 0.561, with a difference of 3.6% in generalisation of the model in all population. The degree of explanation by the MRT-P25 model is only satisfactory, so more predictor is required.

Details of the regression of MRT-P25 are listed in table 6.5.

Table 6.5 Details of Regression of MRT-P25 with All Possible Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	345.211	9.385		36.783	.000
plot ratio	-20.367	3.129	-.676	-6.508	.000
AvgIntervalDepthLow	.682	.281	.252	2.429	.020

As shown in table 6.5, to achieve a lower MRT-P25, higher plot ratio and small average low interval depth is required. Higher plot ratio will cause more shadow casting from the building at 16:00, and this echoes the clustering result by MRT statistical measures. Small average low interval depth actually describes a distribution pattern showing that the building's long facades are parallel to the site boundary. Therefore, the interval depth will be the gable wall of the building, namely a relatively short wall. This pattern allows east-west facing buildings which will cast wide shadow at 16:00 and the least amount of overlapped area between the building shadow and the footprint of the same building. This also echoes the design rule concluded from MRT clustering: building orientation is influential on MRT performance, and the building casts a large shadow and has less overlapped area with

footprint. This could lead to a lower MRT level, for example, on east-west facing buildings.

However, the preferred residential building orientation is 90 degrees of due south. A balance needs to be made between building performance and outdoor performance according to requirement and expectation.

6.8 Quantitative Analysis of Statistical Measures of Wind Speed and Building Distribution Indices

6.8.1 Regression Parameter Selection

The independent variable of building distribution adopted in the regression is identical as that for mean radiant temperature. The selected distribution indices are decided based on the parametric study of building distribution in Chapter 3. Detail of representatives of indices refers to table 6.4.

The regression dependent variables are selected wind speed statistical measures. According to the parametric study of wind speed measures, P20, P70 and maximum are selected for the reason of holistically representing the wind speed data distribution. The three measures are key turning points on the curve generated by wind speed measures of each sites from low to high level. It could be ensured from grading to WS-Max that maximum wind speed is not dramatically increased due to vortex and the turbulence of high-rise buildings which may cause hazard for residents. As a result, in this research, P20, P70 and maximum are expected to be as high as possible.

Table 6.6 Summary of Statistical Measure Selection for APSH Data

Parameter	Abbreviation	unit	range		Definition	Reason
P20	WS-P20	m/s	0	1.51	20% percentile of WS grid value of a site	first turning point of curve from steep to slow gradient
P70	WS-P50	m/s	0	2.56	70% percentile of WS grid value of a site	second turning point of curve from slow to steep gradient
Maximum	WS-Max	m/s	0	4.18	Maximum of WS grid value of a site	Maximum and end of the curve. Key value of WS

6.8.2 Regression Result of P20 of Wind Speed at 16:00

Regression of P20 of wind speed is carried out with representative building distribution indices. A validation is also operated by regression with all distribution indices. The result of regression with representatives are reasonable and acceptable, therefore, it is adopted for further discussion in this research.

Regression of P20 involves average low facade ratio, residential building area and average high interval depth. These indices reflect the impact from distribution factors of boundary enclosure, site size and building orientation at boundary, respectively. The R square and adjusted R square are 0.542 and 0.507, respectively, with shrinkage of 6.5% for generalisation. The power of explanation on wind speed P20 variation is only satisfactory, so further study is needed to increase sample size and proper distribution index.

Table 6.7 Details of Regression of P20 with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.901	.130		6.925	.000
AvgLowFacadeRatio	-.005	.001	-.462	-4.222	.000
residential building area	-.013	.003	-.563	-4.847	.000
AvgIntervalDepthHigh	.009	.004	.297	2.541	.015

As seen in table 6.7, regression of P20 is negatively correlated to average low facade ratio and residential building area and positively correlated to average high interval depth. Seen from standardised coefficients, RBA has the highest predictive power. The regression shows that to increase low level wind speed, it is necessary have reduced boundary enclosure, smaller site size and a considerable number of buildings with long facades perpendicular to site boundaries.

6.8.3   Regression Result of P70 of Wind Speed at 16:00

Regression of P70 applied is with representative distribution indices. The criterion of P70 is expected to be as high as possible. The regression model has a R square of 0.457 and adjusted R square of 0.415, with a shrinkage of 9.2%. Details of coefficients are listed in table 6.8.

Table 6.8 Details of Regression of P70 with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.683	.141		11.895	.000
AvgLowFacadeRatio	-.005	.002	-.405	-3.328	.002
residential building area (㎡)	-.012	.003	-.444	-3.619	.001
AvgDistance	.009	.003	.322	2.558	.015

As shown by standardised coefficients, P70 is negatively correlated to average low facade ratio and residential building area which is identical to P20 regression, and it is positively correlated to average distance to road. Average low facade ratio and residential building area have a similar predictive power on P70 regression. The tendency could be summarised as, in order to increase P70 it is suggested to have less enclosed boundary, small site scale, and large separation distance between street facing facades and streets.

6.8.4   Regression Result of Maximum of Wind Speed at 16:00

Due to a dense residential ward, excessive maximum wind speed can barely form inside a residential ward. Therefore, maximum wind speed is expected as high as possible in the analysis.

The regression of WS-max has an R square of 0.427 and adjusted R square of 0.398 with shrinkage of 6.8%. Details of regression are shown in table 6.9. WS-max

positively correlated with triangleSD and residential building area. Therefore, to achieve high WS-max value, it is suggested to have clustered distributed buildings and a large site scale. The requirement on site scale is reserved to that based on P20 and P70 regression. Therefore, a balance is needed in design according to requirement priority.

Table 6.9 Details of Regression of Maximum with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	2.968	.097		30.471	.000
TriangleSD	.001	.000	.485	3.855	.000
residential building area (m^2)	.009	.004	.313	2.491	.017

6.8.5   Comparison of Wind Speed Regression Model

Models of three wind speed regressions are listed in table 6.10. As shown, the degree of explanation of the equations is limited to satisfactory. Sample size expansion is required to allow further enhancement on the number of independent variables. The possible distribution index to add in as independent is the building orientation indicator.

Table 6.10 Regression Equation Comparison between P20, P70 and Maximum Regression

Model	Regression Equation	R^2	Sig
P20 With Selected	WS-P20=0.009×aIDH-0.005×aLFR-0.013×RBA+0.901	0.542	.000
P70 With Selected	WS-P70=0.009×aD-0.005×aLFR-0.012×RBA+1.683	0.457	.000
Max With Selected	WS-max=0.001×TSD+0.009×RBA+2.968	0.427	.000

By comparing P20 and P70 regression, P20 is positively correlated to an average high interval depth, while P70 is correlated to average distance to road. This indicates that building facade parallel to road tends to block wind flow and form low level wind speed area behind the buildings. Therefore, building long facades perpendicular to road causing a high value of aIDH would increase wind speed value in low range. Then, P70 is more positively correlated to average distance to road, namely, large open space on the edge of site tends to improve high range wind speed value.

According to regressions, smaller low facade ratio and site scale (represented by low RBA) could contribute to higher P20 and P70.

Regression of WS-max shows reverse correlation with RBA compared to that of P20 and P70. Large scale site tends to have a higher maximum value of wind speed. However, smaller sites would have higher low and high range wind speed. Maximum of wind speed also positively correlates to spatial openness of the site (represented by high SVF).

6.9 Qualitative and Quantitative Analyses of Longwave Radiation from Environment

6.9.1 Qualitative Analysis by Clustering of Statistical Measures of LRE

Longwave radiation from environment (LRE) is another thermal comfort output from the Envi-met simulation of the atmosphere model. It indicates long wave radiation from all possible sources in the environment, which refers to all residential buildings and other buildings in a site in this research. The output of LRE includes LRE maps and data at all grid points of a site. For easier application, the statistical measures of LRE data are used: minimum, maximum, percentiles of P10-P90 and quartiles of P25 and P75. Considering the definition of LRE, low LRE is expected for good thermal comfort.

Cluster analysis of sites by LRE characteristics is operated. Five clusters are grouped of very good, good, fair, poor and very poor LRE performance. Matching the clustering result with site LRE maps and the histogram of LRE data of each site, characteristics of each cluster are consolidated. The LRE map shows higher LRE around building facades in brighter warm colours, and shows as green and blue in open space where it has received less impact from buildings which are the only radiative sources. The histogram of LRE data appears in either single modal or

bimodal. Single modal is a special case of bimodal with the absence of the left peak. The right peak indicates the accumulated frequency of area having high LRE under influence of buildings. The left peak is the result of areas with low LRE due to distant separation from buildings where it lies in open spaces either in-site or on-edge.

The 'very good' cluster A appears bimodal in the histogram, specifically two peaks appear. The location of the right peak is between 100-150 W/m^2. The location of the left peak is at 50 W/m^2 due to the large in-site open space. Therefore, the left peak is merged into the right one. The left peak is much taller than the right peak.

Similarly, for the 'good' performance cluster B, there are two sub-groups. A single peak appears in the histogram of sub-group B1, with a location of 100-150 W/m^2. The sites have neither on-edge open space nor large intact in-site open space, so the left peak is absent. For sub-group B2, its histograms appear bimodal. The right peak locates around 150 W/m^2. The left peak is greatly merged into the right peak only if the open space is relatively large which causes clear bimodal. The left peak is slightly taller than right.

The 'fair' performance cluster C appears bimodal in histograms and has two sub-groups. For sub-group C1, the right peak locates at 220 W/m^2 and the left peak locates around 0 W/m^2. The left peak is dramatically taller than the right, due to the existence of a large on-edge open space which is shown in dark blue of the extreme low LRE value. For sub-group C2, if the sites have a considerate size in-site open space, the right peak of histogram would be located just over 200 W/m^2 and the left peak would be located at 20-30 W/m^2. If the sites have very small in-site open space, the location of right and left peak would be just below 200 W/m^2 and around 100 W/m^2, respectively.

The 'poor' performance cluster D has three sub-groups. Sub-group D1 has a single mode in the histogram located at 200 W/m2. For sites with small in-site open space,

their single peak has a thicker left tail. In sub-group D2, three peaks exist in the histogram. The right is located at 200-250 W/m^2, the middle at 100 W/m^2 (due to existence of considerable size in-site open space) and the left at 20 W/m^2 (due to on-edge open space). In sub-group D3 three peaks exist; however, the histogram appears even and flat due to the even mix of multi-types of building cluster (pure low-rise, mid-rise clusters) and in-site open spaces. The right peak is located at 240 $W/m2$, the middle peak at 150 W/m^2 and the left peak at 100 W/m^2.

For the last cluster of 'very poor' performance of LRE (cluster E), the histograms appear bimodal. The right peak is located at 230 W/m^2 and left at 100-150 W/m^2. Furthermore, the right peak is taller than the left.

By matching the cluster by LRE P10-P90 result with building distribution index grading, the cluster groups significantly overlap with groups of sites by grading of SVF. SVF dominantly and negatively impacts longwave radiation from buildings collected at a horizontal grid over ground at 3m. Namely, high SVF of building distribution would lead to less longwave radiation in between neighbourhoods which may cause extra heat stress.

For the further study, LRE statistical measures could be applied as one of indicators of thermal comfort. Meanwhile SVF could also be directly used as a substitute of LRE. Furthermore, SVF could also represent the level of illuminance on building facade and indoor/outdoor sunlight availability.

6.9.2 Quantitative Analysis by Regression of P70 of Longwave Radiation from Environment

Because of the close relationship between longwave radiation from environment and SVF, regression of SVF is worth mentioning as a comparison. The regression of SVF

refs to the regression of SVF in chapter 5. Here the regression of LRE is still presented as a reference.

An area with a high range of LRE located around building facades is of interest to the research. P70 of LRE is adopted for regression because it is an indicator of the high range LRE and appears to be the intersection of multiple curves of sites drawn by statistical measures of LRE.

A regression of LRE P70 at 16:00 is carried out with building distribution representative indices. The model shows great predictive power by R square of 0.864 and adjusted R square of 0.854. The shrinkage of predictive power in generalisation is 1.2%. LRE-P70 is negatively correlated to SVF and TriangleSD positively correlated to average distance to road. SVF shows the most significant influential power by a standardised coefficient of -0.977. Coefficients are listed in table 6.11.

Table 6.11 Details of Regression of LRE-P70 with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	422.320	15.217		27.753	.000
SVF	-3.779	.244	-.977	-15.488	.000
TriangleSD	-.028	.005	-.398	-6.061	.000
AvgDistance	.516	.121	.286	4.248	.000

6.10   Integration of Thermal Comfort Regressions and Application in Design

Comparison of Regression Equations

In this chapter, regressions by indicators of three thermal comfort aspects are operated: P25 of mean radiant temperature; P20, P70 and maximum of wind speed; and P70 of longwave radiation. Standardised coefficients of above-mentioned five regressions are listed in table 6.12.

As seen from the comparison of standardised coefficients, the MRT-P25 result is more independent from regression of wind speed and longwave radiation aspects. MRT-P25 is correlated to distribution indices of plot ratio and low interval depth at the boundary. An average low interval depth has only appeared in regression of MRT-P25. WS-Max and LRE-P70 share one index (TSD) from the distribution factor of the amount of high-rises with the regression of MRT-P25. Their requirements on the high number of high-rises are concordant for optimised performances: higher plot ratio would help with low MRT-P25 and higher TSD or clustered distribution would help with high WS-Max and low LRE-P70.

Three wind speed regressions share indices regarding built amount and boundary enclosure. The requirements on the residential building area (representative index from group of built amount) for optimised wind speed performance are identical as seen from the WS-P20 and WS-P70 equations. Low RBA or small site scale would lead to higher WS-P20 and WS-P70, but WS-max shows conflict by requiring high RBA or large site scale to achieve high max wind speed. WS-P20 and P70 also share an index of average low facade ratio and expect it to be low or open at the boundary for a higher wind speed.

Except for the shared indices between wind speed regression equations, WS-P20, P70 and max are separately correlated to an average high interval depth, average distance to road and TSD. This is a reflection of their respective data location on the site map, and the influence of corresponding distribution indices at that position. The indices and location pairs are aIDH and low wind speed behind front row buildings and aD and high wind speed in on-edge open space. TSD and max wind speed only occurred in open spaces.

Except for the aforementioned TSD shared by WS-max and LRE-P70, wind speed regressions of WS-P70 and LRE-P70 regression share a variable of aD. They both

predicted high in-edge open space to achieve higher level performance. LRE-P70 is an index which highly relies on site spatial openness to achieve high level performance.

Table 6.12 Standardised Coefficient Comparisons between
MRT-P25, WS-P20, WS-P70, WS-max and LRE-P70 Regressions

Index Group		High-rise		Built Amount	Interval depth		Set back	Boundary Enclosure	Openness
Models	Expect	TSD	PR	RBA	aIDL	aIDH	aD	aLFR	SVF
MRT-P25	Small		-.676		.252				
WS-P20	Large			-.563		.297		-.462	
WS-P70	Large			-.444			.322	-.405	
WS-max	Large	.485		.313					
LRE-P70	Small	-.398					.286		-.977

In summary, the comparison of five regressions in thermal comfort aspects, except for WS-P20 and WS-P70 preferring small site scale and WS-max preferring larger site scale, the remaining requirements on distribution by multiple regressions for improved thermal comfort performances are concordant: clustered distribution, high plot ratio, less enclosed boundary and more spatial openness.

Application of regression equations in design practice

For the purpose of the easy application of the regression result into design procedures, an equivalent compromise table of building distribution indices is converted.

When thermal comfort performance improvement occurs in every step size of the human perceptive scale, the equivalent changes of building distribution indices which need to be correspondingly compromised are calculated and listed as design reference (table 6.13).

Table 6.13 Equivalent Change of Building Distribution to One Unit Improvement of Thermal Comfort Performance Variables

Independent Variables			MRT-P25 -1°C	WS-P20 +1m/s	WS-P70 +1m/s	WS-max +1 m/s	LRE-P70 -10 W/m²
Cluster	Name	Range					
High-rise	TSD	60-1200				2.062	25.126
	PR	0.6-3.6	1.479				
Built amount	RBA	2-63		-1.776	-2.252	3.195	
Interval depth	aIDL	10-36	-3.968				
	aIDH	11-56		3.367			
Setback	aD	13-61			3.106		-34.965
Enclosure	aLFR	0-93		-2.165	-2.469		
Openness	SVF	50-77					10.235

The equivalent compromises in building distribution indices for one unit performance changes are checked with the range of the corresponding indices. All the compromises are practical in design, and could be applied as design references according to the required amount of performance improvement.

6.11   Summary

In this chapter, the following research questions are answered:

1) What are the characteristics of selected variables (MRT, WS and LRE) describing thermal comfort and their proper statistical measures to adopt in further analysis?
2) How do building distribution indices impact the performance variables in a qualitative way?
3) What extent do building distribution indices impact thermal performance variables to in a quantitative way?

In this chapter, four achievements are made through parametric studies of the thermal comfort performance indices, clustering analysis on performance statistical measurements, clustering analysis on simulation data of the selected statistical measurements, and multiple linear regressions.

Firstly, the characteristics of data distribution of the three selected performance metric, mean radiant temperature (MRT), wind speed (WS) and longwave radiation from environment (LRE) are achieved, matching the histogram pattern with the corresponding performance.

Secondly, statistical measurements are selected for each performance indices as a representative indices in quantitative analysis and optimisation.

Thirdly, the design rules are extracted based on qualitative analysis. Design rules are extracted mainly from the clustering analysis of each thermal performance variable and supplemented with a few concluded during analysis of parametric study and regression results. The summaries are as following:

Marco (Dominative Rules)

1. High plot ratio is most dominant in reducing mean radiant temperature (MRT). High sky view factor (SVF) is most dominant in reducing longwave radiation between neighbourhoods (LRE). Both rules relate to extra heat stress.

2. Orientation of buildings manifests a stronger impact on MRT than on accumulated sunlight hours due to the momentary characteristics of MRT.

Meso Scale (Effectiveness of Controlling Indices)

1. Low MRT and high APSH and WPSH are opposite but not in conflict. Balanced combination leading to good sunlight performance and good-fair MRT performance includes: high-rise community in low density, surrounded by large on-edge open space (especially on the south), and proper amount of east-west orientated buildings of considerable height.

2. For increasing overall wind speed: Decreasing enclosure at boundary is more effective than decreasing building height.

3. For increasing overall wind speed: If in low and mid-rise sites, enlarging open space inside site is more effective than enlarging building separation.

4. For increasing overall wind speed: When site boundary is enclosed, low/mid-rise community is better than high-rise community; when reducing degree of enclosure, high-rise community has a higher changing gradient than low-rise community.

5. For increasing wind speed in low wind speed level: When the site is less enclosed, high-rise community is considered better than low-rise community.

Micro Scale Combination Suggestions:

1. Best combination:

a. For a cooler outdoor environment, or a high proportion (30%) of small MRT value, the combination would be a high plot ratio, clustered distribution, open boundary or relative enclosed, and a large residential building area.

b. For a very good overall wind speed performance, smaller, even, clustered distributed, open boundary and no large building mass blocking the dominant wind direction.

2. Extreme combination: Highly enclosed site with a) extreme plot ratio or b) extreme density, evenness and medium plot ratio will lead to very high proportion (40%) of small MRT value, but may cause sunlight availability problems.

3. Combination causing largest MRT reduction: Low/mid-rise community with a) few high-rise buildings or b) with high degree of enclosure at the boundary or c) with long facade and high density could result in largest reduction of MRT from 14:00-16:00, namely the dramatic transformation from sufficient sunlight potential to strong shadow protection during activity slots.

4. Worst combination of overall wind speed: High density, low-rise, long facade array perpendicular to wind direction.

The attributing pairs of building distribution mentioned in the micro scale rule, which lead to improved performance in thermal comfort, are marked in table 6.14 by serial number of the micro rules.

Table 6.14 Building Distribution Index Combinations Suggested for Improved Thermal Performances

Evaluation Of Performances		High Overall Wind Speed			Low Overall MRT	Low Overall LRE
		Low WS level	Middle WS level	High WS level	High proportion of smaller values	Small overall LRE level
Size	Small Site	1b, 12	1b, 12	1b, 12		
Homogeneity	Even Distribution	1b	1b	1b	2b	
	Clustered Distribution	1b	1b	1b	1a, ,	
Plot Ratio	High				Dominant, 1a, 2a,	
	Medium				2b	
Density	Dense Separation				2b, 3c	
Spatial Openness	Large SVF					Dominant
Open Space Existence	Open Space Inside Site	8	8	8	2,	
	Open Space At Dominant Wind Direction (South & East)	9		9		
	Wider Building Separation	8	8	8		
Boundary Enclosure	Open Boundary	1b	1b	1b	1a	
	Less Enclosed	7a,		7b	1a	
	Highly Enclosed				2a, 2b, 3b,	
Facade Length	Long Facade				3c	
Building Orientation	Not Perpendicular to Dominant Wind	1b	1b	1b	2,	
Building Height	Low/Mid-Rise			7b	3a, 3b, 3c	
	High-Rise	7a			3a	
Build Amount	Large Residential Building Area				1a	

The qualitative conclusions of this chapter are the series of regression equations and their equivalent application in practice. A consolidation of regression equations forms table 6.15 as a presentation of quantitative results of this chapter.

Table 6.15 Consolidation of Regression Equations of Thermal Performance Variables:
MRT-P25, WS-P20, WS-P70, WS-max and LRE-P70

Model	Regression Equation
MRT-P25	MRT-P25=0.682×aIDL-20.367×PR+345.211
WS-P20	WS-P20=0.009×aIDH-0.005×aLFR-0.013×RBA+0.901
WS-P70	WS-70=0.009×aD-0.005×aLFR-0.012×RBA+1.683
WS-Max	WS-max=0.001×TSD+0.009×RBA+2.968
LRE-P70	LRE-P70=0.516× aD-3.779×SVF-0.028×TSD+422.320

Chapter 7 Meta-Model Construction for Multi-Domain Multi-Objective Optimisation (MD-MOO)

This chapter expatiates on the preparation work for a global multi-domain, multi-objective optimisation (MD-MOO). It is the preparation of three meta-models for three performance domains. It is the bridge from single domain prediction model to a multi-domain optimisation model. Meta-model, namely the single domain multi-objective function (SD-MOF) here in this research, is the simplified substitution of the complex performance predictive model for each performance index, applied in the iteration of MD-MOO. To make this iteration calculation practicable in workload and time cost, it is necessary to have a simplified meta-model (namely SD-MOF).

The meta-model adopted in this research is the black-box model trained with an enhanced generic regression neural network (GRNN-GWO) based on data pairs in case study simulation and prediction. It has the advantage of fast calculation speed and low initial training data amount which fits the requirements of the scenario of this research.

Three meta-models are needed for three performance domains, acoustic, sunlight availability and thermal comfort. For each meta-model or SD-MOF, its component objective functions, or the single domain single objective functions (SD-SOF), which describe the relationship between single-domain performance index and related arrays of morphology parameters, need to be constructed as a base of training.

To summarise, the content of this chapter is I.introducing the background and methodology of meta-model training in MD-MOO, II.explaining the constructing SD-SOFs in each domain, III.training and assessment of meta-model or SD-MOF for three domain, and IV.discussion about the meta-models. The flowchart of this chapter is as seen in the figure 7.1.

I.Background and Methodology
of Meta-model in MD-MOO

II.Construction of SD-SOFs for Acoustic, Sunlight
Availability, Thermal Comfort Domains

1. Background of Application of Meta-Model in MD-MOO

2. Methodology of Construction Meta-Model for MD-MOO

3. Confirmation of Morphology Parameters and Performance Indices for SD-SOF for Three Domains

4. Sensitive Analysis and Selection of Morphology Parameters and Performance Indices for SD-MOO

5. Mathematical Expression of SD-SOF

IV.Discussion

9. Discussion on Update of Single-Domain Meta-Models

10. Discussion on Weight Selection in Meta-Model Construction Progress

11. Discussion on Sample Size Expansion by Latin Hypercube Method

III.Construction of Three Meta-Models

6. Mathematical Expression of SD-MOF and Meta-model

7. Meta-Model Construction Procedures through GRNN-GWO

8. Meta-Model Prediction and Assessment

Figure 7.1 Content Structure of Chapter 7 Meta-Model Construction for MD-MOO,
the Expansion of Box 7 in Overall Content Structure

The structure of this chapter includes 7.1 Background of Application and Construction of Meta-Model in MDO; 7.2 Methodology of Construction Meta-Model for MD-MOO with GWO-GRNN; 7.3 Confirmation of Morphology Parameters and Performance Indices for SD-SOF for Three Domains; 7.4 Sensitive Analysis and Selection of Morphology Parameters and Performance Indices for SD-MOO; 7.5 Mathematical Expression of Single Domain Component Objective Function (SD-SOF); 7.6 Mathematical Expression of Single-Domain Multi-Objectives Function (SD-MOF) and Meta-Model; 7.7 Meta-Model Construction Procedures through Hybrid Generic Regression Neural Network (GRNN) with Grey Wolf Optimiser (GWO); 7.8 Meta-Model Prediction and Assessment; 7.9 Discussion on

Update of Single-Domain Meta-Models; 7.10 Discussion on Weight Selection in Meta-Model Construction Progress; 7.11 Discussion on Sample Size Expansion by Latin Hypercube Method and 7.12 Summary

Acronyms for Chapter 7

ABC	artificial bee colony optimisation
aCAH	average corner area high value
aCAL	average corner area low value
aCAmean	average corner area
aD	average distance to road
aFHmean	average facade height
aFL	average total facade length
aFLmax	average max facade length
aFLmean	average facade length
aFLmin	average min facade length
aFR	average facade ratio
aFSmean	average facade storey
aIAmean	average interval area
aIDH	average interval length high value
aIDL	average interval depth low value
aILmax	average max interval length
aLFL	average low-rise facade length
aLFR	average low-rise facade ratio
ANN	artificial neural network
aOL	average outline length
APSH	annual possible sunlight hour
APSH-IQR	interquartile range of APSH
APSH-P20	P20 of APSH
APSH-P50	P50 of APSH
APSH-P70	P70 of APSH
APSH-THR(0)	ratio of APSH=0h
APSH-THR(413）	ratio of APSH<413h
AR	aspect ratio
aRS	average residential storey
aTCA	total corner area
BD	building density
BPNN	backward propagation neural network
DL	diagonal length of site
DoCE	design of computational experiment
DV	design variable

FPA	foot print area
GRNN	genetic regression neural networr
GRNN-GWO	genetic regression neural network hybrid with grey wolf optimiser
GRNN-PSO	hybrid generalised regression neural network particle swarm optimisation
GWO	grey wolf optimiser
GWO-GRNN	improved GRNN with grey wolf optimiser
GWO-Kriging	improved Kriging with grey wolf optimiser
HCCI	Homogeneous-Charge Compression Ignition
HRBA	high-rise building area
HRBR	high-rise ratio
LHS	Latin hypercube sampling
LMRBA	low/medium-rise building area
LRE	longwave radiation from environment
LRE-P70	P70 of longwave radiation from environment
mae	mean absolute error
mape	mean absolute percentage error
MARS	multivariate adaptive regression splines
MD-MOF	multi-domain multi-objective function
MD-MOO	multi-domain multi-objective optimisation
MDO	multi-domain optimisation
MLR	multiple linear regression
MOCBO	multi-objective colliding bodies optimiser
MOO	multi-objective optimisation
MOPSO	multi-objective particle swarm optimiser
MOSOS	multi- objective symbiotic organism search
MRT	mean radiant temperature
MRT-P25	P25 of mean radiant temperature
mse	mean square error
NSGAII	non-dominated sorting genetic algorithm with elitist strategy
NSGWO	non-dominated sorting grey wolf optimiser
OGBA	over ground building area
OV	objective variable
pdf	probability density function
PR	plot ratio
RBA	residential building area
RBFNN	radical base function neural network
RFPA	residential foot print area
SD-MOF	single domain multi-objective function
SD-MOO	single domain multi-objective optimisation
SD-SOF	single domain single objective function
SF	site shape factor
SOF	single objective optimisation functions
SPL	sound pressure level
SPL-IQR	interquartile range of SPL

SPL-L10	sound pressure level L10
SPL-P10	P10 of sound pressure level
SPL-P40	P40 of sound pressure level
SPL-P70	P70 of sound pressure level
SPL-THR(65)	ratio of SPL<65dBA
sRC	residential circumference
sRSA	residential superficial area
SCR	Residential building superficial area divided by building circumstance
SVF	sky view factor
SVF	sky view factor
TLA	total land area
TSD	standard deviation of triangle area
UHI	urban heat island
WPSH	winter possible sunlight hour
WPSH-IQR	interquartile range of WPSH
WPSH-P30	P30 of WPSH
WPSH-P50	P50 of WPSH
WPSH-P70	P70 of WPSH
WPSH-THR(0)	ratio of WPSH=0h
WPSH-THR(83)	ratio of WPSH<83h
WS	wind speed
WS-Max	maximum of wind speed
WS-P20	P20 of wind speed
WS-P70	P70 of wind speed

7.1 Background of Application and Construction of Meta-Model in MDO

This section introduces the background of the concept of a meta-model as well as its applicative necessity in MDO. Reviews of a meta-model applied in MDO in architecture and other research fields are followed. Multiple construction approaches of a meta-model are reviewed. According to the comparisons, genetic regression neural network (GRNN) method is adopted. Based on the requirement of parametric input of GRNN for meta-model construction, grey wolf optimiser (GWO) is hybridised to GRNN to optimise parameter decision.

7.1.1   Introduction of Meta-Model Concept and Its Necessity for Multi-Domain Optimisation (MDO)

Meta-model is first presented by Sacks, 1989 (Sacks *et al.*, 1989). Meta-model was first developed with the aim of modelling the deterministic output of a computer experiment as the realization of a stochastic process, providing a statistical basis for designing experiments for efficient prediction to expensive computational experiments. The uncertainty of prediction could also be assessed with this model.

Meta-model is an approximate estimation model to substitute the original predictive models with high accuracy in individual sub-systems during MDO progress. It aims to reduce calculation scale during MOO of the whole system. It is essential because the complexity of each of the prediction models from multiple research fields involved in MDO may fail the global search calculation and iteration due to excessive calculation load.

The nature of meta-model for MDO is a mathematical model to fit discrete data of independent and dependent variables with approximation approaches. It could be stated that the real model equates to the meta-model plus approximate errors.

Meta-model is constructed by a regression and fitting method, based on input variable, and output variable dataset from an original prediction model. Once achieving a meta-model, its prediction ability will be assessed to the original simulation or prediction model.

As in this thesis, the reason for adopting a simplified meta-model is to ensure the performability and implementation of the MD-MOO is through considerable times of iterations and convergences. The iteration starts from predicting performance results based on the morphology scheme, then optimal searching based on multi-domain performance results, followed by iterating the predictive and searching procedures based on a newly generated morphology scheme from generic algorithm, until most optimised performance results are collected, and the corresponding morphology schemes are confirmed feasible in practise.

Without a meta-model, this iteration which needs to repeatedly call-in the analysis model, involving predictions related to 23 performance indices in three performance domains and 20 morphology parameters, would result in enormous calculations and time costs. Hence, it is necessary to have a simplified model (in this thesis this refers to single-domain multi-objective function, namely SD-MOF) summarising the relationship between single-domain performance metric and related arrays of morphology parameters.

7.1.2 Reviews of Meta-Model Applied in MDO in Architecture Field and Construction Methods

Meta-model is also mentioned as a surrogate model in some pieces of research. Multiple linear regression and multi layer regression models can be applied as a meta-model of a practical problem, as found in this example (Chen, Yang and Sun, 2017). More complex machine learning approaches are also available to construct a meta-model, for example an artificial neural network (ANN). The Bayesian neural

network is first developed by adopting an unsupervised learning method based on Bayesian probability theory. Evolutionary algorithms are used to train ANN, which is a neuroevolution of machine learning. The meta-model trained by a neuroevolution method would be faster in prediction than the detailed simulation, however the reliability of the prediction is dependent on the constraint definitions and cross validation sample data (Kheiri, 2018).

The meta-models are often applied in research fields involving multiple disciplines, with a number of evaluative metrics from completely different viewpoints which may share part of the design variables. For example in the car crush experiment, the supportive and protective performance of the driver's seat requires simultaneous consideration with its comfort level; two disciplines that share design variables in the structural design of the seat.

Meta-model has recently started to be used in the architecture design field. It is especially widely used in the thermal and building energy design (Ouarghi and Krarti, 2006; Magnier and Haghighat, 2010; Zemella *et al.*, 2011; Gossard, Lartigue and Thellier, 2013; Sun, Han and Feng, 2015).

Several green building meta-models were developed and tested to provide robust building performance predictions for the green building assessment of passively designed high-rise residential buildings in Hong Kong. This work also accurately spots important architectural design factors and prepares for the efficiency of future optimisation. The meta-model construction is based on the EnergyPlus simulation and Monte Carlo regression approach to interpret the relationship between input parameters, which are passive building design factors of building layout, geometry, facade thermophysics, etc, and output indices, which are indoor environmental indices of daylight, natural ventilation and thermal comfort (Chen, Yang and Sun, 2017).

This work concentrates on the sensitive analysis of design variables for regression model fitting. There are the only two papers operate deep research of sensitive analysis in architectural fields. As the inclusion and omission of design variables for the regression equation are highly dependent on the composition of output variables, input variables and sample size, minor differences can have profound influences over the relative importance of an input. Validations for sensitive analyses with two approaches have been operated, which are the multiple linear regression and rank transformations of model responses, to confirm the reliability of variable selection for current usage in design (Chen, Yang and Sun, 2017). A differently sensitive analysis by MLR is solely used as the basis of design variable selection for future MD-MOO.

Latin Hypercube Sampling (LHS) is also used with the Monte Carlo method to expand samples within the design range for the meta-model of a green building design. Valid sample size is also tested. For their regression models of selected passive design parameters and indoor environmental indices, a size over 100 per regression coefficient is determined for acquiring stable statistical estimations (Chen, Yang and Sun, 2017).

Only the Multiple Linear Regression (MLR) model and the Multivariate Adaptive Regression Splines (MARS) model are applied for generate regression fitness models for linear and non-linear problems, respectively. The paper also suggests to apply more complex machine learning methods, like artificial neural network (ANN), for a more accurate meta-model to substitute the simulation model (Chen, Yang and Sun, 2017).

Most widely accepted training methods for meta-models used for MDO are Kriging, improved version of Kriging, various versions of GRNN, i.e. BP-GRNN, RBA-GRNN, GWO-GRNN, etc.

The GRNN is a probabilistic neural network proposed by Donald (Specht, 1991). As a memory based, one-pass learning network with highly paralleled structure without iteration, it estimates function values of continuous variables and converges them to the regression surface based on kernel regression network (Wang et al., 2016). The algorithm has good ability of smoothly transiting from sparse sample data even in multi-dimensional design space. Hence, GRNN has the advantages of a high level of fault tolerance and robustness, and relatively good prediction even under the condition of a lack of sample data. For instance, as in this thesis of only 43 real practice samples available, GRNN is a suitable approach for meta-model construction.

Applied in the field of mechanical engineering for ethanol fulled Homogeneous-Charge Compression Ignition (HCCI) engine, a hybrid generalised regression neural network (GRNN)–particle swarm optimisation (PSO) model (namely GRNN-PSO model) performs a rapid optimisation within 75ms, and allows customised weight adjustment for optimisation by a developed tool (Bendu, Deepak and Murugan, 2017). The model is used to optimize three input parameters: the charge temperature, engine load, and EGR rate.

The well-known Grey Wolf Optimizer (GWO) is applied for a single objective optimisation of one parameter in the GRNN training procedure. Grey Wolf Optimizer (GWO) was developed by Mirjalili etc. (Mirjalili, Mirjalili and Lewis, 2014).

The operational procedures are:

1. Initialise the wolf pack population.

2. Generate solutions for grey wolves and the prey stochastically.

3. Calculate the fitness for positions of each wolf according to objective function.

4. Update the position of each wolf with respect to the position of alpha, beta and gamma wolf.

5. Calculate absolute distance between the current best solution to the prey position.

6. Terminate when convergence condition is met.

A new non-dominated sorting grey wolf optimiser (NSGWO) algorithm was developed in 2018 (Jangir and Jangir, 2018), and applied in the research field of multi-objective optimisation over engineering design and economic constrained emission despatch problem with integration of wind power. The mechanism and validation of the efficiency and effectiveness of the proposed algorithm NSGWO was expatiated. The results showed that NSGWO is comparable in execution time, general distance, diversity metric from aspects of high coverage and fast convergence over standard unconstraint, constraint and engineering problems, against algorithms of multi-objective colliding bodies optimiser (MOCBO), multi-objective particle swarm optimiser (MOPSO), non-dominated sorting genetic algorithm II (NSGAII) and multi-objective symbiotic organism search (MOSOS).

7.1.3   Meta-Model Construction Method Comparison and Adoption of GWO-GRNN

To construct a meta-model of single domain multi-objective optimisation (SD-MOO), various approaches are available. Kriging, improved version of Kriging, various versions of GRNN, i.e. BP-GRNN, RBF-GRNN, GWO-GRNN, etc. are widely used. To choose the most appropriate approach for this research, a pioneer test on meta-model construction based on the initial data was conducted.

Six model training approaches are tested and compared for their predictive ability to this research question. These approaches are: Kriging, backward propagation neural network (BPNN), radical base function neural network (RBFNN), generic regression neural network (GRNN), improved Kriging with grey wolf optimiser (GWO-Kriging) and improved GRNN with grey wolf optimiser (GWO-GRNN).

The selection of meta-model training approach is based the assessment of attempted meta-models' predictive accuracy. The pioneer meta-models are trained and examined adopting 17 design variables and 23 objective variables from the previous studied

building distribution parameters and environmental performance indices in Matlab2017. The constitution of the DVs and OVs are listed as below.

The 23 objective variables for acoustic are SPL-IQR, SPL-P40, SPL-P70, SPL-THR(65); for sunlight are APSH-IQR, APSH-P50, APSH-P80, APSH-THR(413), APSH-THR(0), WPSH-IQR, WPSH-P50, WPSH-P80, WPSH-THR(83), WPSH-THR(0); for thermal are LRE-P80, LRE-IQR, MRT-P80, MRT-IQR, MRT-THR(28), WS-P50, WS-P80, WS-IQR, WS-THR(0.3). The 17 design variables are RBA, HRBA, RBFA, TLA, SCR, PR, TSD, SVF, aRS, aD, aTCA, aIAmean, aOL, aILmax, aLFR, aFLmax, aFR.

It worth to mention that the pioneer training was tested simultaneously with the procedure of parametric study, hence the adopted building distribution indices and performance parameters have slight difference to the final meta-model training. Since the pioneer meta-model is the same research question under similar numbers of indices and parameters, the assessment and selection of the training approach is still applicable and generalisable to the final meta-model training for three performance aspects.

All 23 OVs are assessed according to the three predictive errors. Limited by the length of the thesis, here only the acoustic objective variables SPL-P40(N2), SPL-P70(N3), and SPL-THR(65)(N4), SPL-IQR(N5) are presented and analysed to indicate the comparison between the predicted values and original sample objective value. The full procedure is identical as in section 7.5.

The variations of performance variables in acoustic field is the most well explained by building distribution parameters, as seen in the multiple regression equations, compared to the sunlight and thermal fields. The R^2 for acoustic performance parameters are generally higher than the other two aspects, especially thermal. Because the simulation models for three aspects, including detailed setup condition,

involved indices and exclude indices, are simplified in different degrees, the involved design variables could not fully explain the variation of the performance variables, especially in thermal aspects. With consideration of this fact, since a training approach is needed for accurate predictive ability for all three aspects, the fitness of acoustic meta-model is more suitable as an assessment criterion of training approaches.

The meta-models used 43 samples. A dataset of 35 cases is used as a training group, and of the rest, 8 cases are used as an assessing group.

The assessing criterion measurements of the meta-models are mean absolute error (mae), mean square error (mse), and mean absolute percentage error (mape), mathematically presented as below:

$$mae = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \qquad (7.1)$$

$$mse. = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \qquad (7.2)$$

$$mape = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \qquad (7.3)$$

Where m is number of dataset in assessing group, y is the real data, $\hat{y}$ is predicted data.

The predictive efficiency of tested acoustic meta-models is shown in table 7.1. It is obvious the predictions of BPNN and RBFNN are below satisfactory. Hence, a graphical comparison only implies among Kriging, GWO-Kriging, GRNN and GWO-GRNN meta-models (Figure 7.2).

As seen from the curves in figure 7.2 compared acoustic meta-model predicted value against the original sample value, Kriging and GRNN have advantages on different functions of the 4 SOFs. The Kriging applied with GWO also has no significant improvement. All 4 training approaches have certain fluctuations over the original values. GWO-GRNN has relatively good stability among all construction methods

tested with less far-outlying prediction values. The prediction value of the meta-model constructed by GWO-GRNN is generally close to the real value of the four objective variables evaluated with the 8 samples of assessment group.
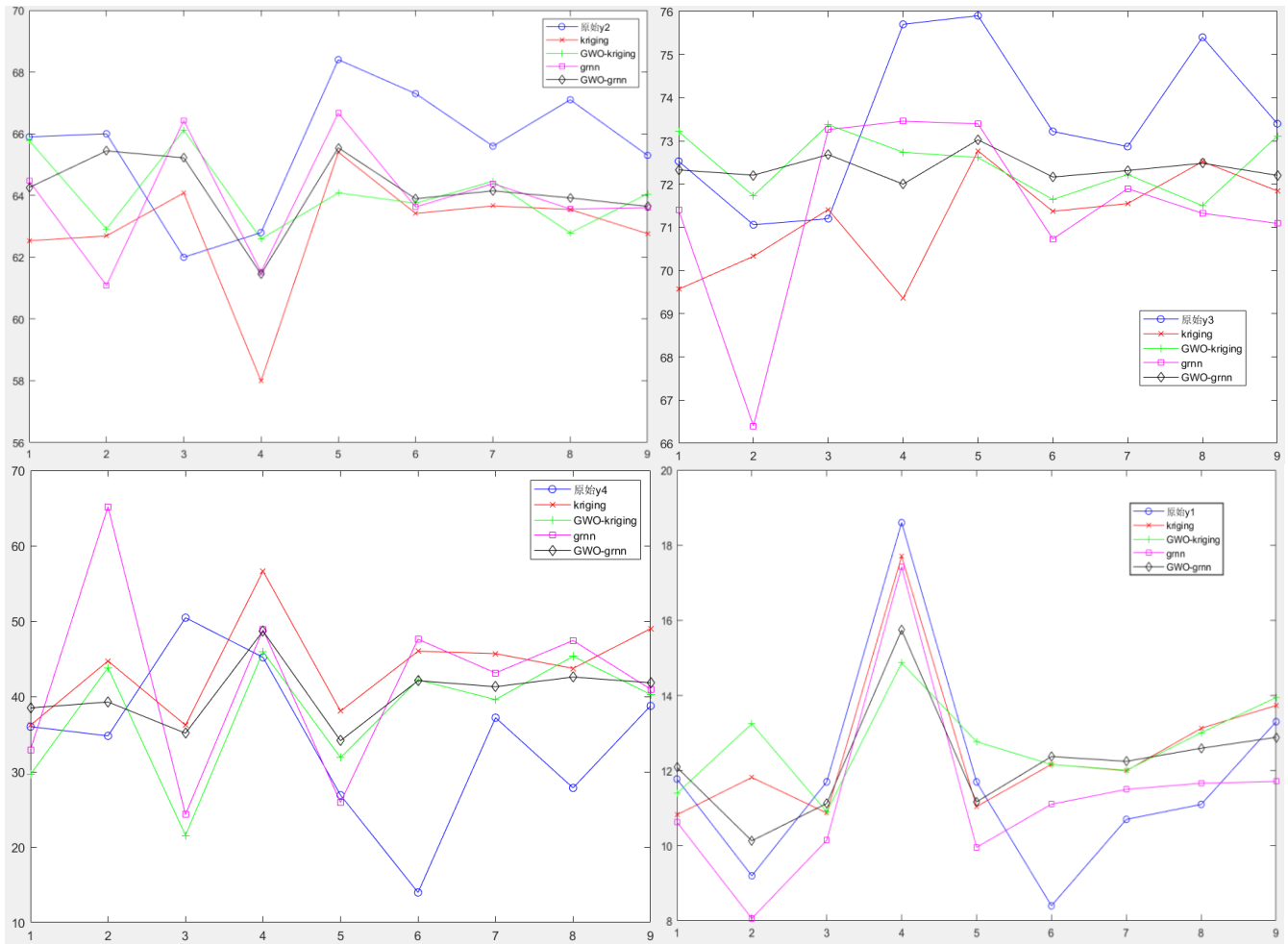
Figure 7.2 Comparisons of Acoustic Meta-Model Predicted Value of SPL-P40 (Up Left), SPL-P70 (Up Right), SPL-THR(65) (Down Left) and SPL-IQR (Down Right) by Kriging, GWO-Kriging, GRNN and GWO-GRNN over Original Acoustic Simulation Model

Small values for three errors are expected as signs of good fitness in meta-model. According to the three error comparison, GWO-GRNN outperforms the BPNN, RBFNN, Kriging, GWO-Kriging and GRNN approach in training the meta-model from the same data set.

Table 7.1 the Assessing Measure of Sub-objectives of Acoustic Meta-Model Constructed by
BPNN, RBFNN, Kriging, GWO-Kriging, GRNN and GWO-GRNN

BPNN	MSE	MAE	MAPE
SPL-P40	30.7532	5.0223	0.0982
SPL-P70	30.7194	5.0316	0.0409
SPL-THR(65)	3218.0963	54.8371	1.9936
SPL-IQR	9.5956	2.1303	0.2033
RBFNN	**MSE**	**MAE**	**MAPE**
SPL-P40	44.6288	5.8224	0.0879
SPL-P70	35.6026	5.2328	0.0707
SPL-THR(65)	846.8544	24.9624	0.9628
SPL-IQR	7.0957	2.1260	0.1948
Kriging	**MSE**	**MAE**	**MAPE**
SPL-P40	10.7272	3.1627	0.482
SPL-P70	8.3638	2.3316	0.0313
SPL-THR(65)	223.5790	12.6257	0.5106
SPL-IQR	3.3148	1.4973	0.1472
GWO-Kriging	**MSE**	**MAE**	**MAPE**
SPL-P40	8.0891	2.4255	0.0367
SPL-P70	4.5262	1.6965	0.0228
SPL-THR(65)	203.4581	10.1054	0.4141
SPL-IQR	6.1907	2.0413	0.1851
GRNN	**MSE**	**MAE**	**MAPE**
SPL-P40	7.7870	2.5756	0.0393
SPL-P70	6.4756	2.3369	0.0317
SPL-THR(65)	297.90	12.902	0.5167
SPL-IQR	2.5012	1.3962	0.1290
GWO-GRNN	**MSE**	**MAE**	**MAPE**
SPL-P40	5.5417	2.1455	0.0326
SPL-P70	4.0948	1.6805	0.0226
SPL-THR(65)	150.97	9.2261	0.3973
SPL-IQR	3.3703	1.4040	0.1287

For objective SPL-P40 SPL-P70 and SPL-THR(65), meta-models trained by GWO-GRNN acquires smallest MSE, MAE and MAPE value compared to the other 5 approaches with great reduction on the errors. Take MSE reduction as an example, for SPL-P40 meta-model, to the most GWO-GRNN reduces the prediction error through RBFNN approach by 87.58%; to the least GWO-GRNN reduces the error by GRNN

approach by 28.83%. For SPL-P70 meta-model, to the most GWO-GRNN causes an 88.50% reduction on MSE compared to RBFNN; to the least GWO-GRNN causes a 9.53% reduction compared to GWO-Kriging approach. For SPL-THR(65) meta-model, to the most GWO-GRNN provides 95.31% reduction on MSE compare meta-model from BPNN approach; while to the least GWO-GRNN provides 25.81% reduction on MSE compared to GWO-Kriging approach. For SPL-IQR meta-model, GWO-GRNN approach has similar predictive efficiency as Kring approach. While GRNN approach slightly overperform GWO-GRNN, as seen in table 7.1 the three errors are slightly smaller than which by GWO-GRNN.

To sum up, GWO-GRNN significantly stands out in predictive efficiency in meta-model training compared to the rest 5 training approaches in 3 out of 4 acoustic performance objectives. And GWO-GRNN equally performs in the SPL-IQR meta-model to the other approaches. Therefore, it is proper to select GWO-GRNN training method as the tool to construct the single domain and multi-domain meta-models as the base of the multi-domain mulit-objective optimisation in this research.

7.2 Methodology of Construction Meta-Model for MD-MOO with GWO-GRNN

In this section, firstly the definition and mechanism of grey wolf optimiser (GWO) is explained, followed by the definition and mechanism of genetic regression neural network with grey wolf optimiser (GWO-GRNN).

7.2.1   Definition and Mechanism of Grey Wolf Optimiser (GWO)

Grey wolf optimiser (GWO) is a swarm intelligent algorithm, first presented by Mirjalili (Mirjalili, Mirjalili and Lewis, 2014). GWO imitates the characteristics of

hunting acts of grey wolves (tracking, encircling, chasing and attacking) to accomplish optimisation searching. Similar to particle swarm optimisation (PSO), artificial bee colony optimisation (ABC), GWO is a statistical optimisation that it is easy to operate with less parameters and high robustness. GWO has better performance in convergence speed and calculation accuracy; hence it has gained extended attention in recent years in multiple research fields as an effective optimisation tool of non-linear continuous questions.

In the optimum searching process of GWO, the hunting action of a wolf pack is usually lead by wolf α, with β and δ's occasional attendance. In the wolf pack, α, β and δ are most close to the preys, whose locations are used by other wolves to orientate to the prey. In the searching space, the locations of the optimal solutions (the preys) are unknown.

To mathematically express the hunting, I assume α (best candidate solution), β and δ has a better understanding of the prey's location. Hence, the first three optimums saved so far force the rest of the wolves (including ω) to update their locations according to the optimums.

A mathematical description of an individual wolf tracking the prey is noted as below:
Use equations 7.4 to calculate distance between individual wolf and α, β and δ respectively.
Use equation 7.5 to update locations of the wolves.
Use equation 7.6 to calculate the moving direction of individual wolf to the prey.

$$D_\alpha = |C \cdot X_\alpha - X|, D_\beta = |C \cdot X_\beta - X|, D_\delta = |C \cdot X_\delta - X| \qquad （7.4） \quad X_1 =$$

$$X_\alpha - A \cdot (D_\alpha), \ \ X_2 = X_\beta - A \cdot (D_\beta), \ \ X_3 = X_\delta - A \cdot (D_\delta) \qquad （7.5）$$

$$X_{(t+1)} = (X_1 + X_2 + X_3)/3 \qquad （7.6）$$

Where,

D is the distance of individual wolf to the potential prey presented by position of the α, β, δ wolves; X_i ($i = α, β, δ$) and X are the position vector of the α, β, δ wolves and a grey wolf respectively,

$C = 2 \cdot r_1$, r_1 is the random value in [0, 1].

X_n (n=1, 2, 3) refers to the updated location respectively to the leader wolves; $A = 2a \cdot r_2 - a$, where a decreases linearly from 2 to 0 over the iteration, r_2 is the random value in [0, 1].

$X_{(t+1)}$ refers to the location update of the prey in next iteration, where t is the current iteration.

Parameter A and C urge GWO to explore and exploit the searching space. Along the decreasing of A, half of the iteration is used for exploring and the rest for exploiting the design space. C is the random value of [0, 2] which provides random weight for increasing or decreasing distance between wolf and prey. These ensure the randomness for GWO in exploring and escaping from local optimum solutions.

7.2.2 Definition and Mechanism of General Regression Neural Network With Improvement By Grey Wolf Optimiser (GWO-GRNN)

Genetic regression neural network (GRNN) is a type of radial basis network applied for function approximation. It is a memory-based, one-pass learning algorithm of highly parallel structure. This algorithm is especially suitable for regression problems without justification of linearity in assumption.

Its basic idea is to provide estimation of continuous variables and fit them to the connotative linear or nonlinear regression surface. The required input metrics in GRNN are the system input vector x, system output vector y and an unknown parameter of the possible regression relationship. This approach allows the appropriate function form being unknown and non-specified, by expressing it as a

probability density function (pdf). When the x and y of the joint pdf is accessible, the conditional pdf and the expected y would become predictable. The predictive regression result is expressed in a parallel neural network structure. After acquiring the network, the new, estimated value of y could be calculated with new input of x. The computational time cost would depend on the propagation efficiency in the four layers of the network.

Advantages of GRNN concentrate on the following aspects:

1. Capable of achieving smooth transition among points even only providing sparse multi-dimensional data.

2. Only having one adjustable parameter to control the overall generalisation.

3. Having simple structure of one hidden layer and the same number of neurons as the training data pairs.

4. Would globally converge instead of trapping at local optimums.

5. One-pass training process regardless of sample size without iterations.

6. The prediction is limited within an observed range due to the nature that the projective value of GRNN is the weighted average of training samples.

In this research, the construction of GRNN is operated in Matlab through the function of *newgrnn*, with calling format of:

$$NET = NEWGRNN(P, T, SPREAD)$$

The construction of GRNN by *newgrnn* function requires three inputs: the matrix of input vector, the matrix of target vector and spread of radial basis function. In the structure, Q groups of input vectors consist of R×Q dimensional matrix, Q groups of objective classification vectors consist of S×Q dimensional matrix.

Newgrnn function creates a network of two layers. Only the first layer has biases. The first layer contains *radbas* neurons in which weighted input and net input are calculated by *dist* and *netprod* functions, respectively. While the second layer contains *purelin* neurons in which the weighted input and net inputs are calculated by

normprod and *netsum* functions, respectively. (For more explanation of the Matlab program, see appendix B)

The spread parameter is the distance between an input vector and a neuron's weight factor, and it is user defined. It is used to decide the value of a neuron's weight vector under which the neuron's weighted input would be spread accordingly. The spread limits the expanding speed of radial base function, default set as 1.

The mechanism in the first layer of the GRNN by *newgrnn* refers to calculating weighted average between target vectors of which design input vectors are closest to the new input vector. Therefore, the network would respond significantly to the target vector paired with the nearest input vector. When spread is small, the radial basis network function would be steep, and the neuron of weight vector closest to the input would acquire larger output than other neurons. As a result, the regression function would closely, but less smoothly, fit to the sample. When spread is large, namely when more neurons contribute to the average, the fitting network function will be smoother but with increased errors. Therefore, if the close fitting of the data to the approximate function is required, adopt a smaller spread than the typical distance between input vectors, and vice versa.

A good balance between close and smooth regression fittings requires the necessity of involving the grey wolf optimiser (GWO) to help specify the spread value. Conventionally, looping of the training procedures is used to search for the best spread value, and cross validation method is used for predictive accuracy examination of models of GRNN. But this method is slow in calculation speed and become tedious when sample size increases. Because the spread value is decided through tests, and varies depending on input data pairs, an optimisation of self-updating in the spread value according to the input data characteristics for every construction of GRNN model is operated. The optimisation operator adopts grey wolf optimiser (GWO).

GWO is hybridised into GRNN for global optimisation of the spread values according to the dataset characteristics of the sample group for each time meta-model training. Hence, predictive error of the tested meta-model is used as the objective function in GWO. After iterations of optimisation with calling-in meta-model training procedure, the best spread is saved for the final training based on certain samples.

Since the meta-model is a substitution of the original simulation model in MDO system, its accuracy is vital for the validity of the whole optimisation. The main two influential factors on accuracy are the construction method of the fitting equation and the selection of sample points. Multiple construction methods are compared and GWO-GRNN is selected due to the best approaching ability, as shown in section 7.1.3. As for sample selection, in this research only 43 cases are adopted in training and optimisation. Even trained with these sparse samples, the predictive accuracy of the system is acceptable and optimisation results could be cross validated with cases of good environmental performance. However, further sampling design with consideration of existing data is highly suggested to update the meta-model for higher predictive accuracy.

The predictive accuracy of the meta-model is examined by the calculation of error between predicted value and observation value. The generalisation of the meta-model is examined by cross-validation. Cross-validation is an assessment method to examine the ability of generalisation of a research model to an independent sample other than the sample data utilised for its estimation analysis. In cross-validation of the meta-models, the samples adopted are divided into two groups, one group used for meta-model construction, one group for accuracy validation.

7.3 Confirmation of Morphology Parameters and Performance Indices for SD-SOF for Three Domains

In section 7.3, 7.4 and 7.5, how to construct the SD-SOFs for three research domains are explained. The single domain single objective function (SOF) is the basis of forming single domain multi-objective function (SD-MOF) and further, the multi-domain multi-objective function (MD-MOF), or the global objective.

In this research, for SOF and functions constructed by SOF, the independent variable (x) is expressed with environmental performance parameters or building morphology parameters from the viewpoint of the physical meaning of the term; or is expressed with design variable (DV) from the viewpoint of optimisation calculation. Similarly, the dependent variable (y) is expressed with environmental performance indices or statistical derived measurements of performance metrics from viewpoint of physical meaning; or is expressed with objective variable (OV) from the viewpoint of optimisation calculation.

The performance indices and design variables (or performance parameters) are firstly selected for each domain. Then, in section 7.4 according to the sensitive analysis operated, the key performance parameters are spotted. In section 7.5, the component objective functions (or SOF) of each domain's optimisation question (namely SD-MOF) are summarised and mathematically expressed. SOFs of three domains are prepared in the form of functions to be used in the formation of SD-MOF in section 7.6.

7.3.1 Specification of Optimisation Objective Variables of Each Domain

For this integrated environmental performance optimisation, there are components of three sub optimisation systems: in other words, three single domain multi-objective

optimisations (SD-MOO). They are performance optimisations of traffic noise attenuation, sunlight possibility, and thermal comfort in relation to building distribution and arrangement in residential wards in South-East China.

For each SD-MOO, there exists its own objective variable (OV) and design variable (DV). The OVs are identical to the statistical measurements of performance evaluation metrics mentioned in the previous analysis in chapters 3, 4, 5 and 6. Evaluation metrics are discussed and decided in previous chapters for acoustic, sunlight and thermal domain, respectively, with references to design regulations, codes and guidance. The adopted metrics are sound pressure level (SPL) for acoustic, annual and winter accumulative sunlight hour (APSH and WPSH) for sunlight, and mean radiant temperature (MRT), wind speed (WS) and longwave radiation from environment (LRE) for thermal. A series of statistical measures of the selected metrics are generated to display the holistic data distribution of the above metrics at different levels. The introduction of the statistical measure series is to locate the evaluation on critical value fulfilment and performance homogeneity of one metric. The statistical derived measurements are the value of critical percentiles, satisfactory ratio of critical value and interquartile range of metric (for data evenness). These statistically derived measurements or performance indices would be applied as optimisation objectives variables (OV) in SD-MOO and MD-MOO.

The objective variables (OV) in one domain are constrained with variation of other OVs in other domains through the shared DVs. Hence, these OVs of MD-MOO are coupling variables constrained with each other that requires MOO to search the trade-off solutions in-between their constraints.

7.3.2   Specification of Design Variables of Each Domain

Design variables (DV) are building morphology parameters, referring to the design possibilities of the residential ward layout from individual and neighbourhood aspects.

For SD-MOO and MD-MOO, DVs are divided into shared variables and local variables. Regarding to MD-MOO the shared variables could be variables shared by three or two domains, which means the DV has significance to the optimisation objectives in multiple domains but not to all, simultaneously. The local variables refer to those that are exclusively influential in one of the three domains, expressed as Xi. The DVs shared by three domains are noted as X0. The DVs shared by any two out of three domains are expressed as XNS, XST and XNT respectively for variables shared by noise attenuation (N) and sunlight availability (S), shared by sunlight availability (S) and thermal comfort (T), and shared by noise attenuation (N) and thermal comfort (T).

7.4 Sensitive Analysis and Selection of Morphology Parameters and Performance Indices for SD-MOO

A large number of building morphology parameters is extracted and a long list of DVs is selected based on the analysis in Chapter 3. Similarly, plenty of choices from simulations exist as OVs for each domain. However, to ensure accuracy and avoid redundancy in optimisation, the number of DVs and OVs needs to be limited. Hence, a sensitive analysis is essential for selection of necessities.

Due to the fact that the number of the DVs and OVs is considerably large for the optimisation progress under the available sample size of 43, a sensitive analysis is conducted to screen the most influential DVs and OVs for optimisation.

7.4.1 Sensitive Analysis Results For Selection of Performance Indices (OVs)

As previously shown in chapter 2, performance evaluative metrics in three domains are decided, and the performance indices which formed by the statistical measurements of these metrics in various level, are demonstrated. Based on the

parametric analysis in chapters 4, 5 and 6, the most influential statistical measures are selected according to their critical positions in the data distribution of the corresponding performance metrics. The rationale of the selected statistical measures of the performance matrices refers to the corresponding sections of Chapter 4, 5 and 6.

The applied OVs for acoustic domains are SPL-P10, P40, and P70 indicating traffic noise level at low, medium and high range. Except for absolute noise level, ratios of satisfactory level (which is 65dBA) and interquartile range are also applied as acoustic OVs, respectively for indication of ratio of standard fulfilling and homogeneity of performance level.

Similarly, for sunlight domain the applied OVs are APSH-P20, P50, P70, and WPSH-P30, P50, P70 as indication of level distribution at low, medium and high level for accumulated sunlight hours in annual and winter span. In addition, APSH-THR(413), APSH-THR(0), WPSH-THR(83) WPSH-THR(0) are four standard-fulfilling ratios, as well as APSH-IQR and WPSH-IQR as representatives of homogeneity. SVF is also applied as objective considering from daylight aspect, except for sunlight aspect.

For thermal domain, WS-P20, P70 and max are used as indicators of low, medium and high level of wind speed, and MRT-P25 and LRE-P70 are used as critical measurements of mean radiant temperature and longwave radiation from environment.

Details of the OVs are listed in table 7.2 with lower and upper bounds and corresponding optimisation target.

Table 7.2 Details of Adopted Optimisation Objectives for Three Domains

Objective Variable	Abbrev	Number	Range		Optimisation Target
			Lower	Upper	
P10 of sound pressure level	SPL-P10	Y19	0	90	min
P40 of sound pressure level	SPL-P40	Y20	0	90	min
P70 of sound pressure level	SPL-P70	Y21	0	90	min
Ratio of SPL<65dBA	SPL-THR(65)	Y22	0	100	max
Interquartile range of SPL	SPL-IQR	Y23	0	70	min
P20 of APSH	APSH-P20	Y1	0	1653	max
P50 of APSH	APSH-P50	Y2	0	1653	max
P70 of APSH	APSH-P70	Y3	0	1653	max
Interquartile range of APSH	APSH-IQR	Y4	0	1653	min
Ratio of APSH<413h	APSH-THR(413）	Y5	0	100	min
Ratio of APSH=0h	APSH-THR(0)	Y6	0	100	min
Sky view factor	SVF	Y7	0	100	max
P30 of WPSH	WPSH-P30	Y8	0	348	max
P50 of WPSH	WPSH-P50	Y9	0	348	max
P70 of WPSH	WPSH-P70	Y10	0	348	max
Interquartile range of WPSH	WPSH-IQR	Y11	0	348	min
Ratio of WPSH<83h	WPSH-THR(83)	Y12	0	100	min
Ratio of WPSH=0h	WPSH-THR(0)	Y13	0	100	min
P25 of mean radiant temperature	MRT-P25	Y14	280	360	min
P20 of wind speed	WS-P20	Y15	0	4.2	max
P70 of wind speed	WS-P70	Y16	0	4.2	max
Maximum of wind speed	WS-Max	Y17	0	4.2	max
P70 of longwave radiation from environment	LRE-P70	Y18	0	380	min

7.4.2 Sensitive Analysis Results For Selection Of Performance Parameters (DVs)

In this research, the sensitive analysis is operated with multiple linear regressions conducted in the corresponding sections of chapters 4, 5 and 6 for acoustic, sunlight and thermal domains. The sensitive coefficient adopts the correlation coefficient, and R^2 of the linear regressions. For each OVs, the included morphology parameters in the corresponding regression equation are the most influential explaining factors of the variation, which are selected as DVs for following SD-MOF and MD-MOF construction.

The MLR results are consolidated in table 7.5. The DVs which are significant to the variation of each OV, are marked in the column of each OV. These DVs appeared in

the table 7.6, if they existed in multiple columns, they appear as shared DVs in MD-MOO; if they existed in only one column, they appear as local DVs in MD-MOO. Although the regressions of objective functions with certain DV are used for DVs selection in MD-MOO, the shared DVs could not be named based solely on it and would not be directly or accurately spotted, because the accuracy of multiple variate linear regression operated is limited by the sample size on the allowed number of adopted DVs. Furthermore, the GRNN based meta-model would be non-specified in its form. Hence, involving all 23 DVs with different influences significance on the OVs is necessary without the current need to clarify the attribute as local or shared.

As confirmed by table 7.6, 20 DVs out of 38 initially selected morphology parameters in previous parametric studies in chapter 3 passed the sensitive analysis and are adopted for further meta-model training and MD-MOO. The selected DVs and corresponding feasible ranges are listed in table 7.3:

Table 7.3 Details of Adopted Design Variables for Three Domains in Optimisation

Morphology Parameter	Abbrev	Number	Adopted No. in Matlab Program	Range	
				Lower	Upper
total land area	TLA	X1	V1	2	42
foot print area	FPA	X2		0.4	9
over ground building area	OGBA	X3		2	65
residential building area	RBA	X4	V2	2	63
residential foot print area	RFPA	X5		0.3	7
average residential storey	aRS	X6		4	36
low/medium-rise building area	LMRBA	X7		0	35
high-rise building area	HRBA	X8		0	63
plot ratio	PR	X9	V3	0.6	3.6
building density	BD	X10	V4	0.07	0.4
aspect ratio	AR	X11	V5	0.5	4
average distance to road	aD	X12	V7	13	61
average outline length	aOL	X13	V8	139	606
average total facade length	aFL	X14		86	383
average max facade length	aFLmax	X15		31	113
average min facade length	aFLmin	X16	V10	12	82

average facade length	aFLmean	X17		18	92
average low-rise facade length	aLFL	X18	V11	0	219
average facade ratio	aFR	X19	V9	0	100
average low-rise facade ratio	aLFR	X20	V12	0	100
average max interval length	aILmax	X21	V13	18	86
average interval depth low value	aIDL	X22	V14	10	36
average interval length high value	aIDH	X23	V15	11	56
average interval area	aIAmean	X24		183	3729
average corner area	aCAmean	X25		0	6519
average corner area low value	aCAL	X26		0	913
average corner area high value	aCAH	X27	V16	0	5837
total corner area	aTCA	X28	V17	0	26073
average facade storey	aFSmean	X29		2	23
average facade height	aFHmean	X30		7	69
residential circumference	sRC	X31		900	19000
residential superficial area	sRSA	X32		17000	480000
high-rise ratio	HRBR	X33	V18	0	1
diagonal length of site	DL	X34		203	897
site shape factor	SF	X35	V19	1.5	4.5
standard deviation of triangle area	TSD	X36	V6	60	1200
sky view factor	SVF	X37		50	77
sound pressure level L10	SPL-L10	X38	V20	0	90

Therefore, 23 OVs and 20 DVs are used in meta-model training and following MD-MOO.

7.4.3   Confirmation Of Preset Constraints of OVs and DVs

For meta-model training, except for design space, OV, DV and their relationship function, the available pre-set constraints between OVs and DVs are also required. They could be expressed in the form of equality or inequality to describe the relationship between OVs, between DVs, and the value range of OVs and DVs.

The hypothetical constraints between OVs are the coupling relationship between them,

which could be seen in the trend of simulation results but remain unknown. Hence, constraints between OVs will be absent in the meta-model training, and will be defined as a conclusion of the MD-MOO.

The pre-set constraints of DVs considered for meta-model training and MD-MOO are the inequality of relative magnitude between the DVs, according to the definition of the variable. Because the 38 initially selected DVs are independent building morphology parameters, in chapter 3, factor analysis has been operated for all available morphology parameters, in order to spot the correlation between them. One or two key parameters are selected from each group of factor analysis to form the 38 initial DVs. Therefore, only the relative magnitude is described according to the definition, for example, by implication, the residential building area should be larger in value than the footprint area; the average outline length (at one side of a site) should be larger than the average lower facade length (at one side of a site) in value, etc. The constraints among adopted DVs are consolidated in table 7.4.

Table 7.4 Inequality Constraints among Adopted Design Variables

Constraints	Implication
$X4 \geq X2$	RBA$\geq$FPA
$X12 \leq X13$	aD$\leq$aOL
$X13 \geq X16$	aOL$\geq$aFLmin
$X13 \geq X18$	aOL$\geq$aLFL
$X13 \geq X21$	aOL$\geq$aILmax
$X13 \geq X22$	aOL$\geq$aIDL
$X13 \geq X23$	aOL$\geq$aIDH
$X16 \leq X18$	aFLmin$\leq$aLFL
$X22 \leq X23$	aIDL$\leq$aIDH

Table 7.5 Design Variable Involvement in Regressions of Optimisation Objectives

Design Variables		Sunlight availability Objective Variables													Thermal Comfort Variables					Acoustic Variables				
Abbrev	number	APSH-P20	APSH-P50	APSH-P70	APSH-IQR	APSH-THR(413）	APSH-THR(0)	SVF	WPSH-P30	WPSH-P50	WPSH-P70	WPSH-IQR	WPSH-THR(83)	WPSH-THR(0)	MRT-P25	WS-P20	WS-P70	WS-Max	LRE-P70	SPL-P10	SPL-P40	SPL-P70	SPL-THR(65)	SPL-IQR
		Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Y21	Y22	Y23
TLA	X1																			Y		Y		
RBA	X4															Y	Y	Y						
PR	X9	Y	Y	Y	Y			Y	Y	Y	Y	Y	Y		Y									
BD	X10											Y		Y						Y	Y		Y	Y
AR	X11						Y	Y																
aD	X12			Y	Y			Y			Y	Y					Y		Y	Y		Y		Y
aOL	X13																				Y		Y	
aFLmin	X16																			Y	Y	Y		Y
aLFL	X18												Y	Y										
aFR	X19						Y																	
aLFR	X20	Y	Y	Y					Y	Y	Y					Y	Y							
aILmax	X21				Y																			
aIDL	X22														Y								Y	
aIDH	X23														Y									
aCAH	X27				Y																			
aTCA	X28																						Y	
HRBR	X33				Y								Y											
SF	X35																			Y	Y		Y	
TSD	X36	Y	Y	Y	Y	Y		Y	Y	Y	Y		Y	Y				Y	Y					
SPL-L10	X38																			Y	Y	Y	Y	

7.5 Mathematical Expression of Single Domain Component Objective Function (SD-SOF)

For acoustic domain, the optimisation objective is to achieve the quietest and even distributed acoustic performance in residential wards. The objective consists of three sub-objectives: quiet noise level, even noise level distribution, and high ratio of area with satisfactory noise level. To represent the three sub-objectives, three groups of OVs are adopted as P10, P40, P70 of sound pressure level, Interquartile range of SPL and ratio of SPL<65dBA. As the order number referred in table 7.3, the single objective functions (SOF) of acoustic domain are mathematically expressed as:

$$N_1 = min\ (Y_{19}), Y_{19} = f\left(X_{Nm}, X_i\right)$$
$$N_2 = min\ (Y_{20}), Y_{20} = f\left(X_{Nm}, X_i\right)$$
$$N_3 = min\ (Y_{21}), Y_{21} = f\left(X_{Nm}, X_i\right)$$
$$N_4 = min\ (100 - Y_{22}), Y_{22} = f\left(X_{Nm}, X_i\right)$$
$$N_5 = min\ (Y_{23}), Y_{23} = f\left(X_{Nm}, X_i\right) \qquad (7.7)$$
$$s.t.$$
$$g(X_{Nm}, X_i) \le 0$$
$$h(X_{Nm}, X_i) = 0$$
$$g(Y_n) \le 0, h(Y_n) = 0$$

Where,
acoustic MOF is consists of 5 SOFs, noted as N_s. N represents Noise.
s is number of SOFs, s=1, 2, …, 5.
X refers 20 design variables (DV), namely X is expressed as a 20 dimension vector.
X consists of X_{Nm} and X_i, respectively refer to shared variables and local variables.
m, i are the number of shared variables and local variables respectively. m+i=20,
Y refers to objective variables (OV). n is the order number of Y, n=19, 20, …, 23.
g is the inequality constraints of X and Y, h is the equality constraints of X and Y.
Constraints of X refer to table 7.4. Range of X and Y refer to table 7.3 and 7.2.

It is worth noting that the OV of Y22 is the solving maximum. For the purpose of compatibility of solving the minimum of SOF in MD-MOO and maintaining the value of SOF larger than zero for a weighted linear summation in construction of MD-MOF, Y22 needs to be transformed when constructing N4. Because the definition of N4 is a

percentage of satisfactory area of noise level, 100-Y22 is used when constructing N4 to ensure the function has positive values, with a reasonable physical meaning of ratio of unsatisfied noise level.

Regarding the sunlight availability domain, the optimisation objective is to achieve long and even-distributed accumulative sunlight hours in winter and year-round in the residential wards, by appropriate organised building distribution. The SOF is divided into seven sub objectives of 1. long accumulative sunlight hours year round (APSH); 2. performance evenness of APSH; 3. large ratio of area with satisfactory APSH; 4. sky view openness for daylight, view comfort and attenuation of urban heat island effect (UHI); 5. long accumulative sunlight hours in winter (WPSH); 6. evenness of WPSH; 7. large ratio of area with satisfactory WPSH.

Hence, seven groups of OVs are adopted in single objective optimisations, which are 1. P20, P50, P70 of APSH; 2. interquartile range of APSH; 3. ratio of APSH＜413h, ratio of APSH=0h; 4. sky view factor (SVF); 5. P30, P50, P70, of WPSH; 6. interquartile range of WPSH; 7. ratio of WPSH＜83h，ratio of WPSH=0h.

Therefore, in accordance with the order number of design variables listed in table 7.3, the single optimisation objective functions (SOF) for sunlight availability domain are expressed as the equation:

$$S_1 = min\ (1653 - Y_1), Y_1 = f(X_{Sm}, X_i)$$

$$S_2 = min\ (1653 - Y_2), Y_2 = f(X_{Sm}, X_i)$$

$$S_3 = min\ (1653 - Y_3), Y_3 = f(X_{Sm}, X_i)$$

$$S_4 = min\ (Y_4), Y_4 = f(X_{Sm}, X_i)$$

$$S_5 = min\ (Y_5), Y_5 = f(X_{Sm}, X_i)$$

$$S_6 = min\ (Y_6), Y_6 = f(X_{Sm}, X_i)$$

$$S_7 = min\ (100 - Y_7), Y_7 = f(X_{Sm}, X_i)$$

$$S_8 = min\ (348 - Y_8), Y_8 = f(X_{Sm}, X_i)$$

$$S_9 = min\ (348 - Y_9), Y_9 = f(X_{Sm}, X_i)$$

$$S_{10} = min\ (348 - Y_{10}), Y_{10} = f(X_{Sm}, X_i)$$

$$S_{11} = min\ (Y_{11}), Y_{11} = f(X_{Sm}, X_i)$$

$$S_{12} = min\ (Y_{12}), Y_{12} = f(X_{Sm}, X_i)$$

$$S_{13} = min\ (Y_{13}), Y_{13} = f(X_{Sm}, X_i)$$

$$s.t.$$

$$g(X_{Sm}, X_i) \leq 0$$

$$h(X_{Sm}, X_i) = 0 \qquad (7.8)$$

$$g(Y_n) \leq 0, h(Y_n) = 0$$

Where:

SD-MOF of sunlight availability domain consists of 13 SOF, noted as S_v, S represents Sunlight.

v is number of SOFs, v=1, 2, …, 13.

X refers 20 design variables, namely X is expressed as a 20 dimension vector.

X consists of X_{Sm} and X_i, respectively refer to shared variables and local variables.

m, i are the number of shared variables and local variables respectively. m+i=20,

Y refers to OVs. n is the order number of Y, n=1, 2, …, 13.

g is the inequality constraints of X and Y, h is the equality constraints of X and Y.

Constraints of X refer to table 7.4. Range of X and Y refer to table 7.3 and 7.2.

It is worth noting that similar to N4 in acoustic domain, Y_1, Y_2, Y_3, Y_7, Y_8, Y_9, Y_{10} are all transformed by subtracting from their corresponding superior limit value for the construction of single objective functions S_1, S_2, S_3, S_7, S_8, S_9, S_{10}.

Similarly, for thermal comfort domain, the optimisation objective is to achieve a good integrated thermal comfort level from three aspects at the test moment in early

summer by appropriate organised building distribution. Hence, the three sub objectives form the SD-MOF, which are low mean radiant temperature (MRT), high wind speed (WS) and low longwave radiation from environment (LRE). The three groups of corresponding objective variables are P25 of mean radiant temperature, P20, P70 and maximum of wind speed, and P70 of longwave radiation from environment. These single objective functions (SOF) are mathematically expressed as:

$$T_1 = min \ (Y_{14}), Y_{14} = f\left(X_{Tm}, X_i\right)$$
$$T_2 = min \ (4.2 - Y_{15}), Y_{15} = f\left(X_{Tm}, X_i\right)$$
$$T_3 = min \ (4.2 - Y_{16}), Y_{16} = f\left(X_{Tm}, X_i\right)$$
$$T_4 = min \ (4.2 - Y_{17}), Y_{17} = f\left(X_{Nm}, X_i\right)$$
$$T_5 = min \ (Y_{18}), Y_{18} = f\left(X_{Tm}, X_i\right) \qquad (7.9)$$
$$s.t.$$
$$g(X_{Tm}, X_i) \leq 0$$
$$h(X_{Tm}, X_i) = 0$$
$$g(Y_n) \leq 0, h(Y_n) = 0$$

Where:

SD-MOF of thermal comfort domain is consists of 5 SOFs, noted as T_w, T represents Thermal.

w is number of sub-objective functions, w=1, 2, …, 5.

X refers 20 design variables, namely X is expressed as a 20 dimension vector.

X consists of X_{Tm} and X_i, respectively refers to shared variables and local variables.

m, i are the number of shared variables and local variables, respectively. m+i=20,

Y refers to evaluation variables. n is the order number of Y, n=14, 15, …, 18.

g is the inequality constraints of X and Y, h is the equality constraints of X and Y.

Constraints of X refer to table 7.4. Range of X and Y refer to table 7.3 and 7.2.

Similar to conditions in the acoustic and sunlight domain, T_2, T_3, T_4 is formed with the corresponding superior limit of objective variables subtracts Y_{15}, Y_{16}, $Y_{17,}$ respectively.

7.6 Mathematical Expression of Single-Domain Multi-objectives Function (SD-MOF) and Meta-model

The nature of the meta-model in this research is an estimated model of single-domain, multi-objective optimisation model. There are three meta-models constructed, respectively, for acoustic, sunlight availability and thermal comfort used in MD-MOO. Hence, the generic form of single domain multi-objective optimisation function (SD-MOF) of three domains is given as equation:

$$Find\ X = \left\{ X_{sk}, X_{lj} \right\}^T$$
$$Min\ F\left(X_{sk}, X_{lj}, D_m \right)$$
$$s.t. \qquad\qquad (7.10)$$
$$G\left(X_{sk}, X_{lj}, D_m \right) \leq 0$$
$$H\left(X_{sk}, X_{lj}, D_m \right) = 0$$

Where:
F represents the single-domain multi-objective function (SD-MOF).
D represents the compounded objective function in each domain, m is the order number of D, m=1, 2 , 3.
X refers 20 design variables, namely X is expressed as a 20 dimension vector.
X consists of X_{sk} and X_{lj}, respectively refer to shared variables and local variables.
k, j are the number of shared variables and local variables respectively. k+j=20,

For this research, D1, D2 and D3 refer to single-domain multi-objective optimisation (SD-MOO) objective variables (OV) for acoustic, sunlight and thermal domain. On the basis of the SD-SOF provided in equation 7.7, 7.8, 7.9, the compound of D_m is a weighted linear summation of SOFs in corresponding domain. The weight for each sub-objective could be adjusted according to design emphasis and preference. Here in this research, the same weight is used for each SOF in one domain, namely SD-MOF equals to average summation of SOFs. The general expression of the single-domain multi-objective optimisation function (SD-MOF) is given by equation:

$$D_m = \alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \cdots + \alpha_n \cdot f_n \qquad (7.11)$$

Where:
D represents the single-domain multi-objective optimisation objective, m is order

number of domains, m=1, 2, 3.

α_n is the weight of each sub-objectives in one domain.

f_n is the function of sub-objective in one domain.

n is the number of sub-objectives in one domain.

For acoustic domain, $f_n = N_m$, n=m=1, 2, …, 5.

For sunlight domain, $f_n = S_m$, n=m=1, 2, …, 13.

For thermal domain, $f_n = T_m$, n=m=1, 2, …, 5.

Therefore, based on the definition, the meta-model of the single-domain multi-objective optimisation model (SD-MOF) is the estimation of the real model, generally expressed by equation:

$$\begin{aligned} F(X) &= \widehat{F}(X) + e \\ &= Min\ F\left(X_{sk},\, X_{lj},\, D_m\right) \end{aligned} \quad (7.12)$$

Where:

F(X) is the real single-domain multi-objective optimisation model.

$\widehat{F}(X)$ is the meta model, in other word estimated model.

e is the error of estimation.

7.7 Meta-Model Construction Procedures through Hybrid Generic Regression Neural Network (GRNN) with Grey Wolf Optimiser (GWO)

This section explains the background of the generic regression neural network and grey wolf optimiser. The construction of the meta-model for MD-MOO adopts the hybrid version of GRNN with GWO. In the end, the meta-model construction procedure through GRNN-GWO is expatiated.

For meta-model construction by GRNN, the parameter adjustment needs empirical decisions and is tested and verified iteration by iteration to assure the prediction accuracy of the meta-model. Due to the fact that the normal grid searching method has shortcomings, including lower accuracy and slow searching speed, GWO is applied to improve the parametrical adjustment and to raise the prediction accuracy of the trained meta-model, in detail, to determine the most influential value of parameter

SPREED in meta-model construction progress. SPREED is a decision parameter of avoiding optimisation trapped in local optimum.

The basic idea of parametrical optimisation by hybrid GRNN with GWO is to take MSE between the meta-model prediction and the sample output value in the test group as the optimisation objective; to optimise the objective through wolf pack searching for the best parameter, which equals the score of alpha wolf at optimal position. At last, the following meta-model training would adopt the best parameter value to achieve best prediction ability. It is noteworthy that the best parameter value is optimised based on this particular group of training and testing samples. When input pair updates, the optimal parameter value would be updated for maximum adaptation to the new data.

The procedures of hybrid GRNN with GWO for meta-model construction are listed as below:

1. data collection and consolidation

As mentioned in previous sections, three meta-models would be separately constructed. Hence, the input data pairs, or input and output data as mentioned in Matlab programs, are prepared according to three performance domains. For acoustic meta-model training, the 20 building distribution parameters and 5 corresponding acoustic performance evaluation indices form the data set of 43 pairs. Similarly, for sunlight meta-model training, 43 pairs of input data of 20 design variables and 13 sunlight objective variables (OVs) are used. For thermal meta-model training, 43 pairs of input data of 20 design variables and 5 thermal OVs are utilised.

In addition, 43 pairs of data are divided into training group of 35 samples and testing group of 8 samples. The samples of the testing group are randomly selected from these 43 pairs. Although a planned sampling within the design space for testing group

data evenness is preferred, as limited by the length of the thesis, Latin Hypercube sampling is not expanded for testing sample conformation.

2. Parametrical optimisation for meta-model training by GWO

The detailed procedures are:

1) Input and output data are normalised to range of -1 to 1 according to their own range. The normalisation and reverse normalisation rules are written into the network for following work of reversing the optimal solutions to real range. The min-max normalisation method is used. The range input and output are projected to -1 to 1.

2) Define the design variables and optimisation objective for GWO

Define the optimisation by GWO as one dimensional, namely one parameter is optimised. The GRNN parameter cmd is defined as design variable (DV) and is represented by the position of the searching agent in GWO. Define the range of parameter need optimisation to be 0.01 to 3. Theoretically the range is 0 to 10, but is empirically limited below 3 as it would fall below 3 for this input pairs.

Calculate the prediction result of the test group input from the trained meta-model and compare it with real output value from the test group data. Then conclude the MSE of the prediction result. Define the MSE as the optimisation objective (OV) of the GWO.

3) Initialisation of optimisation parameters and iteration rules for GWO

Define the population of the searching agent or wolves to 10 and maximum iteration to be 20. Initialise positions and scores of alpha, beta, delta and omega wolf, as well as other parameters utilised in GWO.

Define iteration rules of GWO as:

- When the position of the searching agent exceeds the lower or upper bound of parameter, bring back position to lower or upper bound, respectively.

- When MSE smaller than alpha score, update alpha score to current MSE and corresponding position.

- When MSE larger than alpha score but smaller than beta score, update beta score to current MSE and corresponding position.

- When MSE larger than alpha and beta score but smaller than delta score, update delta score to current MSE and corresponding position.

Following the above rules, imply encircling the prey and updating searching agents' position by GWO algorithm until convergence to acquire best cmd parameter for GRNN training.

3. Meta-model training by parametrical-optimised GRNN

Firstly, normalise input and output data, then adopt normalised data and optimised parameter value for training.

4. Prediction accuracy testing of meta-model

Calculated MAE, MSE and MAPE of normalised prediction error and error of reverse normalised prediction result for examination. Plot the predicted value of each output variable out of the meta-model with test samples input against real test sample output value for examination.

Following the above procedures, three meta-models are trained and saved for NSGAII optimisation.

The flowchart of meta-model construction by hybrid GRNN with GWO is shown in figure 7.3
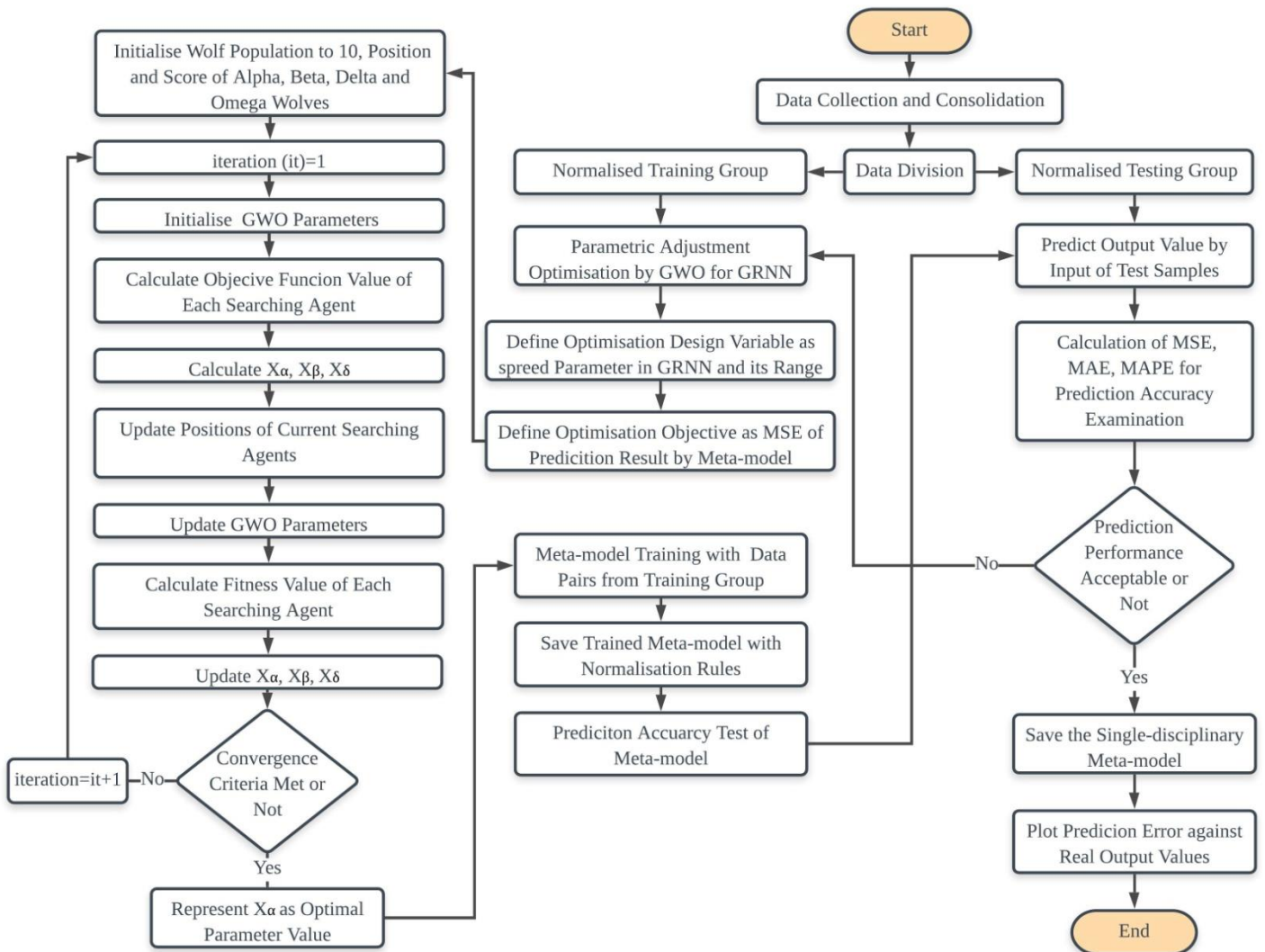
Figure 7.3 Flow Chart of Meta-Model Construction by Hybrid GRNN with GWO

7.8 Meta-Model Prediction and Assessment

Following the procedures in section 7.7, meta-models for acoustic, sunlight and thermal domains are trained. The optimised speed parameters of three models are 1.64142, 1.29922 and 2.67449 respectively, for acoustic, sunlight and thermal meta-models.

A cross validation is operated for the three meta-models. The prediction accuracy of three meta-models is examined with MSE, MAE and MAPE as listed in table 7.6. It could be concluded that the prediction accuracy of the three meta-models are

generally acceptable yet improvable with larger amount and evener distributed sampling in the design space.

Table 7.6 prediction accuracy of three meta-models

Domain	Output Variable	MSE	MAE	MAPE
Acoustic	N1	6.637259	2.182325	0.035639
	N2	4.711982	1.899117	0.028916
	N3	4.919004	1.828059	0.024576
	N4	133.9259	8.947044	0.133195
	N5	4.15153	1.696238	0.150948
Sunlight	S1	9680.504	81.86659	0.074472
	S2	11263.9	97.29724	0.183144
	S3	13815.06	106.8723	0.364811
	S4	14736.13	76.83916	0.092022
	S5	8.702903	2.430394	0.157984
	S6	1.439899	0.97184	Infinitesimal
	S7	35.67356	5.190814	0.166869
	S8	450.3489	16.54532	0.059854
	S9	707.046	24.46897	0.130088
	S10	1689.775	35.21798	0.428748
	S11	1983.315	36.49883	0.147937
	S12	21.93137	3.921292	0.129008
	S13	11.42707	2.518546	0.125862
Thermal	T1	342.1748	18.39599	0.054104
	T2	0.083576	0.193744	0.064314
	T3	0.04489	0.167185	0.068991
	T4	0.020623	0.122625	0.194721
	T5	397.3569	16.00637	0.099136

Prediction output variable values of the 8 test samples are plotted against the corresponding real output values for acoustic, sunlight and thermal in figure 7.4, 7.5 and 7.6. The acoustics meta-model presents relatively good fitness and sunlight meta-model is also acceptable. Comparatively, thermal meta-model shows good fitness for wind speed sub-objectives but less satisfactory for mean radiant temperature (MRT) and longwave radiation from environment (LRE).

The reason for this is because in this research the simulation models of the samples for thermal aspect are highly simplified compared to the real experiment conditions. Except for the calculation mechanism and principle of the simulation software which is fixed, the simulation models consist of detailed setup conditions, involved design variables and excluded influential factors, etc. The decision of how to construct the

simulation models would have impact on the depth of simulation, and further, the prediction of the practical scenarios based on these simulations. In this research, firstly, the setup condition only contains two earth thermal material differences which are asphalt and loamy soil; the building material are all set as default. The only changing design variables are building morphology indices. This leads to the simulation result of MRT and LRE value has small variation range. Lack of objective value variation will increase regression difficulties and reduce the meta-model predictive accuracy. Secondly, the only building distribution indices is not fully covering the MRT and LRE variation in the multiple regression and general regression neural network training, and other undiscovered key explanative indices may not be included as key design variables in thermal aspect, just as the SPL-P10 applied as a design variable representing the sound source factor in acoustic regression and meta-model training. This also leads to the fitness of the trained thermal meta-models not as high as the acoustic and sunlight meta-models, because the building distribution indices could not fully explain the variation of the performance variables.

Therefore, the response of MRT and LRE to the building distribution indices as design variables is limited. It is suggested that in future work exclusive design variables for MRT and LRE should be added and included in meta-model training.
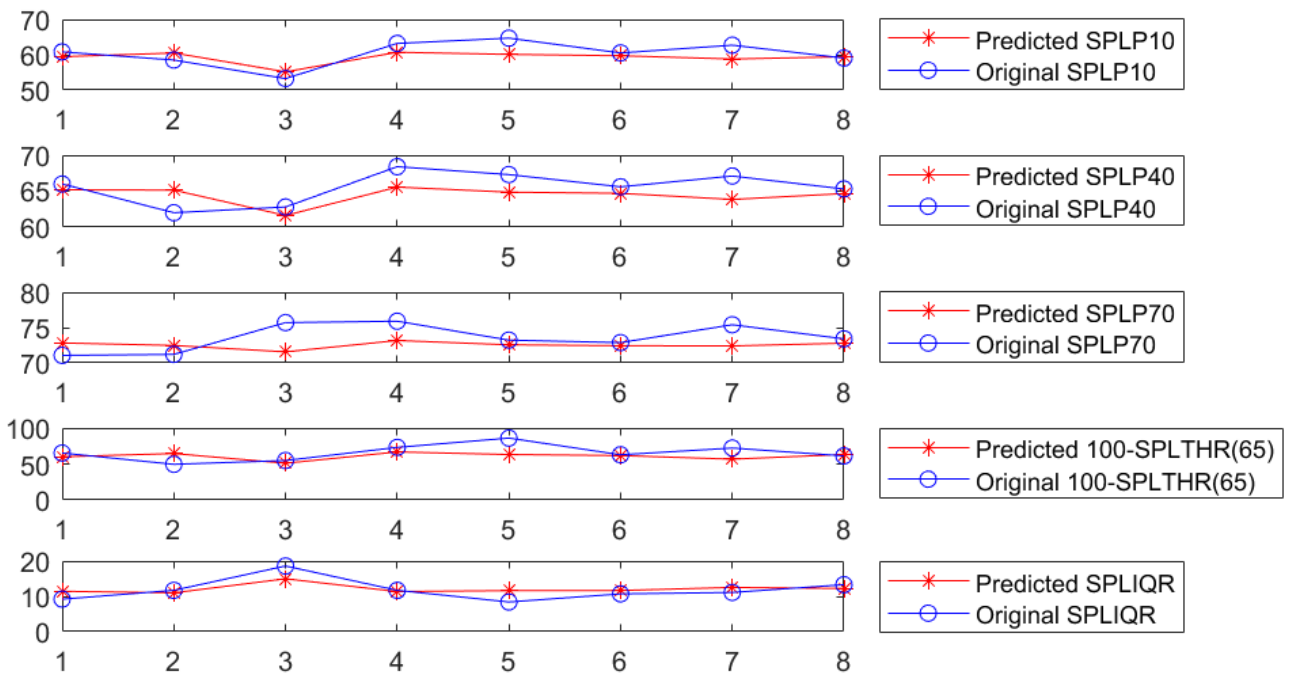
Figure 7.4 Prediction Errors of Output Variables (N1-N5) for Acoustic Meta-Model Tested by 8 Samples
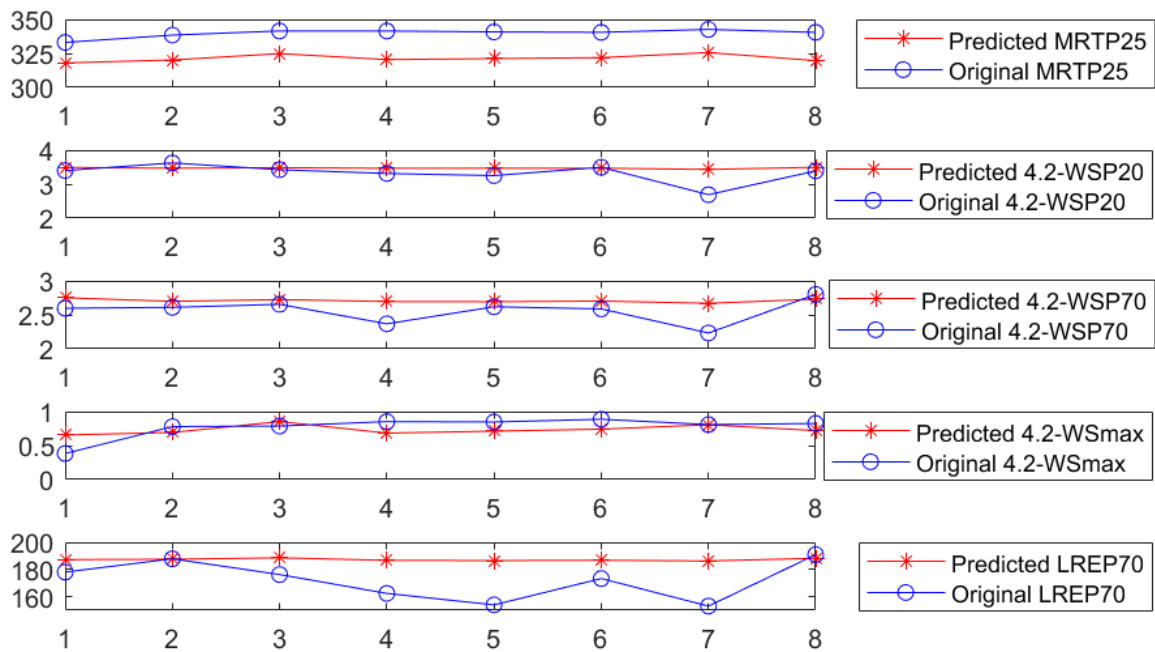
Figure 7.6 Prediction Errors of Output Variables (T1-T5) for Thermal Meta-Model Tested by 8 Samples
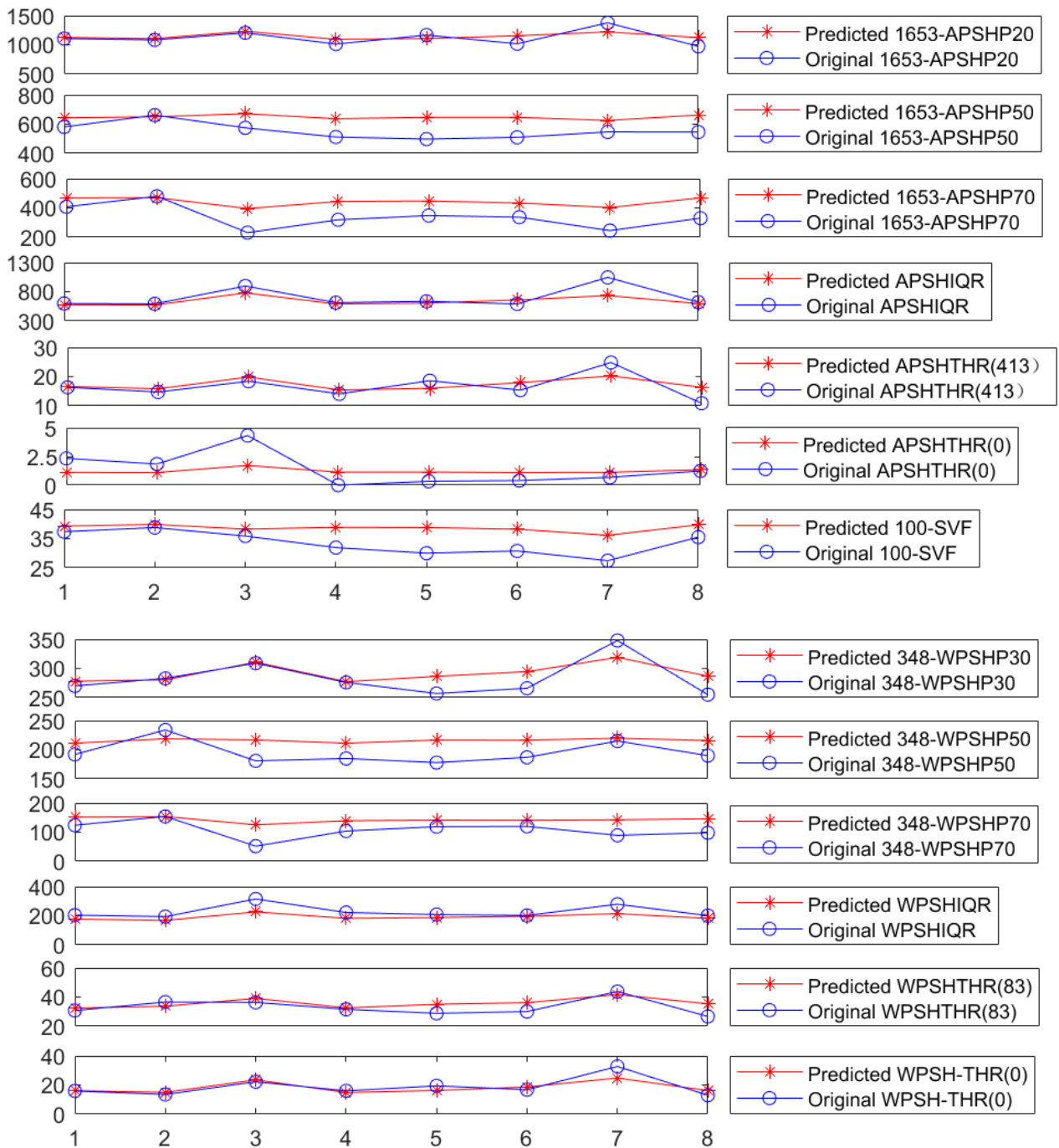
Figure 7.5 Prediction Errors of Output Variables (S1-S13) for Sunlight Meta-Model Tested by 8 Samples

7.9 Discussion on Update of Single-Domain Meta-Models

A few concepts are worth further discussion regarding the validity of meta-model construction in this research, and in the possible direction of future work. They incorporate the update of single domain meta-model and single domain optimisation.

7.9.1　Meta-Model Update Iteration and Requirement

To further improve the prediction accuracy of the meta-model except for iteration of GWO, an iterative loop is required for meta-model updating. The basic idea would be after examining prediction accuracy, if convergent constraints are met, for example below accuracy expectation, a new iteration of sampling of design variables homogeneously from design space guided by sampling method would be required. The newly selected samples would be simulated in the high accuracy model, which is the performance simulation models for various domains, and real output of sub-objective values would be collected. Then, the new sample pairs of input and output would be utilised in addition to previous samples in training and testing the new generation of meta-model. Repeat the iteration of sampling, simulation, training and testing until convergence constraints are met.

Nevertheless, in the context of this research, for the purpose of iteration of meta-model training, newly generated samples with evenly distributed DV values need to be formed into a 3d urban model. Hence, the model could be imported into performance simulation software packages for predicting performance metric output. Therefore a high-dimensional parametric model of residential ward with details responding to 38 design variables or 20 key design variables or building morphology parameters is needed. However, the development of an urban scale high-dimensional parametric model is still lacking in technical support from both algorithm and available parametric modelling software. Also, the development of this parametric

model requires considerable time and effort. Limited by the length of this thesis, its development is not expanded. For future work, if the urban scale parametric model generation tool is available, after even selection of new sample in design space by DoCE, new 3d residential ward models would be generated to be imported into simulation packages, for instance CadnaA, Rhinoceros+Grasshopper and Envi-met, for evaluation. The new generation of result pairs would be utilised for meta-model updating.

7.9.2   Single Domain Multi-Objective Optimisation

There is a possibility to execute collaborative optimisation strategy of three level hierarchical structure, if needed in research. When the update loop is implied for meta-models, SD-MOOs for acoustic, sunlight and thermal domains could also be operated before MD-MOO to form a hierarchical optimisation structure. In the context of this research, the three level structure includes single domain simulation, single domain optimisation and multi-domain optimisation, which is in fact the structure of collaborative optimisation strategy.

Single domain optimisation (SD-MOO) is constituted by iterating the loop of high accuracy simulations, meta-model (low accuracy model) construction and optimisation calculation of meta-model objectives. In the context of this research, SD-MOO allows improvements and updates on the building morphology scheme before MD-MOO, based on performance requirements in only one domain. This provides the opportunity of introducing optimisation emphasis, through adoption of optimal results of SD-MOO in order to narrow down MD-MOO constraints of allowed searching space. In future work, SD-MOO is suggested to be optional before MD-MOO, if sample amount allows.

It is worth noting that de-coupling process is required for MD-MOO design variables, when SD-MOO is implied and isolated from the whole MDO system. As mentioned

in 9.1.1, the shared DVs are coupling variables because they are partly shared in cross domain optimisation; this means the optimal combination of DVs in one domain may leed to worse performance in another domain. De-coupling is the approach of introducing auxiliary variables or constraints to limit the variation of DVs when applying SD-MOO with respect to corresponding changes in other domains. The focus of this research has been set to MD-MOO, namely integrated performance, instead of SD-MOO. Meta-models of single domain are only adopted as input sources for MD-MOO. Therefore, the de-coupling of design variables is not exploited, as single domain optimisation is not expanded in this research.

7.10   Discussion on Weight Selection in Meta-Model Construction Progress

Two sections are discussed here which are the weights for various domains in multi-domain multi-objective optimisation (MD-MOO), and the weights for various sub-objectives or single domain single objective functions (SD-SOF) in a single-domain meta-model.

7.10.1   Weight for SD-MOF in MD-MOF

The weight of each domain for MD-MOO determines the grade of importance of one domain compared to other domains in MD-MOO. It is capable of determining the emphasis balance of one MD-MOO. In research scopes including mechanical and aeronautical engineering, the selection references of the weight of domain are empirical values applied in previous studies, current research emphasis and model adjustment requirement (i.e., accuracy improvement of multi-domain model), etc. Therefore, the weight selection is one aspect of the parametric adjustment in optimisation procedures which displays the experiences of the disciplinary researchers and subjective selections.

However, an optimisation approach to define the weight is also operable, for scenarios that cannot empirically and subjectively decide weight due to lack of research emphasis or previewed knowledge, or that theoretical and data-based assignment of weights for domains is necessary. To achieve overall good performance, the weights of domains in MD-MOO need to be decided on a case-by-case basis, respecting the case conditions. For example, for different inputs of building and neighbourhood morphology and surrounding layout conditions, a well performed environment may refer to absolute different weight combinations among performance domains. The existence of weight combination of domains will show its maximum influences when the optimisation algorithms is solving minimum or maximum value solutions. Therefore, the optimisation on weight selection will appear transformational over the optimisation results of DVs. The optimisation of domain weights could be applied by introducing an optimiser for the weight in the iteration of MD-MOO progress, just as the same way of introducing grey wolf optimiser (GWO) for parametric optimisation in MD-MOO.

In this research, NSGAII is applied as optimisation algorithm with specification of identical weights for three domains. Multiple weight combinations are tested in optimisation, and various groups of Pareto solution sets are recorded. The result shows in different weight combinations; the Pareto solution sets have a few differences but are not significantly different. The reason behind this is that NSGAII searches for non-dominated solutions instead of minimum or maximum value solutions of the objective function. This means all solutions in Pareto are incomparably well performed, with consideration from various optimisation emphases. In other words, the possible weights of domains or possible optimisation emphasis have all been expressed in the Pareto solution set.

Therefore, in this research, the optimisation emphasis in domains would be effortlessly implemented through Pareto solution screening with constraints with domain emphasis. For instance, applying solution selection criteria of limiting the

inferior and superior bound of the sub-objective value. Detailed approaches of solution selection refer to chapter 9.

The theoretical and data-based weight adjustment by optimisation is still promising for future study in MD-MOO. It is also especially supportive, as suggested in this research, that optimisation weight should be adjusted in respect to the surrounding layout conditions of a residential ward and local social norms.

7.10.2   Weight for SD-SOF in SD-MOF

Except for weight for various domains, within each domain, the sub-objectives are linearly summed up with weights to construct a single-domain multi-objective optimisation model or SD-MOF. Currently, because no design emphasis on performance metric existed, the weights of sub-objectives in one domain are equal fractions of a total of 1, with the number of sub-objectives as denominator.

It is tested that in the MD-MOO framework of GWO-GRNN meta-model + NSGA II optimisation, weight adjustment on sub-objectives would show significant influences on the final results. Hence, in future application of this multi-domain optimisation model, a pioneer study would be preferred regarding the performance metric significance in overall performance assessment under specific building morphology and surrounding morphology context. The weight of SD-SOF should be decided according to design preference and emphasis to achieve the model of better adaptability to specific design requirement.

7.11 Discussion on Sample Size Expansion by Latin Hypercube Method

7.11.1 Necessity of Data Expansion and Design of Computer Experiment (DoCE)

As stated, meta-model requires updates to ensure accuracy. However, due to a lack of high-dimensional parametric modelling technique to generate for meso-scale building distribution models to expanded samples for optimisation inputs, it is difficult to operate a meta-model update iteration. Notwithstanding the lack of meta-model update iteration, the predictive accuracy of the meta-model adopted in this research is satisfactory. The meta-model is validated by consistence of optimal solutions and the initial samples of distinctive performance. For future work, a meta-model update is highly suggested. The updating starts with the data expansion of sample size, and collection of simulation results of initial and expanded samples. The nature of data expansion is sampling allowed design variable (DV) range with consideration of existing data distribution. Data expansion is operable through the approaches of design of computer experiment (DoCE).

Design of computer experiment (DoCE) is used to generate sample points which could holistically reflect value characteristics of high accuracy analysis models. In the same scale sampling, higher spatial uniformity would improve approximate accuracy of a meta-model. The reason behind this is that when the value characteristic of an approximate objective function is unknown, samples of high spatial uniformity tend to capture the feature points of the approximate objective in higher possibility. Hence, sampling design is vital as a basis of construction of the meta-model from the viewpoint of data full coverage.

7.11.2 Data Expansion with Latin Hypercube Sampling Method for Meta-model Update

A DoCE method should have features including explorative ability and flexibility. Explorative ability means to sample as low amount of points as possible to holistically acquire characteristics of the real analysis model. Flexibility refers to the capability of generating samples of any specific amount for any dimensional question. Spaces filling sampling methods are widely adopted for their ability for exploration and for their flexibility. Space filling sampling concept would improve explorative ability through aiming to project uniformity and space-filling uniformity instead of consideration of random errors as in conventional DoCE.

Latin Hypercube Sampling (LHS) is one of the most applied space-filling sampling methods. It has no limit from number of variables and levels of variables and is capable of generating samples in any specific amount. Except for its flexibility, the simple implementation in program also makes it widely accepted in DoCE. As a uniform sampling method, basic idea of LHS is to control the location of samples (value of design variables) to avoid clustering of samples in a small range, through limiting only one sample is selected from each specifically divided range. Different from conventional orthogonal and uniform design method, no dependency on reference array exists for organising the experiment in LHS.

To ensure the space-filling uniformity out of the generation randomness of LHS, optimised LHS methods are developed. Among the existing optimised LHS methods, various degrees of evenness, symmetry and regularity in sample point distribution could be achieved, with a balance on sampling time cost and efficiency. As suggested by Teng , for typical high dimension question, as in this research, it would be easier to achieve distribution evenness in space, but one would also need to compromise for sampling quality and efficiency.

In engineering and research practice, quite often the pioneer studies provide an experiment dataset for the later stage of research, but further research requires larger

amount and more holistically representative sampling over the experiment space. However, the pioneer studies could have been invested in experimental and time cost, so that its data should be included in the later stage research. Therefore, the data expansion according to the experiment space with consideration of the existing dataset is obviously contributive in practice. Here Latin hypercube sampling is appropriate for the data expansion with the consideration of the existing dataset.

The mechanism of LHS with the consideration of the existing dataset is to divide the design of space of the original design variables into n+m intervals, where n and m intervals locate the existing samples and expanded samples, respectively. This expansion maintains existing samples at their original locations, and stochastically inserts expanded samples in empty positions.

The general procedures of LHS with consideration of existing data are listed below.
1. Divide the range of design variables into $(n+m)^2$ units, and generate a matrix of $(n+m) \times k$ to save candidate points, in which n refers to numbers of existing sample, m refers to the numbers of expanded samples, k refers to numbers of variables.
2. Search for empty unit in each column k.
3. For each variable column, randomly generate a value in one of the empty rows and record the value in the matrix.
4. The new generated matrix could accommodate more than m points, but when m=2n, the new matrix only has m inserted rows.
5. Only adopt the former m rows of the new matrix as result of data expansion, to ensure the former m rows are fully inserted while the deleted rows are partly filled.

7.12 Summary

This chapter begins by generally introducing the necessity and background of applying a meta-model for multi-domain, multi-objective optimisation. Then the methodology of construction of a meta-model via the selected method of genetic regression neural network hybrid with grey wolf optimiser (GWO-GRNN) is expatiated. The confirmation and selection by sensitive analysis of all utilised design variables (DVs) and objective variables (OVs) are discussed. The construction of the three meta-models for three performance domain are explained, including the mathematical expression, training procedures and meta-model assessment. Finally the discussion of the meta-model constructions are mentioned together with the possible future works.

The contribution of this chapter first prepared the meta-model of an urban scale research, aiming to replacing the original simulation models in the following complicated MD-MOO. The involvement of a large number of building morphological parameters and environmental performance metrics in a meta-model are a pioneer exploration in architecture and urban design field for instance of MDO. Secondly, the hybrid of GWO with GRNN is an innovative attempt in application of GRNN for more autonomous self-learning in a meta-model training. Moreover, its adoption in architecture and urban design field provides a new view aspect for future works regarding meta-model and optimisation.

Chapter 8 Multi-Domain Multi-objective Optimisation with Elitist Non-Dominated Sorting Genetic Algorithm (NSGAII)

In this chapter, integrated global optimisation of three acoustic, sunlight availability and thermal performance domains is operated. In other words, the multi-domain multi-objective optimisation (MD-MOO) is executed for the cross domain multi-objective problem over building morphology parameters, including single building and neighbourhood morphology attributes. Optimisation strategy and algorithms are first introduced, then the methodology of the selected MD-MOO optimisation algorithm Elitist Non-Dominated Sorting Genetic Algorithm (NSGAII) is expanded. Afterwards, the set up and procedures of the MD-MOO via NSGAII for the three-domain global optimisation of this research are expatiated. Optimised solutions will be analysed in detail in Chapter 9.
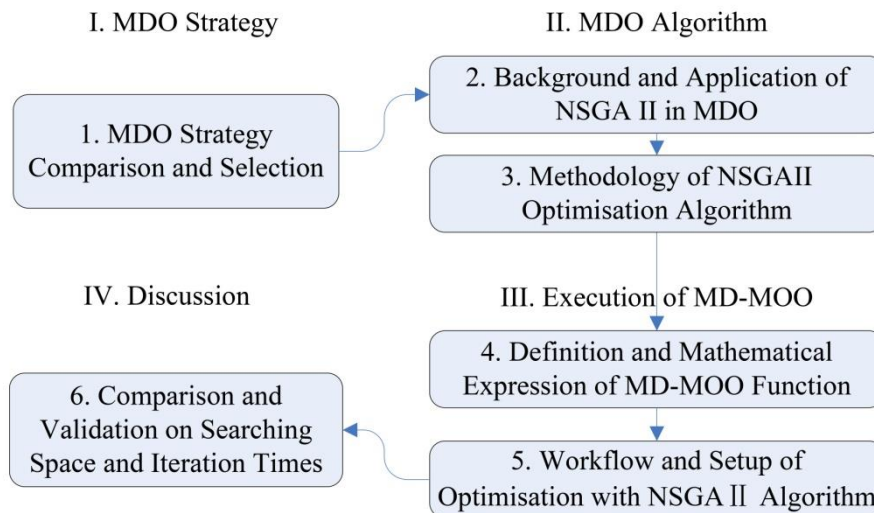
Figure 8.1 Content Structure of Chapter 8 MD-MOO with NSGAII,
The Expansion of Box 8 in Overall Content Structure

The structure of this chapter is: 8.1 MDO Strategy Comparison and Selection; 8.2 Background and Application of NSGAII in MDO; 8.3 Methodology of NSGAII Optimisation Algorithm; 8.4 Definition and Mathematical Expression of MD-MOO Function; 8.5 Workflow and Setup of Optimisation with NSGAII Algorithm; 8.6 Comparison and Validation on Searching Space and Iteration Times

Acronyms for Chapter 8

AAO/SAND	appositional analysis optimisation/simultaneous analysis and design
AIO	all-in-one method
BLISS	bi-level integrated system synthesis
CO	collaborative optimisation
CSSO	concurrent subspace optimisation
DV	design variable
GA	genetic algorithm
GRNN	genetic regression neural networ
IDF	individual disciplinary feasible method
MDF	multi-disciplinary feasible method
MD-MOF	multi-domain multi-objective function
MD-MOO	multi-domain multi-objective optimisation
MDO	multi-domain optimisation
MOO	multi-objective optimisation
NSGA	non-dominated sorting genetic algorithm
NSGAII	non-dominated sorting genetic algorithm with elitist strategy
NSGWO	non-dominated sorting grey wolf optimiser
OV	objective variable
SD-MOO	single domain multi-objective optimisation

8.1 Multi-Disciplinary Optimisation (MDO) Strategy Comparison and Selection

8.1.1 Introduction of Multi-Disciplinary Optimisation Strategy

Optimisation strategy of MDO refers to the structural method and calculation approach responding to the complexity in optimisation progress. MDO strategy consists of single-layer and multi-layer optimisation methods.

Single layer optimisation only conducts optimisation at a global system level, while in sub-system of multiple disciplines, only analysis and calculation within the field are conducted without any optimisation. Commonly used single layer method includes individual disciplinary feasible method (IDF)(Cramer *et al.*, 1994; Lee, 2004),

multi-disciplinary feasible method (MDF, also called all-in-one method, AIO)(Cramer *et al.*, 1994), and Simultaneous Analysis and Design (AAO or SAND)(Haftka, 1985; Cramer *et al.*, 1994). Aforementioned four single-level MDO strategies only have one optimiser, therefore the optimisation is finished in one step. This means there is only analysis but no optimisation for each of the sub-disciplines.

However, there exists the need for optimisation autonomy in each discipline in real practice from the experts and researchers in the corresponding fields. Hence, multi-level optimisation is extensively used in large scale complex systems of practical problems, which allows system level and sub-system level optimisation in various degrees of freedom. Multi-level optimisation strategy proposes concurrent optimisations of sub-systems of multiple disciplines on their own local variables, together with global optimisation of design variables and coordination among system and sub-systems. It allows optimisations separately in each discipline and afterwards coordinates sub-systems. Multi-level optimisation could reduce the calculation scale of global optimisation and loosen the coupling relationship among disciplines so that concurrent yet coordinated designs become possible in each system.

Multi-level optimisation could be classified by the collaborative strategy into a multi-level distributed system based on IDF strategy and MDF strategy. The former coordinates the sub-systems by applying consistency constraints of coupling variables at system level; while the latter fulfils the consistency requirement of coupling variables by multi-disciplinary analysis. The widely used multi-level optimisation methods include concurrent subspace optimisation (CSSO)(Sobieszczanski-Sobieski, 1982, 1988; Renaud and Gabriele, 1994; Park and Lee, 2001), collaborative optimisation (CO) (Braun, 1996; Tappeta and Renaud, 1997), bi-level integrated system synthesis (BLISS) (Sobieszczanski-Sobieski, Agte and Sandusky Robert, 1998; Sobieszczanski-Sobieski *et al.*, 2002), etc. In 21 century the main focus is improving the established single-level and multi-level MDO methods (Chen, 2021).

8.1.2 Comparison of Multi-Disciplinary Optimisation Strategies

Improvement of optimisation strategies of MDO is further required in current research and practice. At present the interests of improving the strategies focus on 1) application of approximate calculation approaches, i.e., response surface method; 2) hybrid of existing strategies, i.e., BLISCO method; 3) adoption of intelligent algorithm, i.e., genetic algorithm.

The most widely used MDO strategies, which are not only single level but are also multi-level strategies, are compared regarding to their advantages, disadvantages and scope of application (table 8.1).

Put into the scenario of integrated environmental optimisation of this research, it is impossible at present for an available simulation tool to handle a holistic multi-domain simulation covering aspects of acoustic, sunlight and thermal with considerable accuracy, which would be most helpful for system level optimisation iterations. The reasons are firstly, the complexity of involved design variables (DVs) and objective variables (OVs) and secondly, the huge differences in the estimation mechanisms and simulation model requirements of the various performance objectives in multi-domain predictions. Although Rhinoceros+Grasshopper appears to be the most promising at handling multi-domain analysis, even optimisation, currently the simulation algorithms for some fields (i.e., outdoor acoustic) are still too simplified and immature to achieve sufficient accuracy. Besides, this research interest is not aiming to develop a platform for handling multi-domain analysis. The predictions of the three individual performance domains: traffic noise attenuation, sunlight availability and thermal comfort, are accomplished in simulation software packages separately, with their inputs and outputs datasets collected, respectively. Therefore, MDO strategies containing global analysis are not executable for this research topic.

The nature of the multi-domain environmental optimisations determines that global optimisation is the focus of this research, and that independent optimisations in separate disciplines are also required for disciplinary autonomy. Because the independent simulations and analysis in three research fields are time and effort consuming, the substitutions of the original simulation models of high accuracy from various software by the meta-model of each domain is essential.

As seen from the multiple linear regressions of performance indices from three domains, the shared design variables are much less than independent design variables in all regression functions. This means that among the domains it has less coupling relationships, while the system variables are more than coupling independent variables in number. Matching with table 8.1, the overall condition of this research suggests an adoption of collaborative optimisation strategy of MDO.

Table 8.1 Comparison of MDO Strategies

	Advantages	Disadvantages	Scope of Application
Multi-disciplinary feasible (MDF) (Cramer *et al.*, 1994)	Simpler structure. Global analysis together with sub-system analysis.(Yi, Shin and Park, 2008) Huge advantages in search global optimums, due to sufficient consideration of multi-disciplinary coupling relation. (Balesdent *et al.*, 2012; Mohammad Zadeh and Sadat Shirazi, 2017)	A large amount of data exchange between disciplines due to global analysis. May have low efficiency in optimisation due to system analysis calculation. High optimisation time cost if requires large scale research question and high accuracy of disciplinary analysis. (Yi, Shin and Park, 2008)	Adequate for scenarios of less design variables. Of less disciplinary analysis calculation amount.
Individual disciplinary feasible (IDF) (Cramer *et al.*, 1994; Lee, 2004)	Avoiding global cooperative analysis among disciplines. Conducting concurrent independent analysis in each discipline and no consideration of disciplinary coupling improves the optimisation efficiency.(Yi, Shin and Park, 2008) Coordination of the independent analysis of sub-discipline and meeting consistency requirements of cooperative disciplines is expressed by introducing auxiliary coupling variable as design variables. (Balesdent *et al.*, 2012)	If having overmuch system variables with coupling relationship, the added auxiliary variable would reduce optimisation efficiency.	Fit for scenarios of less independent variable with coupling relationship. (Balling and Wilkinson, 1997) It is not fit for a large scale system. (Price, Keane and Holden, 2011)
Appositional analysis optimisation (AAO) (Haftka, 1985; Cramer *et al.*, 1994)	Adoption of optimisation machine to ensure the global feasibility of design solutions. Cooperative relationship among disciplines and meeting consistency requirements of cooperative disciplines are expressed by auxiliary equation constraints. Independent disciplinary analysis available. (Balling and Wilkinson, 1997; Balesdent *et al.*, 2012)	Optimise all variables including design variables, condition variables and input/output coupling variables that may cause redundancy in optimisation. (Padula, Alexandrov and Green, 1996)	Fit for scenario of high cost in system analysis and disciplinary analysis. (Yi, Shin and Park, 2008) Not widely used. Simplification on the constraints could derive multi-disciplinary

			feasible method (MDF) and individual disciplinary feasible method (IDF).
Collaborative optimisation (CO) (Kroo *et al.*, 1994; Braun, 1996; Tappeta and Renaud, 1997)	Simpler structure and easier to construct. The objective function of the system layer optimisation is identical to the original objective function of the research. Coordination of concurrent sub-system optimisation through application of consistency constraints at system level. Objective function of each sub-system is to minimise the discrepancy between the sub-system and the objective scheme distributed from the parent system. Sub-systems only exchange data with parent system. High grade of autonomy in disciplinary optimisation which is favourable by designers in individual field (Alexandrov and Lewis, 2002; Yi, Shin and Park, 2008; Balesdent *et al.*, 2012; Mohammad Zadeh and Sadat Shirazi, 2017)	If having many design variables, searching of global optimal in system objective optimisation and satisfying consistence constraints becomes difficult (Mohammad Zadeh and Sadat Shirazi, 2017) Convergence is not yet validated mathematically. (Balling and Wilkinson, 1997)	Fit for scenario of: low level of coupling relationship among disciplines, (Chen, Zhang and Khalid, 2002; Yi, Shin and Park, 2008) System variables significantly more than coupling variables among disciplines. (Marriage and Martins, 2008)
Bi-level integrated system synthesis (BLISS) (Sobieszczanski-Sobieski, Agte and Sandusky Robert, 1998; Sobieszczanski-Sobieski et al., 2002)	Flexibility, autonomy and artificial interference available. Separates design variables into global/system design variables and local/disciplinary design variables. (Mohammad Zadeh and Sadat Shirazi, 2017) Constructs approximate function of original objective function by Taylor's expansion. Alternately optimises the global and local variables depending on the guidance of gradient change. (Chen, Zhang and Khalid, 2002; Balesdent *et al.*, 2012)	Convergence is impacted by selection of initial value. Sensitivity analysis and partial derivative calculation greatly increase calculation cost. (Yi, Shin and Park, 2008)	Fit for scenario of few system variables, more disciplinary design variables, and not being high level of nonlinearity and non-convex. (Yi, Shin and Park, 2008)

8.1.3 Collaborative Optimisation Strategy

Collaborative optimisation (CO) is the most widely researched and used MDO strategy which was first proposed (Kroo *et al.*, 1994) as distributed collaborative optimization algorithm with two-level structure. It hierarchically disassembles the optimisation system into optimisation at a system level and a sub-disciplinary level on the basis of consistency constraints and the idea that sub-systems only exchange data with parent system. The objective function of the system layer optimisation is identical to the original objective function of the research, while the objective function of each sub-system is to minimise the discrepancy between the sub-system and the objective scheme distributed from the parent system. The overall concurrent optimisation is coordinated by the system layer through consistency constraints over sub-systems. Collaborative optimisation has a high grade of autonomy in disciplinary optimisation which is favourable by designers in individual fields. However, its convergence is not yet mathematically validated.

Comparing to other optimisation strategies, collaborative optimisation strategy has the advantages of:

1. Sufficient degree of freedom in design for each sub-discipline, allowing independent disciplinary optimisation resulting in higher influence power of sub-systems to the whole system, making it welcomed by researchers in multiple disciplines,

2. In the scenario of large-scale global systems, the sensitive level to coupling relation among disciplines is reduced.

3. Avoiding time-consuming multi-disciplinary analysis by introducing auxiliary design variables, Also avoiding calculation of the solution of the non-linear coupling equation set so that coordination among disciplines is expressed though consistency constraints at system level.

4. Barely having a requirement on the initial value of variables, and having high stability and approximation convergence.

Meanwhile, the short coming of collaborative optimisation concentrates on the convergence at system level. Because the introduction of consistency equation constraints into system level optimisation, the non-linearity significantly appears in system level optimisation, hence quite often there would not exist a Lagrangian multiplier at optimal point, so that the Kuhn-Tucker condition could not be made – therefore, it is not mathematically validated. It is possible that CO strategy is trapped at local optimal, because the consistency constraints at system level is not directly related to design variables under optimisation, therefore resulting in non-smooth and discontinuous features in constraint functions.

The current researched approaches improving convergence problem of CO are mainly concerned with three aspects: meta-model, optimisation strategy and optimisation algorithm. With application on the Kriging model, a system level meta-model is constructed (ZHANG, Bailin and Weihua, 2010). Response surface method is used for the improvement of CO (Sobieski, Manning and Kroo, 1998). Penalty function method at sub-system level and relaxation factor method are used to reduce the jumping feature of feasible domain (Braun and Kroo, 1997). The Optimisation strategy at system level is improved by the application of a linear weight coefficient, to transfer a multi-objective question into a single-objective question (Tappeta and Renaud, 1997). CO is also hybridised with linear programming to solve the multi-objective question (McAllister *et al.*, 2005). Altered Pareto algorithm has proven that multi-objective genetic algorithm has less sensitivity on objective functions and constraints and the convergence criteria could overcome Lagrangian multiplier (Long *et al.*, 2008)

According to the aforementioned difficulties of the application of CO to the current research, two improvements on the optimisation strategy are adopted:

1. The meta-model of general regression neural network (GRNN) with is adopted for increase on solving efficiency and accuracy.

2. Non-dominated sorting genetic algorithm with elitist strategy (NSGAII) is adopted for avoidance of calculation difficulties on consistency constraints.

Figure 8.2 represents the overall structure of multi-disciplinary optimisation adopting collaborative optimisation strategy. The structure is divided into three layers and data exchange from the layer of single-disciplinary analysis to single-disciplinary optimisation and multi-disciplinary optimisation.
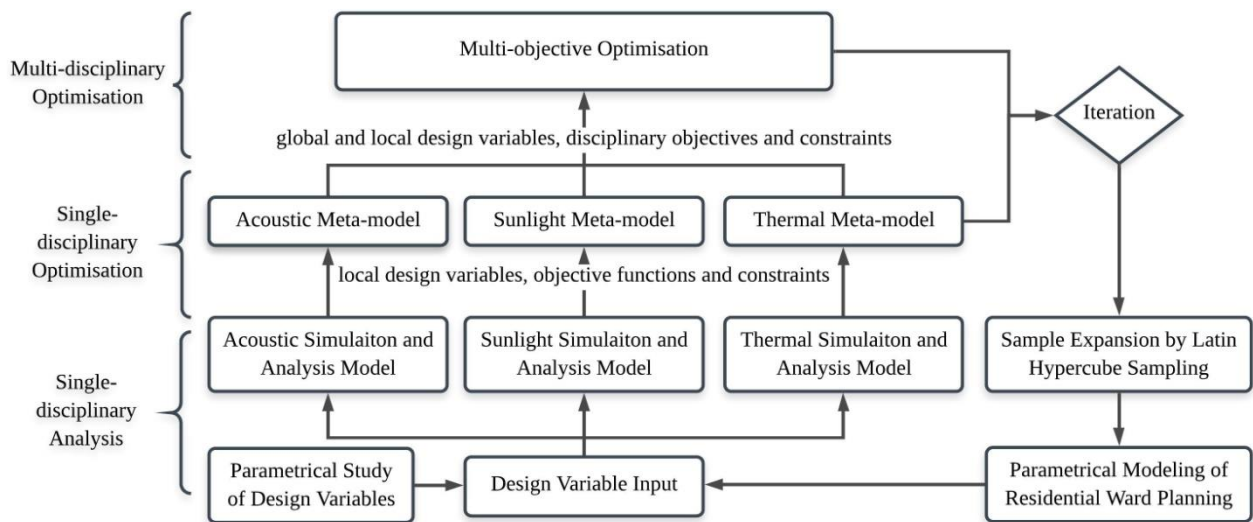
Figure 8.2 Structure of Multi-Disciplinary Optimisation Adopting Collaborative Optimisation Strategy

8.2 Background and Application of Non-Dominated Sorting Genetic Algorithm with Elitist Strategy (NSGAII) in MDO

As the complexity of multi-domain multi-objective optimisation and non-analysability of the meta-model, the advanced design-space searching algorithm is required for identification of the optimal solutions. In this section, the non-dominated sorting genetic algorithm with elitist strategy is introduced with its relationship to NSGA. The background of multiple objective optimisation and development and improvement of algorithms is review in section 2.2. Also the application of multiple MDO optimisation algorithms is review in architectural fields regarding to acoustic, sunlight, thermal and cross-domain scenarios in section 2.3.

Since the selection of proper optimisation algorithm is vital according to the requirement of the optimisation problem. In this research, the final aim is to provide a selectable group of optimised solutions with equal performance level. The definition of NSGA algorithm matches the basic requiremt of the aim of this research.

Non-dominated sorting genetic algorithm (NSGA) is one of the widest used evolutionary algorithms, since its development is aiming to solve multi-objective optimisation with multiple optimums (Deb et al., 2000, 2002; Deb and Goel, 2001). It is an extension version of genetic algorithm (GA) for the optimisation of multi-objective function. It is suitable for the continuous function of multi-objective optimisation (MOO) problems. The aim of NSGA is to increase the adaptive fitness of the candidate solutions to the Pareto front under a set of constraints in the objective functions. The optimisation is achieved through evolutionary operators of selection, genetic crossover and genetic mutation with assistance of meta-models. The candidate solutions are sorted according to the sequence of Pareto dominance, and the similarities between solutions are evaluated on the Pareto front. The sorting and comparison is utilised for promotion of solution diversity (Brownlee, 2011).

Here, Pareto solution is also termed as non-dominated solutions first presented by Pareto in 1966. When having multiple objectives, due to conflicts and incomparability among various objectives, the best solution on one objective could be the worst for another. Hence, when one solution improves performance of one objective function, at least one of the rest functions would be impaired. Hence, for multi-objective optimisation, usually there exists a solution set in which one solution could not outperform the others within that same solution set. In other words, except for these solutions it is impossible to have another solution which improves any objective without impairing at least one another objective. The solutions in this solution set are named Pareto solutions, the relationship of which is specified as: If existing two solutions S1 and S2, for all objectives S1 is better than S2, namely S1 dominates S2.

If S1 is not dominated by any other solutions, then S1 is named non-dominated solution or Pareto solution. A group of optimum solutions from a group of objective functions is named Pareto solution set. The surface formed by scatters of optimised objectives of Pareto solutions is named Pareto front surface.

NSGAII is a multi-objective optimisation algorithm adopted elitist strategy over NSGA developed by Deb et al. (Deb *et al.*, 2002). The basic idea of NSGAII is to select Pareto solutions close to the true Pareto front and uniformly spread on the front during the selection, crossover and mutation progress of GA, according to the non-dominance ranking and crowding distance of each solution. During iterations, the parent and offspring population are combined to ensure elitist solutions are saved.

NSGAII is an improved version of NSGA, which alleviated the previous limitations of high computational complexity, non-elitism approach and the requirement of sharing parameter specification. NSGAII improves the above mentioned limitations from three aspects:

1. Present a fast non-dominated sorting method to reduce calculation complexity.
2. Introduce a crowding degree and crowded-comparison operator to substitute the fitness sharing strategy which needs to specify sharing diameter. Crowded-comparison operator provides comparison criterion after fast sorting, by which members of quasi-Pareto solution set could be uniformly spread out on the Pareto front. This ensures the diversity of the population.
3. Introduce elitist strategy to enlarge sampling space. Combination of the parent population with its next offspring population to form the next generation encourages the well-performed individual in parent population to be inherited into next generation. Furthermore, by saving individuals from the population in various layers, the optimal individuals would faster improve the population.

To summarise, NSGAII possesses advantages in convergence closer to the true Pareto front, size of hyper volume and uniform spread of optimal front. It has been proved as

reliable in building related optimisations.

Hence the optimising aim of NSGA is matching the requirement of the optimisation problem in this research. NSGAII is tested being more effective in searching design space, it is selected as a basis of making improvement in the algorithm.

8.3 Methodology of Multi-Domain Multi-Objective Optimisation with NSGAII

In this section, the mechanism and workflow of NSGAII is explained. The convergence constraints of NSGAII are also stated.

8.3.1   Mechanism of Multi-objective optimisation with NSGAII

NSGAII is the hybrid version of a non-dominated sorting and genetic algorithm. Genetic algorithm appears as the searching engine. The non-dominated sorting method collects and selects the optimal solutions.

With the innovations of fast non-dominated sorting, crowding distance and crowded-comparison operator, the core of NSGAII is formed.

Fast non-dominated sorting is an improved sorting approach compared to NSGA to rank candidate solutions into hierarchical Pareto fronts. The calculation complexity is greatly reduced. In calculating fast non-dominated sorting, two parameters n_p and S_p of each candidate solutions in the population are calculated. Parameter n_p is the counts of solutions dominating solution p and S_p is the set of solutions in population dominated by solution p. The calculation follows the procedures:

1. Find all solutions in population of which n_p=0, and save them in set F1, represent the first Pareto front.

2. For each solution i in F1, the individuals dominated by i are noted as set S_i. Calculate $n_l = n_l - 1$ for each member l in S_i. If $n_l = 0$, save individual l in set F2, as the second front.

3. Repeat above steps until all individuals are hierarchically divided in the population and all fronts are identified.

Crowding distance is the density estimative metric, indicating the density of solutions surrounding a given solution. It is represented with the average side length of the cuboid formed by its nearest neighbours' vertices and only containing the given individual, noted as i_d. The calculation of i_d is noted as below:

Set $i_d = 0$, i=1, 2,…, N

for each objective function f_m,

1. Sort the population based on this objective function f_m value in ascending order.

2. Set the two solutions on the boundary having infinite value of crowdedness, namely $1_d = N_d = \infty$

3. Calculate $n_d = n_d + \left(f_m(i+1) - f_m(i-1)\right) \big/ \left(f_m^{\max} - f_m^{\min}\right)$, n=2，3，…，N-1

After fast non-dominated sorting and calculation of crowding distance, each solution in the population is assigned with two parameters: the non-domination rank n_{rank} and the crowding distance n_d. With utilisation of these two parameters, dominative relationship between any two individuals in the population could be determined through the execution of the crowded-comparison operator. The criterion of determination is i dominating j, namely individual i is better than j if and only if $i_{rank} \leq j_{rank}$ and $i_d > j_d$.

8.3.2   General Workflow of Multi-objective Optimisation with NSGAII

The main procedures of applying NSGAII for optimum searching consists of population initialisation, fitness evaluation, non-dominated sorting, ranking based on crowding distance, elitist selection, bimodal crossover and mutation (Yang *et al.*, 2017). The procedures are stated as following:

1. Randomly generate initial population in size of N and after non-dominated sorting acquire first offspring generation through selection, recombination, crossover and mutation in genetic algorithm (GA).

2. From the second generation, combine the parent generation with the offspring generation. Then execute fast non-dominated sorting, meanwhile calculate crowding distance for members in non-dominated hierarchy. Based on non-domination rankings and the crowding distances between members, solutions are selected by the crowded-comparison operator to form a new parent generation.

3. Generate new offspring generation through general genetic algorithm procedures, until number of iteration is met.

8.3.3   Convergence and Update of Global Multi-objective Optimisation

The global optimisation would stop when convergence constraints are met. Generally speaking, there are three types of convergence constraints: manually stop, stop when feasible, optimal improvement and predictive accuracy standards met, or when optimisation cost standard met. The cost standards are utilised as convergence constraints in this research. The rationale is that the diversity and practical significance of the Pareto solutions would fall below expectation, as shown in the result acquired under convergence constraints of feasibility, optimal improvement and predictive accuracy. The adopted cost standard is the number of iterations.

The predictive accuracy of the MOO is improved through iterations of optimisation procedures, as well as the update of the meta-model of each domain. After the feasibility analysis of Pareto solutions, feasible optimal solutions are acquired. A

reverse encoding should be executed to restore the building distribution patterns according to the value combinations of the DVs in the optimal solutions. The reverse encoding could be realised via a parametric modelling of the building distribution models based by design variables. Once the 3D building distribution models of the feasible optimal solutions are generated by modelling tools, and by feeding the restored distribution models to high accuracy simulation software packages, the simulation results of optimised building distribution schemes are achieved. These data pairs of optimised design scheme and optimised performance results are the basis of updating meta-model trainings and following global optimisation iterations. In this manner, the global system iteration is implemented over a self-updated optimisation system.

Without the update of meta-models of each domain, the global optimisation system would still achieve convergence constraints with acceptable accuracy through iterations of optimum searching. But the update of the meta-model itself contributes to a closer fitness of the estimation of the real simulation model, which provides an accurate basis of global optimisation. This will further improve prediction and optimisation accuracy of the whole system.

However, the difficulties for the system iteration are high dimensional parametric modelling for the high-dimension urban scale residential ward model, with 20 or more design parameters to manage. As mentioned by Marsault, discretisation of space in to votices is the the way to gernate urban forms as used in this research. It is highly limited by the large number of parameters and long calculation time in the process of generation. It is noted that there is no universal mathematical generator of forms which handles a overall urban scenario(Marsault, 2017). As for the length limitation of this research, high-dimension parametric modelling for a meso-scale model is not expanded.

Hence, with the absence of a high-dimensional parametric model, the single-domain optimisation, meta-model updating to which parametric model is obligatory is not conducted in this thesis. However, without the updating, the overall structure of collaborative optimisation of MD-MOO and the iteration and prediction accuracy of MD-MOO would not be affected. The feasible optimal solutions of the MD-MOO are validated through case studies. The details of validation are expatiated in section 9.5.2 and 9.5.3 in Chapter 9.

In future work, the high-dimensional parametric modelling method needs to be explored first, to generate complex residential neighbourhood models. The chunk of parametric modelling needs to connected to simulation software packages of multiple performance objectives in the manner of automatic data input and result record. The chunk of modelling-simulation would be inserted into the MD-MOO system. It will perform as an input resource of data pairs of DVs and OVs, and an output receiver of DV data of the feasible optimal solutions from Pareto solutions. The final convergence constraints of MD-MOO could also be defined as the discrepancy between the updated OV data from modelling-simulation chunk and optimised OV data from MD-MOO, of feasible optimal solutions from Pareto solutions

8.4 Definition and Mathematical Expression of MD-MOO Function

Following the framework of collaborative optimisation strategy, single-domain meta-model is constructed utilising the data pairs imported from high accuracy simulation model of the three domains: traffic noise attenuation, sunlight possibility and thermal comfort in residential wards in South-east China. Single-domain optimisation applies the meta-model as an objective function; while for multi-domain optimisation, the single-domain optimisation objective functions are combined to be the global optimisation objective or multi-domain multi-objective function (MD-MOF).

As for the coupling relationships of three domains, through shared design variables, the global optimisation design variables are separated into shared DVs and local DVs of domain-exclusive building distribution parameters. The MD-MOO is the matrix of weighted single-domain objectives.

The optimisation objective is expressed mathematically as equation:

$$Find\ X = \left\{ X_{sk}, X_{lj} \right\}^{T}$$
$$Min\ \mathrm{M}\left(X_{sk}, X_{lj}, Z \right)$$
$$s.t. \qquad\qquad (8.1)$$
$$G\left(X_{sk}, X_{lj}, Z \right) \leq 0$$
$$H\left(X_{sk}, X_{lj}, Z \right) = 0$$

Where,

M represents the multi-domain multi-objective optimisation function.

Z represents the objective function value of the global evaluation variable.

X refers 20 coupling design variables, namely X is expressed as a 20 dimension vector.

X consists of X_{sk} and X_{lj}, respectively refer to shared variables and local variables.

k, j are the number of shared variables and local variables respectively. k+j=20,

For this research, Z is global optimisation objective, in relation to performance of three domains. Z is a weighted matrix of single-domain optimisation objectives (SD-MOO), expressed as equation:

$$Z = [\omega_1 \cdot D_1, \omega_2 \cdot D_2, \omega_3 \cdot D_3] \quad (8.2)$$

Where:

Z represents global optimisation objective.

D represents single-domain multi-objective objective.

ω_n is the weight of each domain objectives; n=1, 2, 3.

n is the number of disciplinary-objectives in one system.

8.5 Workflow and Setup of Optimisation with NSGAII Algorithm

NSGAII is the optimisation method of design space searching to acquire non-dominated solution sets. Therefore the framework of NSGAII contains objective definition, optimisation setting, main searching loop and Pareto solution output. The detailed procedures are expressed as below:

1. Conform single-domain optimisation objectives

In this research context, single-domain optimisation objectives result in the quietest and evenest noise attenuation performance, highest and evenest sunlight possibility, and best overall thermal comfort including temperature, wind speed and radiation. These domain objectives are constructed by sub-objectives of performance evaluation functions in each domain. The mathematical expressions of three domain objectives D_m refer to equation 7.11.

2. Form global optimisation objective by single-domain objectives

M is the multi-domain objective function which consists of single-domain objective matrix. The weights of three domains are evenly set as 1, so that no design preferences are involved. In future work when design preference existed, proportional weight could be applied to multiple domains, to make summation of weight equals to 1.

3. Define design variables and range in global optimisation

The 20-dimension design variables are applied and are all shared variables. The 18 less significant variables are not adopted in NSGAII optimisation, due to the limitation of meta-model accuracy as trained by less sufficient samples.

The lower and upper bounds of design variables are specified as in table 7.2 and 7.3, to limit searching space. Several optimisation attempts, for instance real design space

searching, partly unconstrained searching and unconstrained searching, are conducted by altering the ranges of the design variables.

4. Set up of NSGAII parameters

Maximum iteration of 50,100, 200, 400 and 600 are specified and tested for different Pareto solution sets. Optimal solution population size is set as 100. Crossover and Mutation percentage are 0.9 and 0.4 respectively. Mutation rate is set as 0.02 with step size of 1. Rules for the calculation of the number of parents and mutants are also mathematically expressed.

5. Initialise parameters in NSGAII

Utilising parameters in NSGAII need initialisation before starting the main loop, which are position, cost function value, rank, domination set, dominated count and crowding distance.

Calculation rules of the parameters are also stated in the Matlab program, which include:

- The searching method of one position in the design space, and calculation of the cost function value.

- The method of operating non-dominated sorting means searching for solutions which is not dominated by any other solutions, namely its dominated counts equals to 0. Note un-dominated solutions in Pareto solution set and rank them.

- The method of calculating crowding distance.

- Sort population: based on crowding distance and based on rank.

6. Implement of main loop of optimisation by NSGAII

After setting up all requirement variables and corresponding information, the main loop of optimisation starts. Within maximum iteration times as set, loop the following steps till convergence:

- Crossover

- Mutate

- Merge population

- Calculate crowding distance

- Sort population

- Truncate

- Non-dominated sorting

- Calculate crowding distance

- Sort population

After the loop is stopped, store Pareto solution set as F1, and display iteration information, plot Pareto front, which is normalised cost function values of Pareto solutions in 3 dimensions with each domain assigned to one dimension.

7. Pareto solutions consolidation and screening

Further data consolidation is required by reverse normalisation of design variable values of Pareto solutions and prediction of their single-domain sub-objectives by corresponding meta-models. In order to understand the physical meaning of Pareto solutions, recovery of the sub-objective value in the real range, namely of the environmental performance evaluation variables, from optimisation results are essential. In this research the 400 Pareto solution size is predefined.

After acquiring Pareto solutions and corresponding performance evaluation variable values in real range, data pair screening is conducted based the proposed constraints between evaluation variables, stated in table 7.3. These constraints represent the numerical coupling relationship between the variables based on their definitions. The Pareto solutions which passed the screening are stored as feasible optimums of the global optimisation solutions (F). Integrated performance plot of feasible Pareto solutions is generated based on solutions in F.

The constraints are not inserted in the optimisation process because adding

optimisation constraints in an NSGAII approach is complicated and greatly increases calculation cost. Involving constraints in optimisation process is obviously inefficient when solution screening is available and valid. As a comparison, in previous research adopting non-dominated sorting grey wolf optimiser (NSGWO) for MD-MOO, the screening is used before non-dominated sorting. The solutions being searched are stored in an archive, then non-dominated sorting is executed over the solution in the archive to select feasible optimal solution. The less dominated solutions will be removed from the archive and new searched solutions will fill the archive accordingly (Jangir and Jangir, 2018).

The flowchart of NSGAII optimisation is shown in figure 8.3.
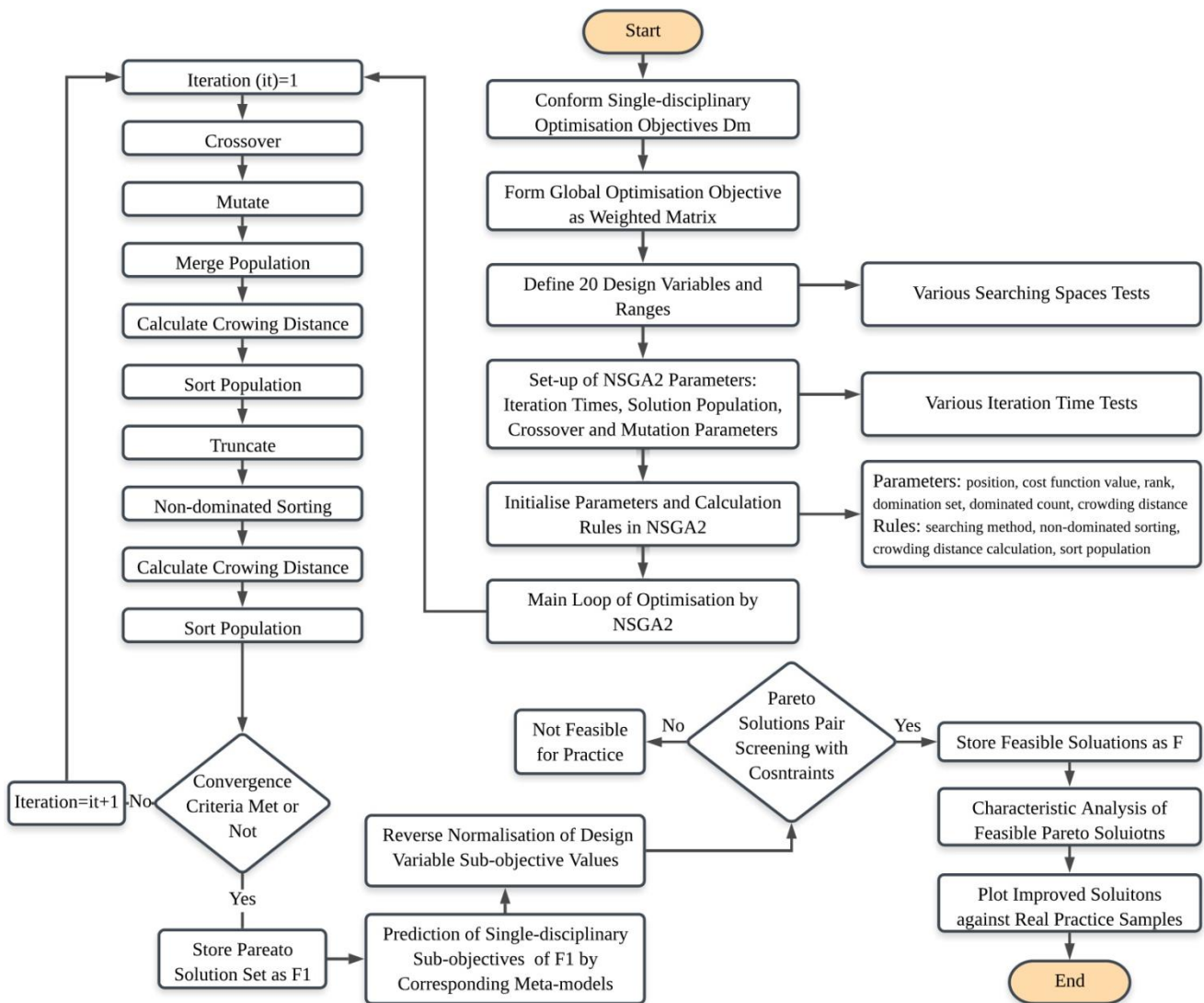
Figure 8.3 Flowchart of NSGAII Optimisation in This Research

8.6 Comparison and Validation on Searching Space and Iteration Times

In the procedure of optimisation, multiple attempts are operated with concerns from two aspects: different searching spaces and various iteration times.

Implied by altering the allowed range of design variables, three types of optimum searching are attempted: real design space searching, partly-unconstrained searching and unconstrained searching. The real design space searching refers to optimum searching only operated in design space limited by the range defined based on physical meaning of all design variables. In this scenario, the ranges reflect the interrelations between design variables as stated in constraints applied in solution screening. Hence, Pareto solutions searched in real design space have the highest proportion of passes in the solution screening - even 100% - whereas the solutions are highly similar in design form to those that lost the diversity of optimal results.

Because this research attempts to discover unbiased optimal design form out of local design habit and conventions, a diversity of optimal solution is expected from the optimisation. Therefore, fully unconstrained searching is also attempted to find any possible optimal solutions that are different from the current design form. The searching result fulfils the expectation of diversity, and clearly reflects the structure of meta-models due to the openness in searching range. Therefore, solutions which passed the constrained screening possess clear characteristic discrepancies. However, a considerable number of solutions with an SPL-L10 value below 50dBA, which is far too low and very idealistic from real conditions in practice, would be excluded from the screening process even if the design form has significant meaning. In other words, searching in unconstrained space results in solution diversity, but a considerable number of solutions are of extreme or near-extreme scenarios. Although these solutions are not directly applicable in practice, they provide novel design patterns

and routines for designers and planners, which is still preferable.

To adjust the result from unconstrained searching, partly-unconstrained searching is operated. While understanding the limitation of fully unconstrained searching, which is less efficient in searching of feasible solutions due to SPL-L10 below 50dBA, the searching space is defined as free for all other variables except for SPL-L10, which ought to be higher than 70dBA. The results show considerable diversity and feasibility when all SPL-L10 over 70dBA. Therefore, the final optimal solutions are searched in partly constrained range and followed by feasible analysis of solutions.

In order to acquire a Pareto solution set of diverse optimal patterns instead of optimum concentration on a single pattern, various iteration times are tested. The nature of searching a non-dominated solution set is to find solutions of discrepancies but all perform incomparably well. Hence, the convergence constraint of this research is not defined as the difference between solutions to below a certain level, but as hierarchical numbers of iterations in order to review the exploration and exploitation level in the Pareto solution set. A good balance of exploration and exploitation level is expected for the requirement of diverse design patterns of good integrated performances.

Iteration of 50, 100, 200, 400, 600 and 800 times are tested. Based on monitoring on the Pareto front plotted 3D for each iteration during optimisation and output data processing after optimisation, the following conclusions are given.

Iteration of 50 times could not sufficiently search the design space. Compared to results from other iterations, the solutions which passed screening of 50 iterations could not ensure presence of the real optimum. The result of 100 iterations shows various patterns of optimum and has a relatively low level of solution concentration on each pattern. Over 70% of solutions searched could pass screening as feasible solutions. The solution set of 200 iterations is similar in characteristics to that of 100

iterations, whereas it has a higher level of concentration of solution on one pattern and close to 100% in passing screening.

For 400 iterations, the solutions are clearly clustered on similar design patterns; nevertheless the proportion of passing screening reduces. The feasible solutions after screening remain considerable diversity. Regarding results of iterations over 600 times, the solutions are highly crowded or sufficiently converged, which is not desirable for this research. The feasible solutions which passed screening usually point to one single optimum with minute variance on the design variable values.

On the basis of the iteration test, 400 iterations are adopted for an appropriate balance of exploration and exploitation for this research and searching population adopts 400 to assure the number of feasible solutions after constraint filtering.

8.7 Summary

In this chapter, at first, it is expatiated that the background of multi-domain multi-objective strategy and optimisation algorithms. Concluded from the comparison, collaborative optimisation strategy is utilised for the overall structure of the global optimisation of this MD-MOO problem. Elitist Non-Dominated Sorting Genetic Algorithm (NSGAII) is adopted as the algorithm. Afterwards, the methodology of NSGAII applied in this research is explained, followed by the expounding of the mathematical expression, workflow and setup of the MD-MOO with NSGAII. Finally, the comparison and validation of the searching space and iteration numbers are discussed.

The contribution stated in this chapter is first the selection and application of a collaborative optimisation strategy in a multiple performance domain scenario in architectural and urban design background. The three layers structure and

characteristics of the strategy suit many similar scenarios in architectural and urban design background regarding the shared input data attributes and simulation procedures.

Secondly, with the insertion of parametric model-automatic simulation chunks in future study, the self-updating global optimisation system built in this research is ready to handle various cross domain multi-objective optimisation. Not only global optimisation, but also local optimisation of each performance domain is also simultaneously capable.

Thirdly, although NSGAII has been widely used in energy consumption and thermal related researches, most are single objective optimisation searching for the only optimal. This thesis is concerned with the concentration of multi-objective optimisation with tools of the Pareto solution set. It allows the collection of several equally well performed design schemes for multiple consideration aspects, which is the closest reproduction to the real design process.

Chapter 9 Analysis and Interpretation over Feasible Pareto Solution Set

Among the mentioned and compared multi-objective optimisation algorithms in chapter 2, the NSGAII is selected as the global optimisation tool. The Chapter 8 presents in detail about the background, methodology and application of NSGAII in this research. In order to enable the NSGAII to search for the best trade-offs among all objectives, the Pareto optimal dominance calculation is used for the comparison of possible solutions. The solutions which perform equally as well as each other are defined as the Pareto solution set, and the plot formed by the solution points is named the Pareto front.

The Pareto solution set is especially suitable in this research to choose from for the special needs of application of the optimal designs. For architectural and urban design scenarios, it would be preferential for designers to refer to and choose from sufficient templates of optimal designs from a choice of several design patterns or styles. The Pareto solution set provides the diversity of optimums with equal performance while also allowing widely different patterns. The iterations could fill the pool of optimums in the Pareto solution set under each design pattern to allow template selection from the designer to attribute discrepancies between similar design patterns.

This section interprets the result of the multi-domain multi-objective optimisation involving three environmental performance domains. The integrated optimisation of this MD-MOO system is presented by the improvement of global performance from the original sample to feasible Pareto solutions, the improvement of single domain performance from the original sample to the feasible Pareto solutions, three design pattern suggestions for residential ward design in South-East China, general design rules, and an interactive webpage tool for easy and fast access to a new design's potential of three domain integrated performance and possible adjustment directions

with design preference. For a flowchart of this chapter see figure 9.1.

I. Analysis Preparation   II. Result Analysis   III. Design Scheme Suggestions

1. Data Transformation for Feasible Pareto Solutions Analysis

2. Analysis over Feasible Pareto Solution Set -MD Performance Improvement

3. Analysis over Feasible Pareto Solution Set -SD Performance Improvement

4. Suggested Residential Ward Distribution Based on Feasible Pareto Solutions

4.1 Spider Diagrams of Performance Balance and Distribution Arrangement

4.2 Suggested Design Pattern A

4.3 Suggested Design Pattern B

4.4 Suggested Design Pattern C

5. Validation of Suggested Design Schemes

6. Web-based Interactive Tool for Application of MD-MOO Results

Figure 9.1 Content Structure of Chapter 9 Analysis and Interpretation over Feasible Pareto Solution Set, The expansion of box 9 in Overall Content Structure

The structure of this chapter is as follows: 9.1 Data transformation for feasible Pareto solutions analysis; 9.2 Analysis of global/integrated multi-domain performance improvement; 9.3 Analysis of single-domain performance improvement; 9.4 Suggested residential ward distribution based on feasible Pareto solutions; 9.5 Validation of suggested design schemes; 9.6 Interactive tool demonstrating scheme selection.

Acronyms for Chapter 9

aCAH	average corner area high value
aD	average distance to road
aFLmin	average min facade length
aFR	average facade ratio
aIDH	average interval length high value
aIDL	average interval depth low value
aILmax	average max interval length
aLFL	average low-rise facade length
aLFR	average low-rise facade ratio
aOL	average outline length
APSH-IQR	interquartile range of APSH
APSH-P20	P20 of APSH
APSH-P50	P50 of APSH
APSH-P70	P70 of APSH

APSH-THR(0)	ratio of APSH=0h
APSH-THR(413）	ratio of APSH<413h
AR	aspect ratio
aTCA	total corner area
BD	building density
DV	design variable
FPA	foot print area
HRBR	high-rise ratio
LRE-P70	P70 of longwave radiation from environment
MD-MOO	multi-domain multi-objective optimisation
MRT-P25	P25 of mean radiant temperature
NDO	normalise domain objective
OV	objective variable
PR	plot ratio
RBA	residential building area
SD-MOF	single domain multi-objective function
SD-SOO	single domain single objective optimisation
SF	site shape factor
SPL-IQR	interquartile range of SPL
SPL-P10	P10 of sound pressure level
SPL-P40	P40 of sound pressure level
SPL-P70	P70 of sound pressure level
SPL-THR(65)	ratio of SPL<65dBA
SVF	sky view factor
TSD	standard deviation of triangle area
WPSH-IQR	interquartile range of WPSH
WPSH-P30	P30 of WPSH
WPSH-P50	P50 of WPSH
WPSH-P70	P70 of WPSH
WPSH-THR(0)	ratio of WPSH=0h
WPSH-THR(83)	ratio of WPSH<83h
WS-Max	maximum of wind speed
WS-P20	P20 of wind speed
WS-P70	P70 of wind speed

9.1 Data Transformation for Feasible Pareto Solutions Analysis

For the convenience and accuracy of optimisation all input data including building morphology data and domain performance data is normalised before meta-model training and MD-MOO. Hence, the final result of Pareto solutions is normalised data in order of magnitude, which requires data recovery to real range following analysis.

In the beginning, the collected data consists of the building morphology parameter value and performance indices value from acoustic, sunlight and thermal fields in their own real range. The performance indices refer to the statistical measurements of one performance parameter, for example SPL-P10 as for SPL.

Then, the above data is normalised with the function of Matlab (equation 9.1). This normalises data in matrix X according to the rules trained and noted as in structure PS, and the normalised data is stored in matrix Y. In this research, all data sets are normalised to range of [-1,1]; in other words, the projection rules noted in the normalised structure are of a minimum value to -1 and a maximum value to 1.

$$Y = mapminmax('apply',X,PS) \quad (9.1)$$

The Normalised building morphology parameter values are noted as DV values. The normalised performance indices values are noted as OV values. For the matrix names for each OV see Matlab manuscripts in Appendices B and C.

The following meta-model training adopts the normalised DV and OV values and the normalised range of variables. During the MD-MOO, the meta-model is calculated which is the estimation of SD-MOF. This is the optimisation function searching for the optimal solution for the single domain multi-objective over all DVs. This single domain multi-objective is Dm mentioned in equation 7.11, which is a weighted summation of sub-objectives in a domain (SD-SOO), or a normalised component of performance metric of the domain, or normalised performance indices of the domain. It also refers to Normalise Domain Objective (NDO) value in chapter 9.

Considering the coupling constraints among DVs, when searching for optimal Dm, its range is the weighted summation of the normalised range of OVs of one domain. Applying the meta-model, NDO values of the original samples are calculated in order to plot feasible Pareto solutions with original samples. The range of the initial design NDO and optimised NDO are also noted for comparison as in table 9.1.

For easy understanding of the feasible Pareto solutions and their promoted performance, it is necessary to acquire the values of their morphology parameters and performance indices. By applying the meta-model, with input of the DV values of all feasible Pareto solutions, the f_n mentioned equation 7.11 could be calculated. The f_n is the normalised component of the performance metric of the domain in total number of 23, noted as N_s, S_v, T_w for acoustic, sunlight and thermal domains, respectively. The normalised value of OVs could be calculated based on equation 7.7, 7.8, 7.9, which are Y_n (n=1, 2…23). Then, reverse normalisation will be operated for al DV and OV values of the solutions, with the function in Matlab as equation 9.2, resulting in a reverse calculation of matrix Y under normalised rules as noted in structure PS; the reverse normalised matrix will be stored in matrix X. For further details, see Appendix C.

$$X = \text{mapminmax('reverse',Y,PS)} \qquad (9.2)$$

Therefore, data pairs of performance indices and morphology parameter values in real range for feasible solution data are achieved. The following analyses of the solutions for design suggestions and rules are based on this data.

9.2 Analysis over Feasible Pareto Solution Set

——Multi-Domain Performance Improvement

From the previous global optimisation of three domains of 400 iterations, data pairs from Pareto solutions of 400 populations are collected. Of these, 232 feasible Pareto solutions have passed the design variable constrained screening. The screening filters out the solutions with extreme design variable values, so that the feasible Pareto solutions have practical combinations of design variable and are applicable to practise.

9.2.1   Integrated Performance Improvement from Global Optimisation and the Evaluative Parameter: Normalised Domain Objective (NDO) of Feasible Pareto Solutions

To demonstrate the feasible Pareto front, the feasible Pareto solutions are plotted as points in a three-dimension coordinate which refers to performances in traffic noise attenuation, sunlight availability and thermal comfort domains. The point position is specified by the three normalised domain objective values in each domain. The normalised domain objective (NDO) is the normalised version of weighted summations of domain component objectives (from optimisation viewpoint) or domain performance indices (from design viewpoint), noted D_m as in Chapter 8. NDO is the global performance assessing index applied in design as it represents the single-domain performance level integrated by the normalised objective value within the component objectives in one domain.

A 3d scatter plot is used to represent the integrated performances of the feasible Pareto front from the global optimisation. The three axes are NDO of acoustic, sunlight and thermal. All three NDO are calculated based on a normalised value. The lower NDO performs best in all three domains. The scatter plot of the Pareto front consists of the feasible Pareto solutions as shown in figure 9.2. The original samples are also plotted according to their scores in each performance domain as a comparison.

As shown on the left of figure 9.2, the Pareto solutions gathered in the form of a leaned surface, with more concentrated solutions plotted on the edge of the surface, especially on the top and left edges.

Because it is expected that the lower the NDO value, the better it will perform solutions in all three domains, this is qualitatively represented on the 3d scatter plot:

The top left corner of the Pareto front indicates the best acoustic and the less optimised sunlight and thermal performances; the top right corner indicates best sunlight performance with compromises in thermal and acoustic aspects; the down left corner shows best thermal and less optimised sunlight and acoustic performances; and the down right corner shows best sunlight and thermal performances but less optimised acoustic performance. The middle sections on edges or inside the surface indicate the solutions of balance but mediocre performances for all three domains.
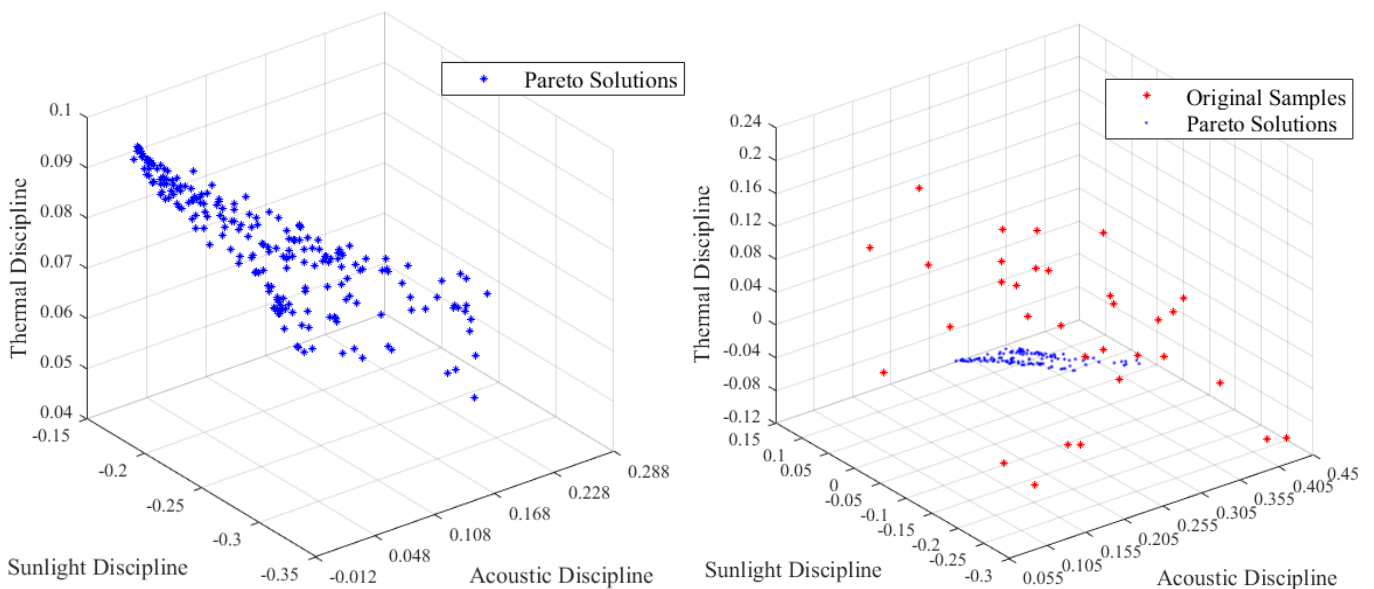
Figure 9.2 Integrated Performances of Feasible Pareto Front from Global Optimisation (left) versus Original Samples (right) in normalised scale of -1 to 1

As seen on the plot, variation trends among the interactive three-domain performances could suggest that:

- Improving acoustic performance has an inverse relationship to individually or simultaneously improving sunlight and thermal performance.

- Improving sunlight performance has an uncertain relationship to improving thermal performance, namely in terms of the concordant or inverse.

In figure 9.3, the normalised domain objective (NDO) values of 232 feasible Pareto

solutions are plotted against 43 original samples for three domains. As can be seen, the value fluctuations of feasible Pareto solutions are limited to a relatively small range and the average value is lower compared to those of the original samples. Because of the trade-off for the inverse requirements in the optimising multiple objectives from the three domains, the average values of NDO of feasible solutions reduces in a smaller amount than that of the samples. Also, the minimum value of NDO which refers to the sum of the extreme values of the domain performance indices is barely approachable for any domain in global optimisation, as shown in table 9.1:

Table 9.1 Range of NDO for Feasible Pareto Solutions and Original Samples

NDO		Acoustic	Sunlight	Thermal
Feasible Solutions	Max	0.249645	-0.16275	0.103747
	Min	-0.01858	-0.33187	0.039043
Original Samples	max	0.441488	0.015151	0.233284
	min	-0.11865	-0.45427	-0.11121

The fluctuation ranges of NDO for feasible solutions in acoustic, sunlight and thermal domains account for 13.4%, 8.5% and 3.2% in the full NDO ranges of each. Similarly, the corresponding fluctuation ranges for original samples account for 28.0%, 23.5% and 17.2%. Hence, for feasible solutions, global performance fluctuation reduces by 52.1%, 63.8%, and 81.4%, respectively, in three domains when compared with original samples.

There is a sharp contrast of the integrated performances between the feasible Pareto solutions and original samples as presented on the right of figure 9.2. The NDO values of feasible Pareto solutions are gathered at a relatively low level in three domains, compared with the original samples noted in red. The Pareto solutions in blue share obvious value decreases on three coordinate axes from the original location of the sample points. Although not reaching the minimum value as in some of the original samples due to the compromises for a multi-domain optimisation, the locations of feasible solutions are below the majority of 43 samples as observed from three coordinate axes.
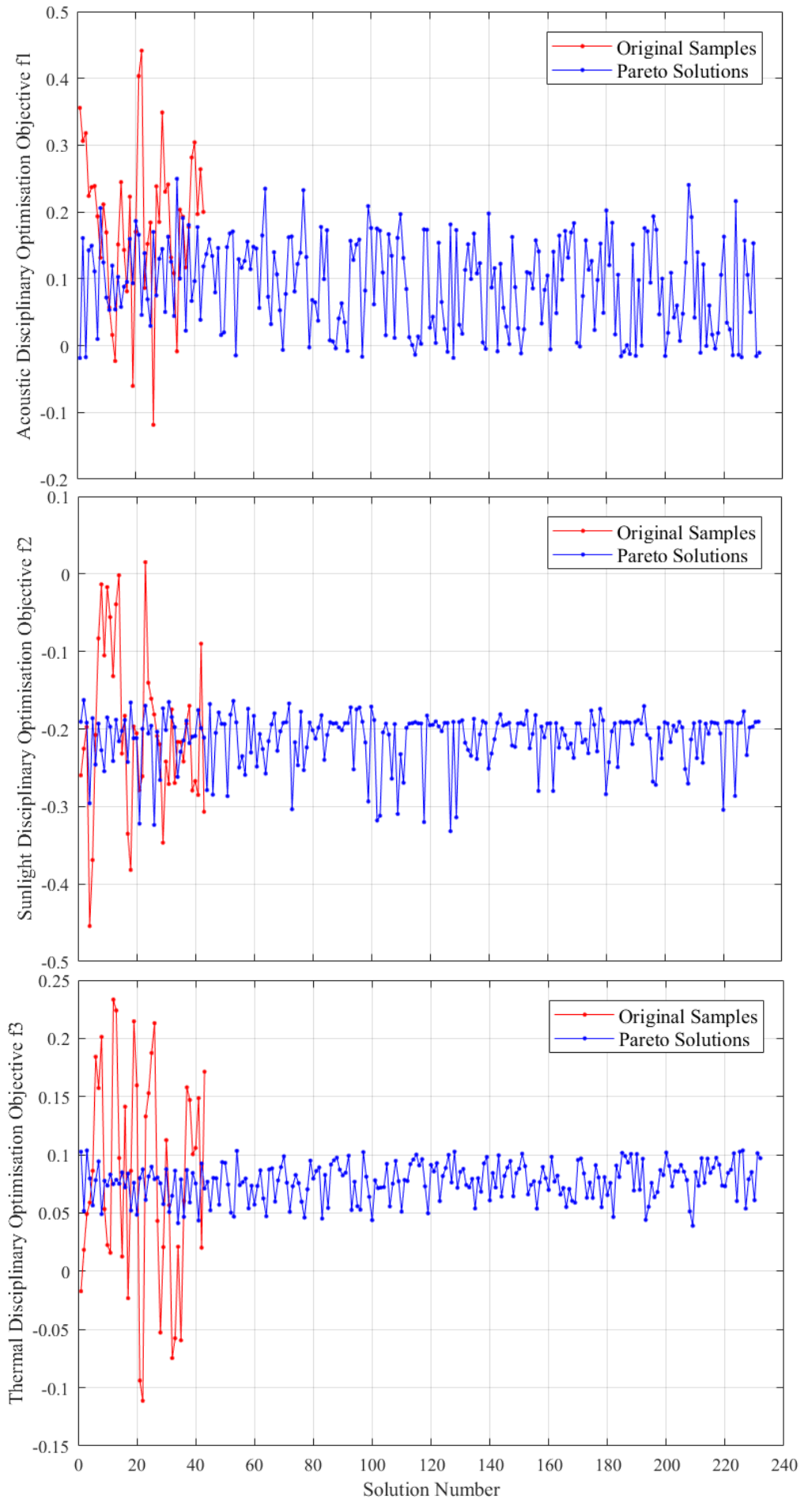
Figure 9.3
Domain
Performance
Improvements
(NDO value) of
Pareto Solutions
Compared to
Original Samples
in Acoustic (up),
Sunlight (middle),
Thermal (down)

9.2.2 Application of Global Optimisation and Normalised Domain Objective (NDO) of Feasible Pareto Solutions

According to the aforementioned analysis, normalised domain objective (NDO) has the descriptive ability to assess the single-domain performance of a design scheme. The NDO ranges of three domains of the feasible Pareto solutions in table 9.1 could be generalised as evaluation criteria of domain performances. If the NDO value of a scheme falls within the ranges, the corresponding domain performance could be ensured with the other two coupling domains considered. Furthermore, a combination of NDO of various domains is capable of describing the optimisation preference of the design scheme. Hence, NDO could be adopted as a novel integrated evaluative parameter for residential ward design results, regarding single-domain performance levels with respect to other coupling domains, and domain balance of a scheme. It could be applied during the selection stage of the optimised design solutions or when assessing the procedure of a novel design scheme.

The implications of lower NDO values in this research constitute an even and quiet acoustic environment, even and long accumulative sunlight hours and high wind speed matching low temperature and radiation. The suggested ranges of NDO for global optimised scheme are [-0.018, 0.250], [-0.332, -0.163] and [0.039, 0.104], respectively for acoustic, sunlight and thermal domains. It is possible to involve other environmental performance domains in future work to be assessed with NDO.

By appling the meta-model and 3d NDO scatter plot in guiding a design, multiple tasks could be achieved for a new and existing design:

- Input a design scheme for fast results within minutes for global and domain performance levels as well as acquiring respondent performance indices values to judge against design standards in the early design stage.
- Plot the point against the feasible Pareto front to check performance balance among domains.

- The positions of several input design scheme could be used to analyse integrated performance improvements.

The generalisation level of NDO requires further validation with sufficient samples. Abundant training samples could improve the predictive accuracy of three meta-models, in other words, accurately predict domain performance level from specific building distribution pattern. Accurate meta-models would improve the feasible solution accuracy calculated from the global optimisation objective function. Therefore, the solidity of the suggested NDO ranges could be enhanced which would contribute to its generalisation.

To ensure a designer-friendly usage of the MD-MOO model, an interactive interface would be easier than Matlab manuscripts. Therefore, an interface is development, described in section 9.7, taking NDO as an evaluative variable and building distribution indices as design variables. By altering design variable combination, solutions are interactively plotted by three NDO values of acoustic, sunlight and thermal domains and are plotted over the reference feasible Pareto front. The tool allows dynamic adjustment of design variables to locate the global performance of the design scheme into the balance zone of good performance. In the interactive chart, the three coordinate axes are assigned the NDO of acoustic, sunlight and thermal domains. The Pareto front is plotted as four curved surface edges. If the test scheme is plotted close enough to the edges, the scheme is considered well performed with domain balance. The position of the test scheme also indicates the domain balance of the design and the possible improvement direction.

In a word, compared to the irregularity of the integrated performances of the original samples, feasible Pareto solutions show great global performances at equal level with various combinations of domain performances. The multi-domain multi-objective optimisation significantly improves the integrated environmental performances of the residential ward schemes from conversional design approach via systematically

searching novel design patterns. The surface of the feasible Pareto front indicates the diversity of feasible solutions. Furthermore, the gathering of feasible Pareto solutions compared to the samples on the surface indicates the effectiveness and efficiency of the MD-MOO.

9.3 Analysis over Feasible Pareto Solution Set ——Single-Domain Performance Improvement

The global optimisation enhances the domain performance at different levels. As shown in figure 9.3, the improvement in acoustic performance reaches its highest level, however it also has a wider fluctuation, while the thermal performance is improved to a very tight range. At the same time, the minimum NDO cannot approach a low value. This section discusses the improvement trend in three domains.
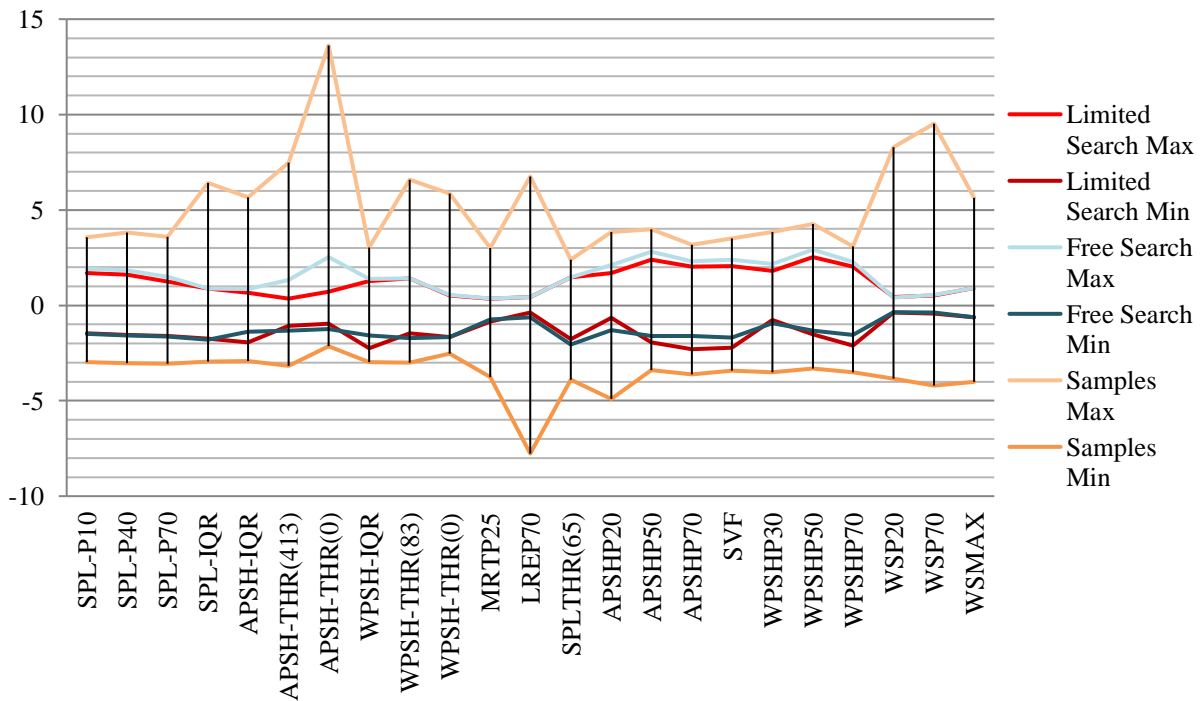

Figure 9.4 Standardised Ranges of Domain Sub-Objectives of Feasible Solutions, All Solutions and Original Samples

The possible ranges of 23 domain sub-objectives (OVs) are compared for unscreened Pareto solutions unlimitedly searched in a physical range of variables, and feasible

Pareto solutions searched in a limited design space of variables. The variations of 23 sub-objectives for two groups of solutions are plotted against the possible range of original samples. The scale of the three groups of values are standardised before the comparison. Figure 9.4 indicates the range of each sub-objective in three domains for two Pareto sets and sample. The outer pair of polylines is the boundary of the samples' ranges, the inner belts in red and blue are range of feasible solutions and free-search Pareto solutions. From SPL-L10 to LRE-P70, the value expectation is low for good performance while from SPL-THR(65) to WS-Max the expectation is high.

As seen in figure 9.4, the left sections of the ranges of the two Pareto solution sets has the clear tendency of approaching the lower boundary of the samples' range. The right sections of the ranges of two Pareto solution sets float close to the upper boundary of the samples' range, especially the right sections of the maximum value lines. This suggests that the global optimisation results improve overall performance through the increase or reduction of the corresponding sub-objective values.

The data variation in each sub-objective of the feasible Pareto solutions from a constrained search is almost overlapping the range of unconstrainted-searched Pareto solutions. This means that even the limited searching space of the feasible solutions is smaller than that of the unconstrained-search solutions. The searching efficiency and optimisation effects in each sub-objective are still approximately equal to the result of the free search.

The variation of design variable values for feasible solutions, free-searched solutions and samples are plotted in figure 9.5. It could be suggested that the majority range of the three groups overlaps. Generally speaking, the range of samples reaches the possible physical boundaries of all design variables except for sound source level presented by SPL-L10. The samples' ranges represent the design space adopted in conventional design habit. The range of feasible solutions is almost fully overlapping the range of the samples, which means the space-limited optimisation has fully

searched the possible practical design space. Even so, for free-searched solutions, the SPL-L10 significantly exceeds the lower limit of the samples' range and the possible range in practise. This means that while the free-searched solutions could reach balanced integrated performance, they nevertheless may contribute by building distribution combination which is out of practical possibility. This is the reason for applying space-limited optimisation to improve searching efficiency in both possible physical boundary and practical design space.
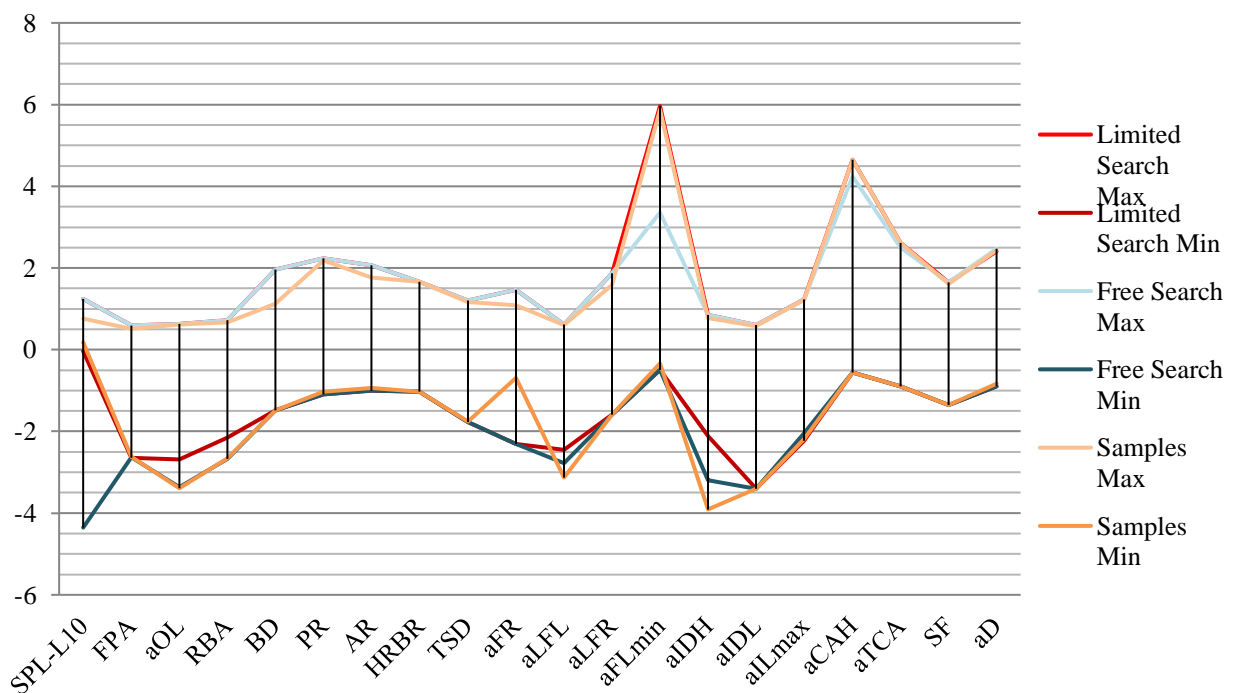
Figure 9.5 Standardised Ranges of Design Variables of
Feasible Solutions, All Solutions and Original Samples

In conclusion, the feasible Pareto solutions collected from limited-space searching could reach identical or even better performances in each of the domain sub-objectives than free-search solutions, with the guarantee of the feasibility of the design variable combination. Furthermore, the allowed design space for performance-balanced feasible solutions is not reduced from that of the conventional design habit. An organisation of the value combination of the same variable and same range could form the suggested design scheme and achieve a domain-balanced and performance-satisfying environment.

9.4 Suggested Residential Ward Distribution Based on Feasible Pareto Solutions

In the global optimisation with NSGAII algorithm, 232 feasible optimised Pareto solutions are selected under the constraints of design variables and objective variables unequalty constraints. After an interpretation of the feasible solution set, the specification of various design schemes within the solution set and an explanation of their distinctions are required.

Therefore, a clustering analysis of feasible solutions is operated based on the characteristics of the value combination of 20 design variables. Three groups of design schemes are revealed. The typical schemes are selected and discussed regarding their building distribution patterns, optimised performance patterns and interactions between the two. The data from the original samples is also incorporated into the mixed cluster of the feasible solution data. This allows the differentiation of well performed and balanced original samples from the more mediocre samples and locating them into the suggested design pattern system. The comparison with outstanding samples is also a cross validation for the feasible solutions.

9.4.1 Cluster Analysis of Feasible Pareto Solutions with Reference to Original Samples

To understand the data distribution and easy application to the practice, the categories of the feasible Pareto solutions are studied. Cluster analysis is applied according to average linkage between groups over all 232 feasible Pareto solutions plus 43 original samples. Combining the feasible Pareto solutions and original samples will allow observation of the similarities and location of the samples within the optimised results.

The cluster is based on the design variable values of the solutions and the samples, since the variation of the design variable of feasible solutions not only covers the full possible physical range but also clearly varies in the form of value combinations.

Attempts of clusters based on objective variable values or performance indices are also carried out. The result does not indicate significant groups as the optimisation solutions are generally even distributed. The feasible solutions are of non-dominated relationships and their environmental performances continuously change in the optimised range. Hence, the discrepancy of domain sub-objectives among feasible solutions is not strong enough to divide them into clusters.

The adopted cluster is based on linkage of design variables similarities in building distribution parameters. 232 feasible Pareto solutions and 43 original samples cluster into 4 significant groups, seen in appendix H. Three main groups are extracted for optimised solutions, and a few sample cases also located in these groups due to similar building distribution pattern. The details about the location of sample cases in optimised schemes will be discussed in following sections of design schemes. The 4th group are fully consisted with sample cases. The sub-group division is decided with consideration of both spider charts of DVs and OVs. The shared physical meaning of each pattern of building distribution and characteristics of performances are used as division reasoning.

As seen in the dendrogram in appendix H, for the feasible Pareto solutions, three main tree branches are clustered. The red, blue and green sections display the part of cluster A, B and C. The typical cases in each cluster with clear meaning in their DV and OV spider diagrams are marked with coloured dots, and list their spider diagram of the right of the dendrogram.

9.4.2   Spider Diagrams of Performance Balance and Distribution Arrangement

Following the selection of suggested design schemes, the performance balances of multiple environmental objectives and the distribution arrangements of various suggested schemes are analysed. The data of feasible solutions and samples is normalised to range of [-1, 1], so that the performance balance of 23 variables and the distribution arrangement of 20 indices could be plotted by two spider diagrams for comparison in figure 9.6. Figure 9.6 (a) indicates the spider diagrams of normalised performance balance, with description of definition and categorisation on the edge rings. The performance balance butterfly chart of typical cases in three design schemes is plotted at centre. Feasible solution NO.218, NO.26 and NO.100 (noted as F218, F26 and F100) is used as typical cases for scheme A, B and C, respectively. Similarly, figure 9.6 (b) presents the spider diagrams of normalised building distribution balance of suggested schemes in three design patterns A, B and C.

For figure 9.6 (a), the upper spider chart indicates the performance balance of three suggested schemes. The top 180 degrees, namely zones 1, 2, 3 and 4, represent performance variables requiring low level for satisfying performances. The bottom 180 degrees or zones 1', 2', 3' and 4', represent variables requiring high level for satisfying performances. Hence, large sector areas in the zones of the bottom semicircle and small sector areas in the top semicircle are the preferred performance diagram pattern of optimised schemes.

As denoted in figure 9.6 (a), light and dark yellow represent negative and positive factors for acoustic performance; light and dark red are assigned, respectively, for negative and positive factors of annual sunlight performance. The pairs of blue and green are assigned, correspondingly, for winter sunlight and thermal performances.

Zone 1 of the top semicircle indicates a negative acoustic performance, including high noise pressure level and noise unevenness. Zone 2 indicates the negative performance of annual accumulative sunlight hours, including sunlight unevenness and the violation of critical value. Zone 3 represents the negative performance in winter of

accumulative sunlight hours, including sunlight unevenness and violation of critical value. Zone 4 shows the negative performance of thermal comfort, including overheating and over-radiation.

Referring to the bottom semicircle, the positive performance of the certain domain appears on the opposite side of the corresponding negative performance. The positive acoustic performance is shown in Zone 1', referring to the fulfilment of critical value. Positive year-round accumulative sunlight performance shown in zone 2' refers to long accumulative sunlight hours and a large sky view. Positive winter accumulative sunlight performance shown in zone 3' contains long accumulative sunlight hours. At last, the positive thermal comfort performance is presented in zone 4' including high wind speed.

As shown in the above performance balance charts, the feasible Pareto solutions share a butterfly pattern in the diagram. Therefore the integrated optimised performance with domain balance could be named as butterfly performance. A practical, balanced and performance-satisfying scheme should have larger hind wings, and small fore wings and short antennae.

The spider diagram of the building distribution arrangement reveals the value combination of design variables or building distribution indices. As coloured in the lower diagram in figure 9.6 (b), in total there are 8 divided zones indicating 8 key distributional attributes. Zone 1 and Zone 2 in the top right section of the circle crowd the design variables about site scale and building amount, respectively. Zone 3 presents degree of building density. Zone 4 refers to variables regarding the existence of high-rise buildings. Zone 5 refers to variables of facade protection, namely degree of enclosure at site boundary which is a strong protection at the facade. Zone 6 refers to the protection of the inner site from the cluster structure of buildings, which is a weak protection at the facade. Zone 7 is the sector referring to variables of facade opening encouragement. Zone 8 represents the variables indicating existence and size

of edge open spaces. Finally, the sound source level is also shown in the spider
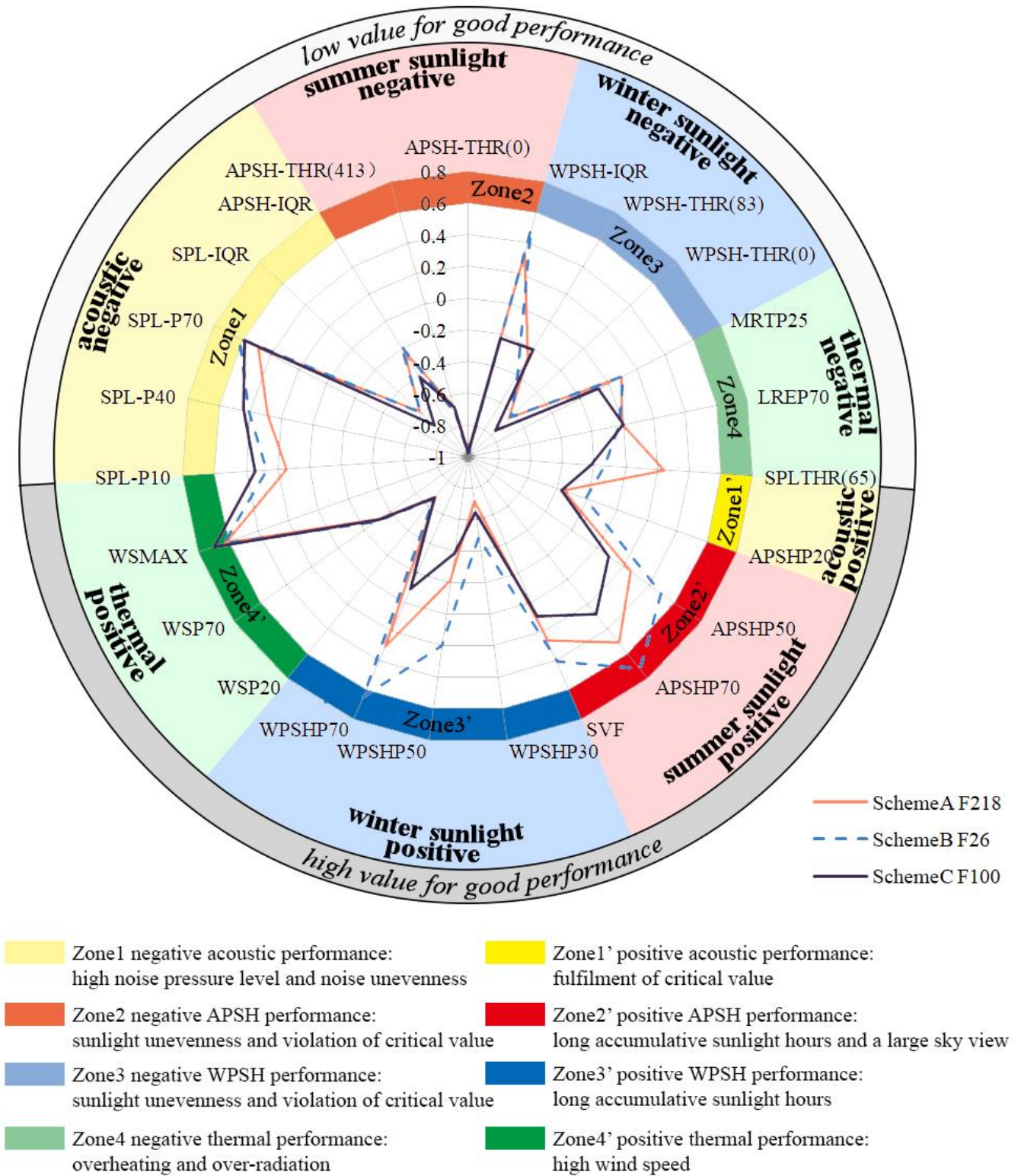

Figure 9.6 (a) Spider Diagrams of Normalised Performance Balance
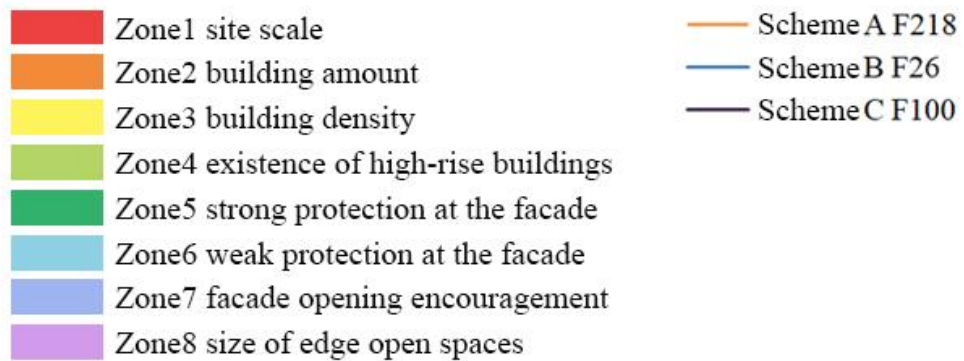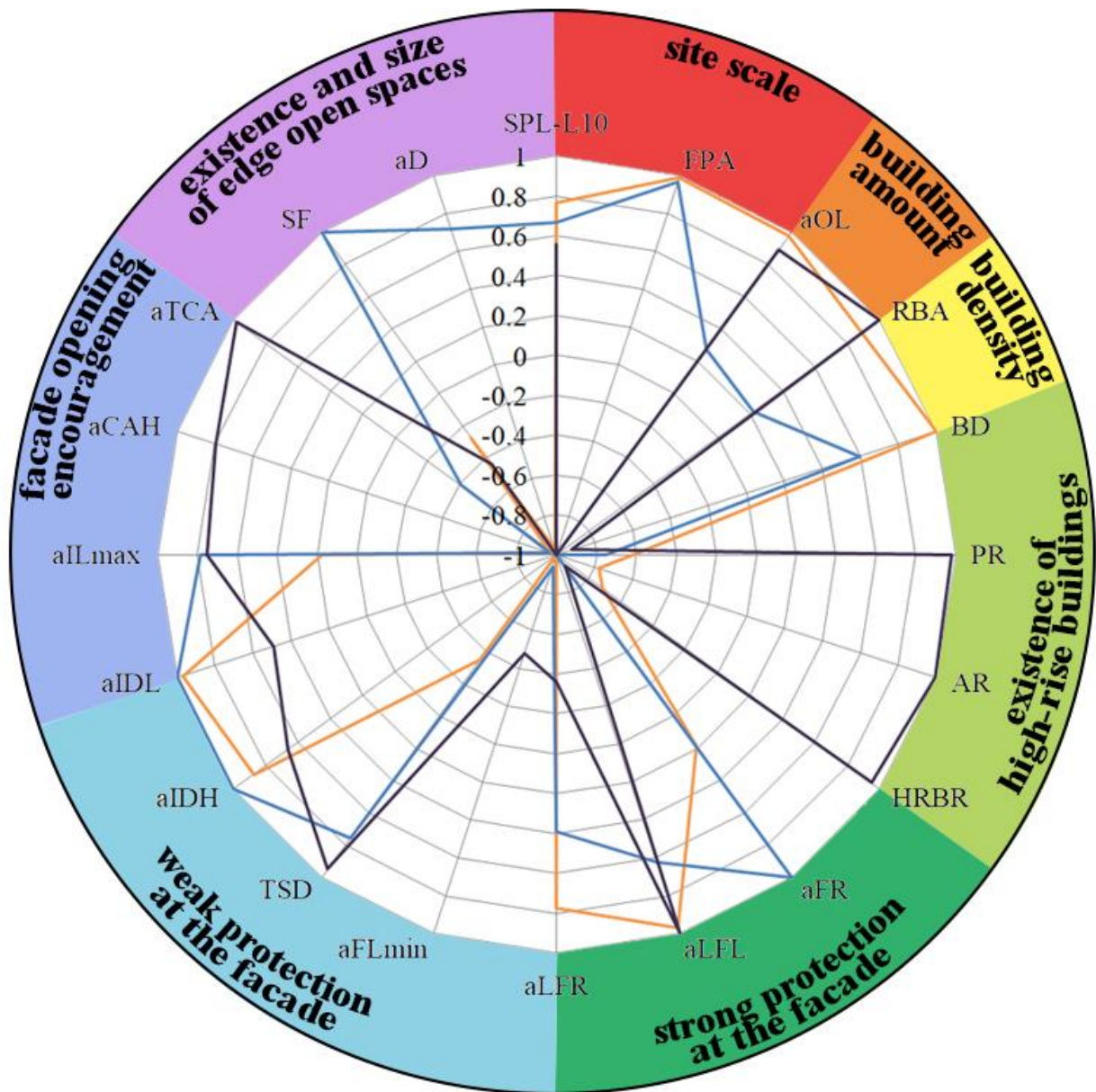of Suggested Schemes in Three Design Patterns A, B and C

Figure 9.6 (b) Spider Diagrams of Normalised Building Distribution Balance
of Suggested Schemes in Three Design Patterns A, B and C

diagram, but is not summarised into a zone. While it is a key indicator of global performance, it is not alterable by building distribution.

According to the characteristics of the design variable, three groups of solutions are divided, and each shares a typical diagram pattern. As shown in the performance balance diagram of the three schemes, the three diagram patterns all resemble a butterfly. Scheme A possesses advantages in acoustic performance, scheme B in sunlight and scheme C in thermal. Judging from the diagram patterns of the three schemes, A is the most domain balanced solution. Correspondingly, the three schemes appear as a three-sector pattern in the distribution arrangement diagram.

The data of the 232 feasible Pareto solutions are consolidated according to the three scheme division. The suggested value range of design variables and performance variables are listed in table 9.2.

In the left part of table 9.2, the range of design variables for three schemes, the sectors with high value level as seen in the spider diagram of design variables are coloured respectively for scheme A, B and C in red, blue and green. While the design variable with lowest value level among the three schemes are marked in bold. Seen from the table 9.2, the scheme A, B and C respectively has 3 sectors, 4 sectors and 3 sectors as various locations of the design variable groups.

In the right part of table 9.2, the range of performance variables for three schemes, the sectors represents the best performances is coloured for each scheme. While the lower and upper bound which represents worst performance for certain domain, are markd in bold. Since the three schemes are all balanced in global performance, seen from the table, scheme A has outstanding performance in acoustic domain; scheme B is better in sunlight performance; and scheme C has advantage in thermal performance.

Table 9.2 Suggested Range for Schemes in Design Pattern A, B and C

Design Variables	A		B		C		Performance Variables	A		B		C	
	Min	Max	Min	Max	Min	Max		Min	Max	Min	Max	Min	Max
FPA	8.88	9.00	5.58	9.00	0.40	9.00	SPLP10	50.06	57.45	52.12	59.46	56.89	62.38
aOL	557.69	606.00	436.71	606.00	222.28	606.00	SPLP40	56.66	62.90	58.75	64.67	62.30	66.81
RBA	59.72	63.00	23.41	63.00	11.35	63.00	SPLP70	65.58	70.50	67.85	73.48	70.00	73.61
BD	12.19	40.00	7.00	40.00	7.00	20.89	SPLTHR65	44.46	66.98	37.68	59.67	27.37	46.81
PR	0.60	3.16	0.60	3.08	0.82	3.60	SPLIQR	11.91	14.68	11.38	14.51	9.98	12.49
AR	0.50	4.00	0.50	3.43	0.89	4.00	APSHP20	509.59	542.27	533.11	657.78	508.43	625.30
HRBR	0.00	0.92	0.00	1.00	0.08	1.00	APSHP50	929.39	1056.54	968.68	1245.65	887.69	1153.58
aFR	25.92	100.00	5.71	100.00	0.00	55.31	APSHP70	1110.04	1268.72	1143.42	1429.52	1043.82	1349.83
aLFL	207.60	219.00	39.70	219.00	45.95	219.00	APSHIQR	544.82	649.51	544.24	675.14	482.74	649.04
aLFR	0.00	100.00	0.00	100.00	0.00	31.87	APSHTHR413）	15.97	16.54	13.83	16.21	13.83	16.40
aFLmin	12.00	47.04	12.00	64.62	12.46	82.00	APSHTHR0	1.13	1.35	0.64	1.21	0.95	1.56
TSD	60.00	1020.16	365.40	1200.00	872.35	1200.00	SVF	57.39	63.92	58.52	70.94	54.55	67.36
aIDH	46.20	56.00	39.17	56.00	28.46	56.00	WPSHP30	47.90	55.94	50.66	92.45	56.14	85.35
aIDL	32.56	36.00	29.95	36.00	10.00	36.00	WPSHP50	116.07	144.83	124.48	211.70	104.37	181.72
aILmax	18.00	86.00	51.25	86.00	48.45	86.00	WPSHP70	180.66	233.18	186.27	290.77	156.45	261.59
aCAH	0.00	437.68	0.00	1773.66	0.00	5837.00	WPSHIQR	167.72	235.72	166.09	260.25	135.42	230.50
aTCA	0.00	17943.46	0.00	25096.46	8754.03	26073.00	WPSHTHR83	35.78	39.36	29.49	36.24	29.48	40.63
SF	1.50	3.78	2.54	4.50	1.50	4.48	WPSHTHR0	14.12	17.90	14.46	18.53	11.84	16.83
aD	13.00	27.58	14.69	57.20	13.00	60.14	MRTP25	319.15	325.05	320.19	324.55	316.96	322.79
SPLL10	70.00	90.00	71.99	87.71	70.00	85.84	WSP20	0.67	0.69	0.69	0.73	0.69	0.75
							WSP70	1.41	1.45	1.44	1.50	1.44	1.52
							WSMAX	3.37	3.50	3.40	3.51	3.47	3.57
							LREP70	187.68	190.73	184.78	188.89	185.20	190.92

9.4.3   Suggested Design Pattern A

Based on the clustering analysis of feasible Pareto solutions as shown in appendix H, scheme A is summarised with a similar building distribution tendency. Distinguishing the detailed discrepancies in design pattern A, 5 design templates are extracted. One original sample also grouped within cluster A for similar design pattern would be compared with the 5 design templates.

The building distribution characteristics of scheme A are expressed in the manner of the combination of the design variables in the spider diagram in figure 9.6(b) and red section in appendix H. Three sectors of fan-shaped areas are formed by 5 design sections with high value level: sector1-1) site scale, 2) building amount and density; sector2-3) strong facade protection; sector3-4) part of weak facade protection and 5) sound source.

These design sections include design variables of sector1: 1) footprint area (FPA) and average outline length of site edge (aOL); 2) residential building area (RBA) and building density (BD); sector2: 3) average facade ratio (aFR), average low-rise facade length (aLFL) and average low-rise facade ratio (aLFR); sector3: 4) average higher interval depth (aIDH) and average lower interval depth (aIDL); and supplemented 5) sound source level represented by L10 of sound pressure level in site (SPL-L10).

These characteristics are also marked in red in table 9.2 which represents design variables of a high level for scheme A. Meanwhile, the smallest value for the lower and upper bound of the design variables is marked in bold. It could be suggested that among the three schemes, scheme A has the smallest lower bound in the design section of 1) existence of high-rise buildings, including variables of plot ratio (PR), aspect ratio (AR) and high-rise building ratio (HRBR); 2) opening encouragement, including average higher corner opening area (aCAH) and average total corner opening area (aTCA); 3) edge open space, including site shape factor (SF) and average distance to road (aD). Except for these three design sections, scheme A also has the lowest level of distribution clusters, presented by the triangle area standard deviation (TSD) and for average minimum facade length (aFLmin).

Shown in figure 9.6(a), the performance evaluative variable pattern of scheme A appears as a butterfly, with advantages in acoustic aspects compared to scheme B and C. In scheme A, SPL levels are lower and SPL-THR(65) is higher. The sunlight performance of scheme A is between that of schemes B and C. Thermal performance of the three schemes is nearly identical. As mentioned in the design strategies concluded in Chapter 4 referring to good acoustic performance, small site and large distance to road should be avoided but high density, boundary enclosure and long facade length are highly encouraged. Therefore, scheme A is capable of achieving an overall good performance in acoustic domain because it has a large site scale which avoids small site, high building density which limited separation distance between buildings, high facade protection and cluster protection which highly encourages boundary enclosure. Lack of edge open space also helps scheme A to achieve a higher ratio of critical fulfilment ratio, whereas scarifies the acoustic performance on street facing facades, namely further acoustic attenuation installation is required at those windows which are impacted.

Looking at the range of feasible Pareto solutions in design scheme A, as shown in table 9.2, it appears that scheme A has the smallest value for the upper bound of accumulative year-round sunlight hours, and in winter and sky view factor (SVF), as well as lower and upper bounds of critical violation ratio year-round. Reaching a higher maximum level of year-round sunlight hours is limited in consequence of a higher building density and a lack of large open space. Similarly, scheme C is limited in reaching a higher minimum level of year-round and winter accumulative sunlight hours due to a high plot ratio. Scheme A is also limited in the level of lower and upper bounds of wind speed due to a high level of facade protection and low level of opening and open space.

An original sample, S22, exhibits a similar design pattern and performance results as templates in scheme A. It is clustered into scheme A in the cluster analysis based on design variable characteristics of all optimised solutions plus original samples as shown in appendix H. As shown in the spider diagrams in appendix H and figure 9.7 marked by a yellowed dotted line, the design variables of S22 share the same three sectors of high level in the diagram as the rest scheme A cases. But the site scale, building amount and facade protection have relatively lower values of the

corresponding variables. Therefore, the other cases each from one sub-group in scheme A outperform the S22 in multi-domain performance balance. The sub-objective performances of S22 appear better in acoustic aspect with lower SPL values and higher ratio of critical value fulfilment. It is grouped in good performance in acoustic qualitative analysis in chapter 4. The performance in thermal aspect of S22 is limited by slightly higher mean radiant temperature and longwave radiation. It is grouped as fair in thermal qualitative analysis in chapter 6. The sunlight performance of S22 is generally similar to the other case in scheme A. It is rated as fair and poor in annual and winter sunlight qualitative analysis in chapter 5 due to small site scale and high building density. The building distribution and domain performance figures are presented in section 9.5. Since the cluster of solutions plus original cases are based on building distribution parameters, and the S22 is the last one adding to the cluster according to group linkage distance, S22 share the key distribution pattern as in scheme A but has discrepancy which results in its performance weakness. It is used as a comparison and a eaay graphic illustration of the possible building distribution pattern.

Based on the tree structure of the dendrogram from cluster analysis of optimised solutions in appendix H, sub-groups are clearly shown. With consideration of the spider diagram patterns of building distribution and corresponding performance, 5 sub-groups in scheme A is divided. Representative cases are selected from each sub-groups according to the significant physical meaning of the building distribution, for example feasible Pareto solution NO. 218 noted as F218 is adopted as template of scheme A1. Comparing the 5 templates, except for the shared part, A1 has highest building density, while A3 has the highest plot ratio. Corresponding to the value boosting in building density for A1, A1 has a lower level of facade protection, but is more tolerable to opening encouragement. It mirrors the design strategies concluded in section 4.6.2, that higher building density could compensate for less facade protections. Similarly, corresponding with higher plot ratio, A3 has a higher level of cluster protection and facade protection. It echoes with the rule of high plot ratio case relies on clusterness and facade enclosure to increase inner protection. A2 is a compromise condition between A1 and A3 - its high BD and PR is diluted by the large scale of site - so on diagram A2, it shows only a sharp increase in site scale and building amount. Correspondingly, A2 shares the characteristics of A1 and A3 in

bottom and bottom left blocks in block area. A4 and A5 are more balanced template schemes sharing fewer extreme values of BD and PR. They also have less sector area at sections of facade protection and cluster protection.
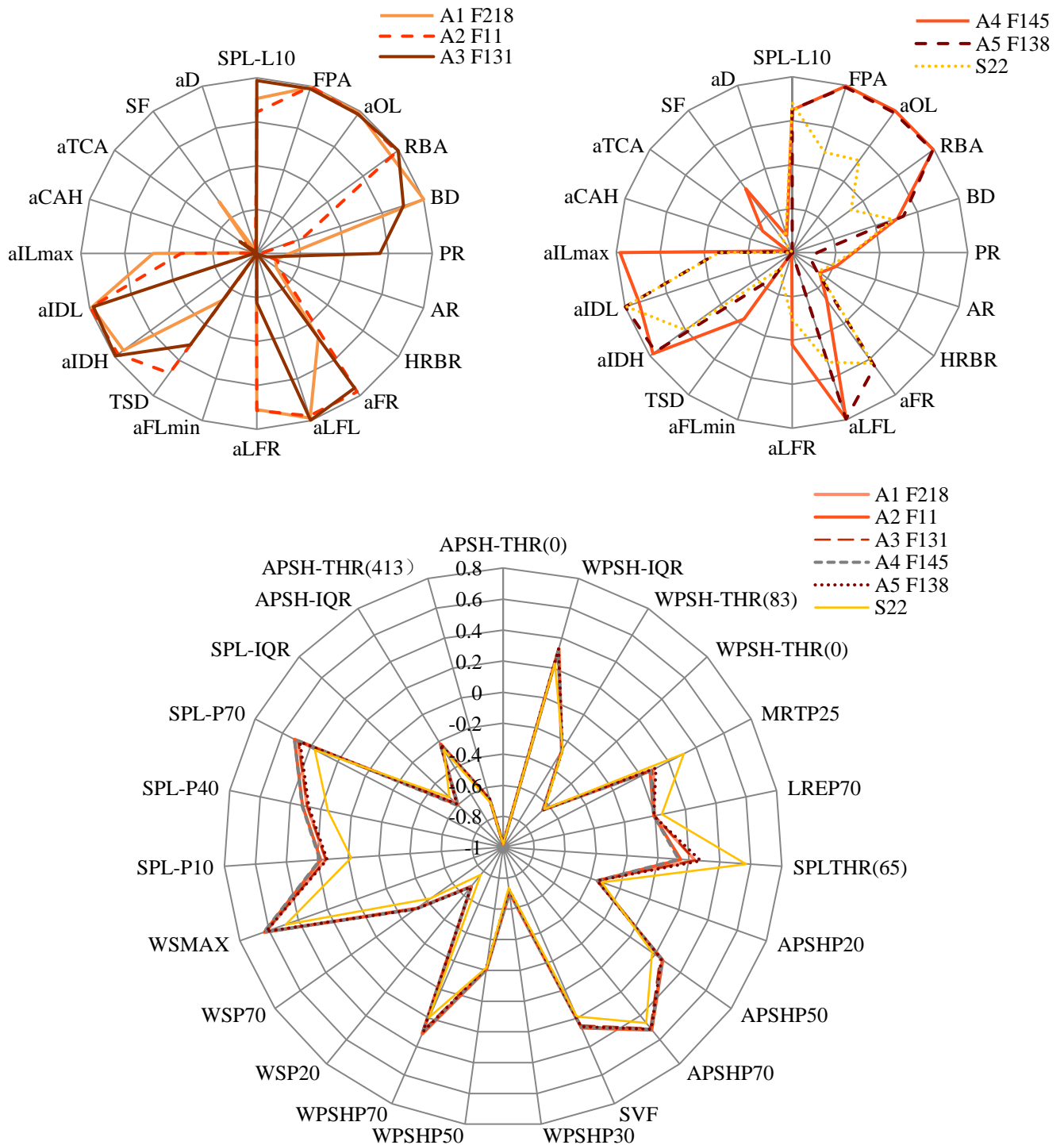

Figure 9.7 Design Scheme A and Corresponding Integrated Performances

Therefore, during the application of the templates in scheme A, if there are special high-level requirements on building density, clusterness, plot ratio, and interval length or balanced valuing, select templates A1 to A5 and adjust the distribution parameters arrangement from the presetting of the templates. If aware of the building distribution characteristics, a selection of the A1 to A5 template could provide a reference of design variable value correspondingly, to acquire a basically balanced performance in early design stage. The adjustment of design scheme could be based on the first orientation of the distribution arrangement and performance expectation. In the following stage, the amount of the adjusted parameter arrangement could be tested for global performance of three integrated domains in the interactive tool in section 9.7.

To summarise, design scheme A could qualitatively be described as embodying a large site scale of considerable building amount with medium building density, and high level in facade protections and cluster protection.

This results in a balanced and satisfying integrated performance with advantages in acoustic aspect, average good performance in sunlight and thermal aspect. For further enhancement in sunlight and thermal performance, introduce in-site or on-edge open spaces and encourage opening, high plot ratio and low building density, respectively. The suggested range of design variable in pattern A is listed in table 9.8 together with corresponding performance level as a design reference.

9.4.4   Suggested Design Pattern B:

The second cluster of solutions is of design scheme B. There are 6 templates in pattern B extracted from the dendrogram and spider diagram shown in appendix H. The butterfly diagrams of performance variables and design indices are presented in figure 9.8.

Generally speaking, the arrangement of the design indices of scheme B appears as 1) large site scale, 2) medium-high building amount, 3) medium building density, 4) medium level of facade protection, 5) high level of cluster protection, 6) medium level in opening encouragement, 7) large edge open space, and 8) medium-high sound source. The involved indices in each sector accordingly are 1) FPA, aOL, 2) RBA, 3)

BD, 4) aFR, Alfl, aLFR, 5) TSD, aIDH, aIDL, 6) aILmax, 7) SF, aD, and 8) SPL-L10. As indicated by the range of design variables in table 9.2, scheme B has a small value on section of 1) existence of high-rises, including variables of PR, AR and HRBR; 2) opening encouragement including aCAH and aTCA; and 3) minimum facade length (aFLmin).

Compared to design schemes A and C, scheme B has advantages in sunlight performance, with extended accumulative hours and a similar ratio of critical value violations. Referring to the qualitative design strategies concluded Chapter 5, the low/mid-rise community with medium density, namely low plot ratio with medium building density is the second-best arrangement for overall sunlight performance as shown in scheme B. It also suggests that a higher plot ratio tends to reduce fully sunlight blocked area, while increasing partly shaded areas. In other words, lower plot ratio would increase P20, P50 of APSH and P30, P50 of WPSH. Facade enclosure would increase the frequency of the lower tail of middle peak, which means it would slightly reduce P20, P50 of APSH and P30, P50 of WPSH. Clustered distribution with in-site open space is mentioned in the design strategy that is beneficial for overall WPSH. Edge open space in scheme B echoes to the design strategy that it is beneficial in boosting high end value (P70 and maximum) of APSH and WPSH. The relative direction of the edge open space to the site impacts the boosting power of the space.

However due to the largely improved value of P70 of APSH and WPSH, APSH-IQR and WPSH-IQR have slightly higher value than which of scheme A and C do. WPSH-IQR is especially high as high end of WPSH-P75 has been greatly improved than the low end of WPSH -P25. The less improvement on the low end of WPSH also appears on that WPSH-THR(0) of scheme B has highest range bound value among which of scheme A and C, namely the heavily sunlight blocked area accounts for a higher proportion of the whole site in scheme B. This is because of the combination of facade and cluster protection and the slightly higher plot ratio.

The thermal performance of scheme B is similar to A, but is less optimised than C. The lower bound of the MRT range of scheme B keeps the highest value. The reason behind this is the density of scheme B limiting the reduction of MRT and increasing of WS. The acoustic performance of scheme B is between A and C, because the

degrees of building density, facade protection, cluster protection, opening encouragement and edge open space are all in the middle-ground between A and C. Although the high level of edge open space reduces acoustic performance, it increases sunlight performance.

As displayed in diagram 9.8, which shows 6 templates of scheme B, in group 1, B2 and B3 share the common characteristics described above, and have increasing levels of building density. The edge open space reduces in order of B3, B1, B2, which is in accordance with the improvement in acoustic performance among the three. The reduction of average distance to road (aD) and average facade ratio (aFR) from B3, B1 to B2 and from B5, B6 to B4, also agrees with the improvement in sunlight performance.

The building distribution trend in B4 and B5 is slightly different from B1-B3. Except for the shared blocks in the diagram, B4 and B5 share significant signs of the existence of high-rise buildings: high PR, AR and HRBR. Existence of high-rises results in improvement in thermal performance - the reduction of mean radiant temperature (MRT) and longwave radiation from environment (LRE), compared to scheme A. B4 and B6 share special conditions in pattern B of extreme low edge open space, which results in a slight improvement in acoustic performance, while also reducing the sunlight performance.

Therefore, to conclude, for designs requiring optimisation weighted on sunlight performance, if there is a special requirement for varying levels of building density, facade protection and edge open space, select template B1-B3 in early early stage as a basis of improvement. If the requirement is for plot ratio and opening encouragement, select from templates B4-B6. Improvement could be texted in the interactive tool introduced in section 9.6, start from input the design variable value of the selected templates, then alter and test the integrated performance.

To summarise, design pattern B could be qualitatively described as representative of a large-scale site, considerable building amount, medium building density, high level in facade and cluster protections, medium opening encouragement, and large edge open space. This results in a balance of the integrated performance with obvious advantages

in sunlight aspect and average good performance in acoustic and thermal aspect. Extra enhancement in acoustic and thermal performance could be achieved by reducing edge open space and building density. The suggested range of design variable in pattern B is listed in table 9.2 together with corresponding performance level.
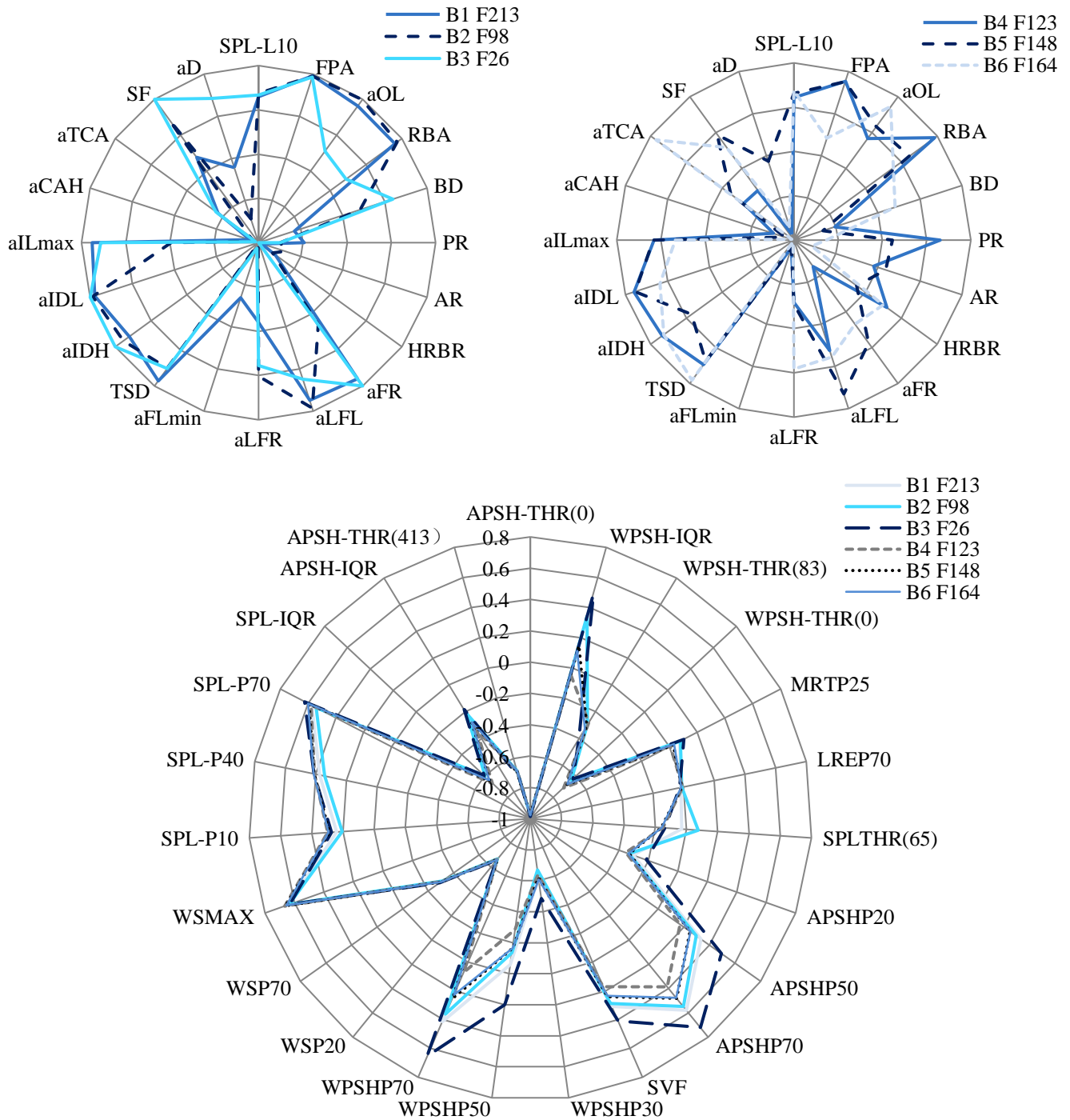

Figure 9.8 Design Scheme B and Corresponding Integrated Performances

9.4.5   Suggested Design Pattern C:

Cluster C is concluded from feasible Pareto solution analysis. The five templates are extracted from solutions in scheme C as seen in the dendrogram and spider diagram in appendix H. In addition, two original samples are also incorporated in cluster C in analysis for their similar design patterns. The spider diagrams of design pattern and corresponding integrated performances are plotted in figure 9.9.

It could be summarised that the design pattern of scheme C is of 1) a large amount of high-rise buildings, 2) medium level cluster protection, 3) high level opening encouragement, 4) mix-arrangement of edge open space. The involved building distribution indices are 1) PR, AR and HRBR; 2) aFLmin and TSD; 3) aILmax, aCAH and aTCA; 4) SF. The range of design variables are listed in table 9.2 of scheme C. It shows that among the three schemes, scheme C has the smallest lower bound value for section of site scale (FPA, aOL), building amount (RBA), and cluster protection (aIDH, aIDL), and the smallest upper bound value for building density (BD) and facade protection (aFR, aLFL).

The corresponding performances of scheme C also appear in a butterfly pattern of balanced and optimised solutions, as shown in schemes A and B. Pattern C has advantages in thermal performance, with mediocre acoustic and sunlight performance. Scheme C also performs best in APSH-IQR WPSH-IQR and SPL-IQR because of less optimised high end value of APSH, WPSH and SPL. Scheme C has a large sector area for 4 sections. The design indices involved for each sector have been discussed in the design strategies concluded in sections 7.4.7, 7.5.5 and 7.10. The optimised integrated performance of scheme C could be qualitatively explained by the design strategies.

High PR is preferable for overall mean radiant temperature, wind speed and longwave radiation from environment. A high level of clusterness is a negative factor for wind speed. Opening in a south-east direction boosts overall wind speed and building mass within 90 degrees of upper wind direction, leading to a fast decrease behind and in front of the building. Opening also improves lower end P30 of MRT. Referring to strategy of variables in the section of edge open space, narrow width in a north-south
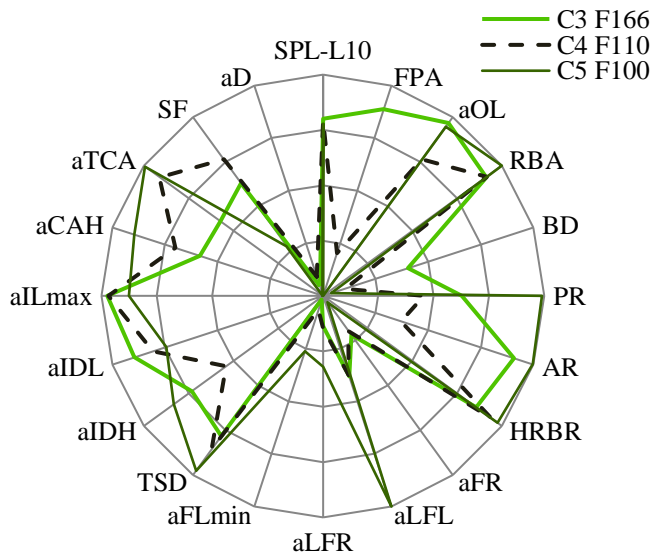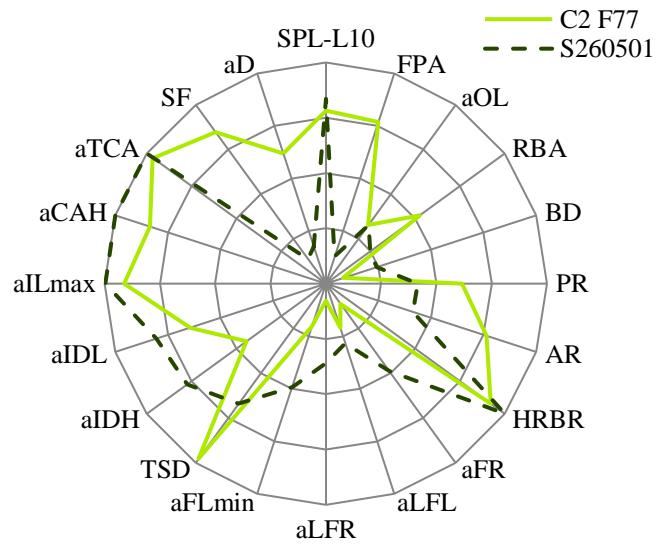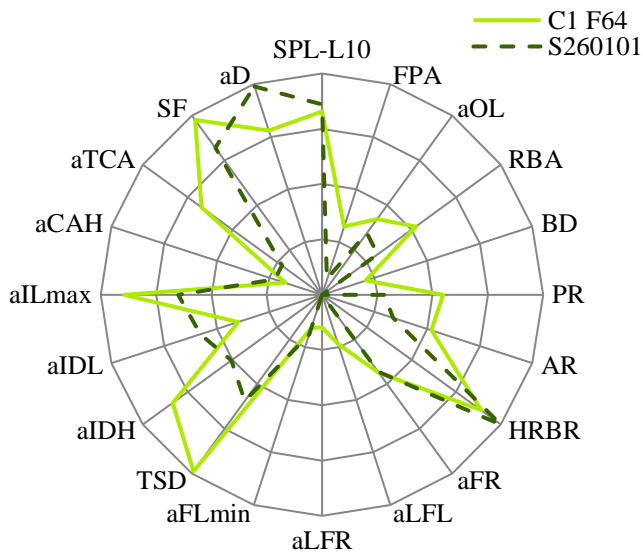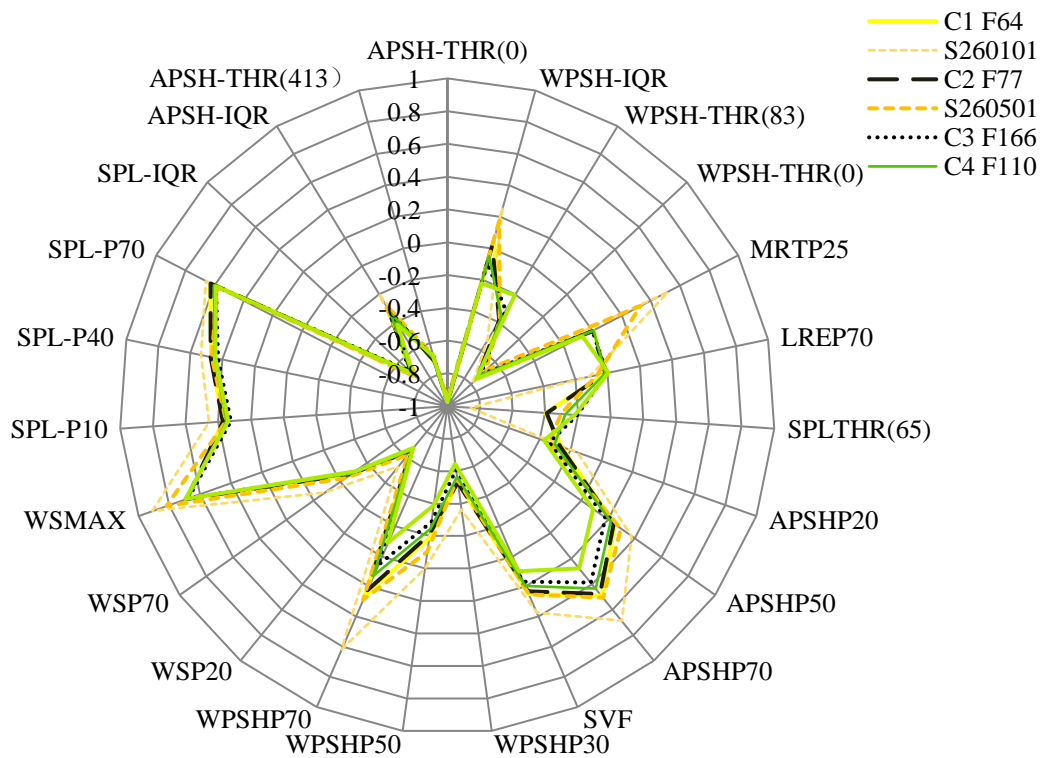
Figure 9.9 Design Pattern C and
Corresponding Integrated Performances

direction has a better MRT performance especially for high PR. It is suggested larger in-site open space helps with overall wind speed. In other words, edge open space would less effectively improve overall wind speed. With these exceptions, scheme C has the characteristics of a smaller site, less enclosure and less facade protection, all of which are beneficial for overall wind speed and mean radiant temperature, as mentioned in the design strategies of thermal comfort.

As indicated in table 9.2, scheme C has the least optimised lower and upper bounds for acoustic performance because of a high level of opening encouragement and edge open space. The lower bounds of APSH, WPSH and SVF are also shown as the smallest - scheme C may result in least optimised sunlight performance due to existence of high-rises.

There extract 5 templates in scheme C from the tree structure of the dedrogram and pattern of spider diagrams in appendix H. Template C1 has a significant boost of value in the amount of high-rise, as presented by the high-rise building ratio (HRBR) and the cluster protection (TSD, aIDH aIDL); and has large edge open space, as presented by long-narrow site shape (SF), large set-back distance (aD) and strong sound source (SPL-L10). A mixed arrangement of opening encouragement also appears in C1 of medium level in the total corner area (aTCA) and max interval length at the boundary (aILmax), but low level in high corner area (aCAH). The sample S260101 almost shares a similar design pattern as C1. They both share a balanced integrated performance as shown in the butterfly diagram of pattern C, while C1 outperforms S260101 in thermal aspect, expecially in MRT performance. Seen from the figure 9.9, the difference in distribution charateristcis between S260101 and C1, which is the low site scale and building density and less long facade protection, leeds to a high value in MRT in diagram of S260101. However, S260101 has better sunlight performance due to low plot ratio, compared to not only C1 but also to all other templates in pattern C.

Compared to the optimised solutions in scheme C, although the original sample S260101 and S2605001 share similar pattern of building distribution to the feasible Pareto solutions in scheme C, the detail discrepancies still cause less balance in global performance as seen in the butterfly chart of performance in figure 9.9. They are

compared and apllied as a media of direct graphical illustration of the possible distribution pattern.

Template C2 has a very close design pattern to C1, except for its high level of opening encouragement and low level in interval depth. The performances of C1 and C2 are almost identical. A sample S260501 shares the same design pattern as C2, it also presents a similar performance butterfly with advantages in sunlight aspect due to small plot ratio.

Template C3, C4 and C5 show similar trends in a high level of cluster protection and existence of high-rises, but appear to be additionally boosted in site scale and building amount. As shown in the butterfly diagram, sunlight performances reduce from C4, C3 to C5 for 10% at the most. This accords with the plot ratio increases from C4, C3 to C5 in design pattern diagram 3. Nevertheless, MRTP25 reduction is also related to the increase in plot ratio as shown in the butterfly diagram.

To conclude, in early design stage thermal performance is more preferred in balanced requirements, there are 5 templates to choosen from according to tendency on building distribution pattern. When selecting templates from scheme C suitable for high plot ratio design, if a higher cluster protection is preferred, refer to template C1; if higher opening encouragement is required, refer to C2; and if there is a particular requirement for a larger site scale and building amount, refer to C3, C4 and C5.

To summarise, for future designs which require a large number of high-rise buildings, the templates in scheme C are appropriate. The design pattern of scheme C could be qualitatively described as a strong presence of high-rise buildings, as a high level of cluster protection, opening encouragement and edge open space, and as tolerable to various site scales and building amounts. The advantage of performance is shown in thermal comfort while compromising on overall sunlight performance. To further enhance acoustic performance, the reduction of opening and open space combined with an increase of facade protection is needed. To improve sunlight performance, reduction of plot ratio is essential. The corresponding range of design variables and performances are listed in table 9.2 for design scheme C.

9.5 Validation of Suggested Design schemes

9.5.1   Validation of Suggested Schemes by Original Samples in Identical Solution
        Clusters

After the cluster analysis, three design schemes are extracted from the three clusters with several original samples in the identical cluster due to accordant design patterns and integrated performances. These original samples are utilised for the validation of practical feasibility and the performance effectiveness of the suggested schemes.

Sample S22 resembles design scheme A in figure 9.10. Its building distribution, simulation result maps of noise propagation, annual accumulative sunlight, winter accumulative sunlight, sky view factor, mean radiant temperature, wind speed and longwave radiation from environment are presented.

Templates in Scheme A most resemble S22 and S1801 in design pattern. A possible distribution of scheme A would have multiple smaller clusters with relatively high facade protection for street facing sides of the cluster, while wider separation distance exists between clusters. The slim high-rises mainly locate in the in-site and edge open spaces on the dominant wind direction.

The rhythmic cluster-facade protection and wide-deep interval opening not only ensures the acoustic performance by obstructing traffic noise propagation, but also allows higher speed wind flow to circulate in site. The corner opening with narrow-facade high-rises on the upper wind direction also introduces wind into the site. The edge open area facing the narrower street or smaller sound source allows proliferation in year-round and winter accumulative sunlight hours, with control on traffic noise propagation within a satisfactory level. Medium density matching medium plot ratio within clusters results in balances of sufficient annual accumulative sunlight hours and excessive mean radiant temperature due to overheating by radiation. The coupling configuration of density and plot ratio in clusters also ensures winter accumulative sunlight hours, sky view factor and longwave radiation from the environment remain within a satisfactory level.

Another possible distribution 2 of scheme A would appear as figure 9.11, taking maps of S1801 as a sample. Except for the shared distribution characteristics and performances, distribution 2 has a different form of cluster protection. The clusters are not highly enclosed or directly wrapped by low-rise buildings, nor do they have a U-shape building mass. The relatively long facade of medium-rise residential buildings in even distribution and medium density consists of the facade protection effect. It demonstrates another possibility of forming cluster and facade protection.

There is no original sample clustered into the same group with suggested design scheme B. However, matching the values of design variables of samples to ranges of scheme B, a few samples share characteristics of pivotal design variables which determine the main structure of the distribution with scheme B. For instance, sample with alteration on facade ratio, edge open space etc. could be adopted as validation cases of scheme B on its practical feasibility and performance effectiveness.

There are two samples in the same cluster as design scheme C, S260101 and S260501, which are utilised as validation cases for scheme C.

Figures 9.12 and 9.13 for possible patterns 1 and 2, show distribution characteristics of high plot ratio, cluster distribution, boundary opening encouragement and edge open space. Comparatively, pattern 1 accomplishes cluster protection relying on longer residential building facades, while pattern 2 relies on long facade together with the long wrapping of low-rises. Furthermore, pattern 1 is of a large site scale, but pattern 2 is small which is in accordance with the tolerance of the various site scale of scheme C.

Because facade enclosure is not a good accompaniment with high plot ratio scheme in the context of good integrated performance, the open distribution of scheme C consequently leads to a satisfactory acoustic performance. Nevertheless, it results in the best thermal performances in aspects of mean radiant temperature, wind speed and longwave radiation, as well as distinct APSH level and satisfying WPSH level. Furthermore the edge open space significantly increases the high end level of APSH and WPSH.
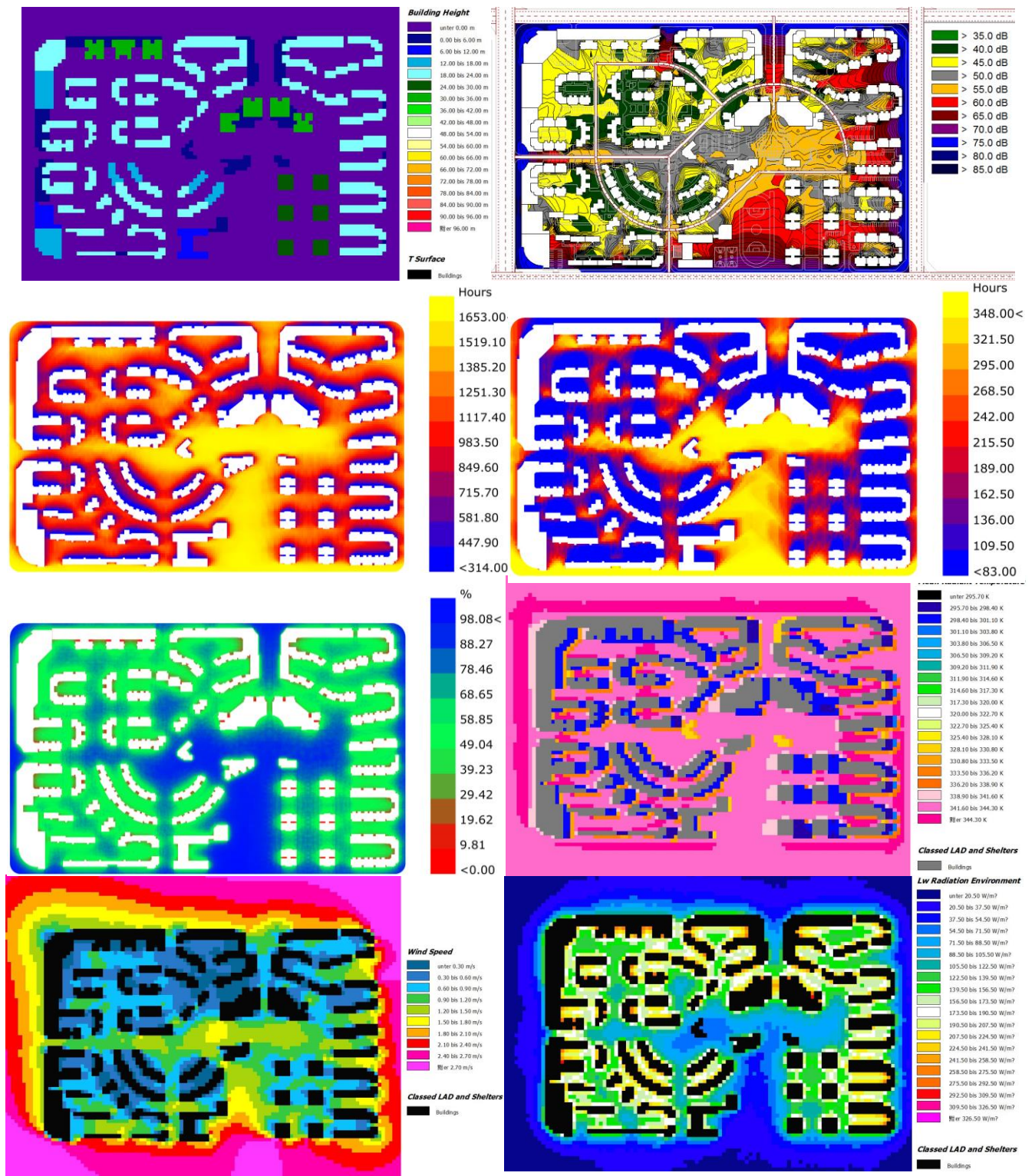
Figure 9.10 Possible Design Pattern 1 of scheme A (Left 1) and Maps of Multiple Simulation Results, Reference on S22 (Right 1 SPL, Left 2 APSH, Right 2 WPSH, Left 3 SVF, Right 4 MRT, Left 5 WS, Right 5 LRE)
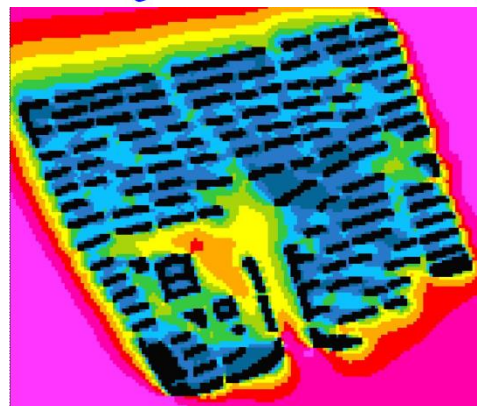
Figure 9.11 Possible Design Pattern 2 for Scheme A (Left 1) and Maps of Multiple Simulation Results, Reference on S1801 (Right 1 SPL, Left 2 APSH, Right 2 WPSH, Left 3 SVF, Right 4 MRT, Left 5 WS, Right 5 LRE)
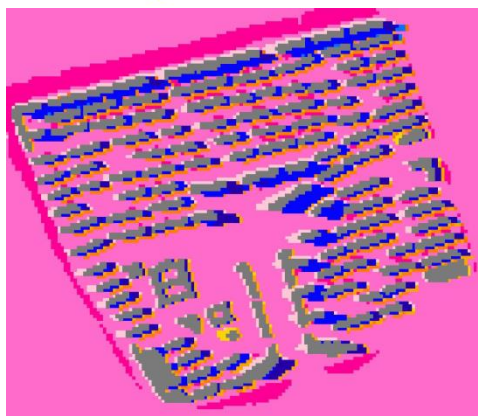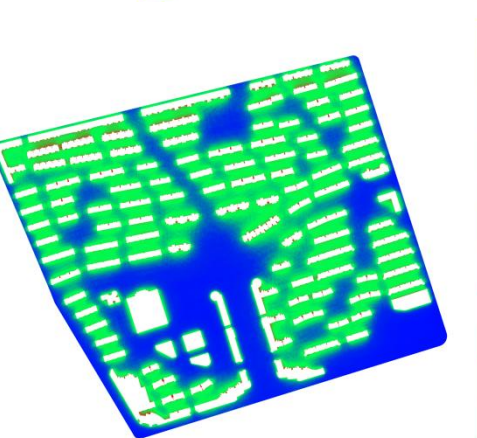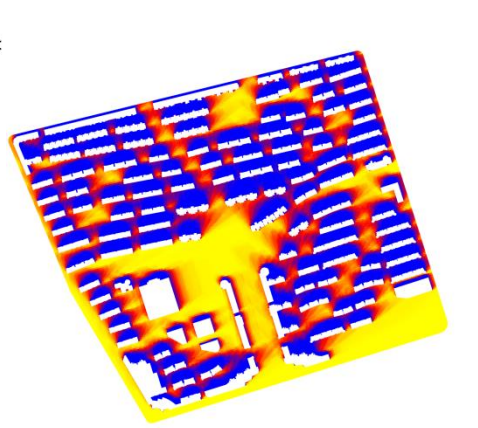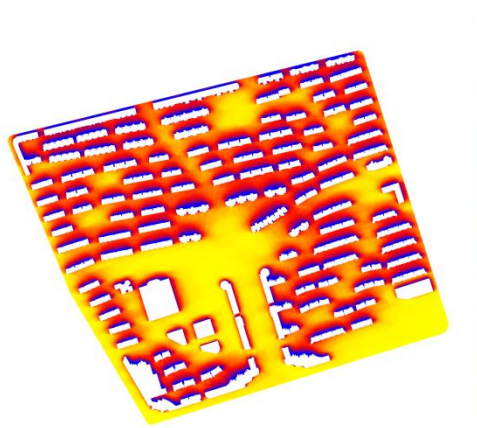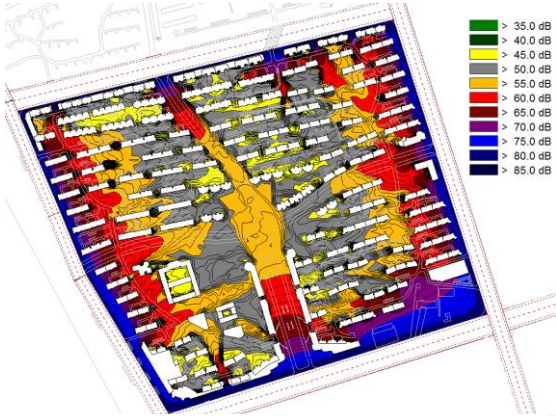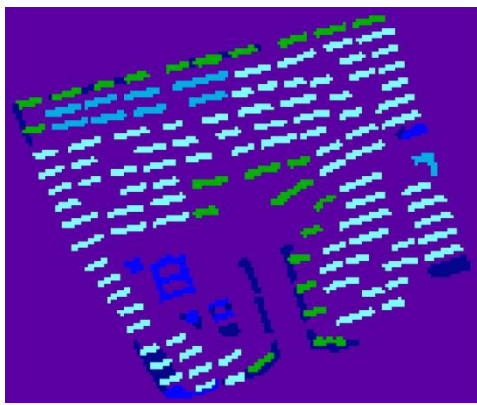
9.5.2 Discussion on Parametric Modelling of Scheme Templates for High Accuracy Simulation

To visualise and validate suggested design schemes, the best approach is to model the templates in each scheme, and to feed the template models for domain simulation in separate high accuracy predictive software. However, this approach requires high dimensional parametric modelling technique to accomplish ward planning models based on value combinations of at least 20 variables. According to the extraction process and definitions of the 20 design variables, it is obvious that the 20 variables are only indicators of the pivotal distribution characteristics of the design pattern. In other words, 20 variables are the necessary condition of a description of a specific distribution, rather than a sufficient condition to generate a specific distribution. High dimensional modelling approaches and further exploitive requirements for sufficient modelling parameters are difficult to achieve. This makes it impossible to generate parametric modelling in scale of residential ward planning at the moment. Due to the length limit of this research, original samples of similar distributions are utilised for suggested design scheme validation by their same distribution pattern and corresponding optimised and balanced performances.

In future work, parametric modelling in high dimension for large scale models, i.e., residential ward planning, should be explored. After achieving an applicable technique of generating large scale building cluster models with complex interactions and relationships, not just suggested design scheme could be visualised and modelled. The models of optimised results could become back-substitutions in domain simulation software for performance tests to validate the predictive accuracy of optimisation models or to be archetypes of new designs.

With the help of high-dimensional parametric modelling, further updating of meta-model and global optimisation results is still encouraged which improves the projection accuracy from design pattern to optimised integrated performance of solutions. In other words, it would enhance the confidence and generalisation level of suggested design schemes.

Figure 9.12 Possible Design Pattern 1 for Scheme C (Left 1) and Maps of Multiple Simulation Results, Reference on S260101 (Middle 1 SPL, Right 1 SVF, Left 2 APSH, Middle 2 WPSH, Right 2 MRT, Left 3 WS, Middle 3 LRE)
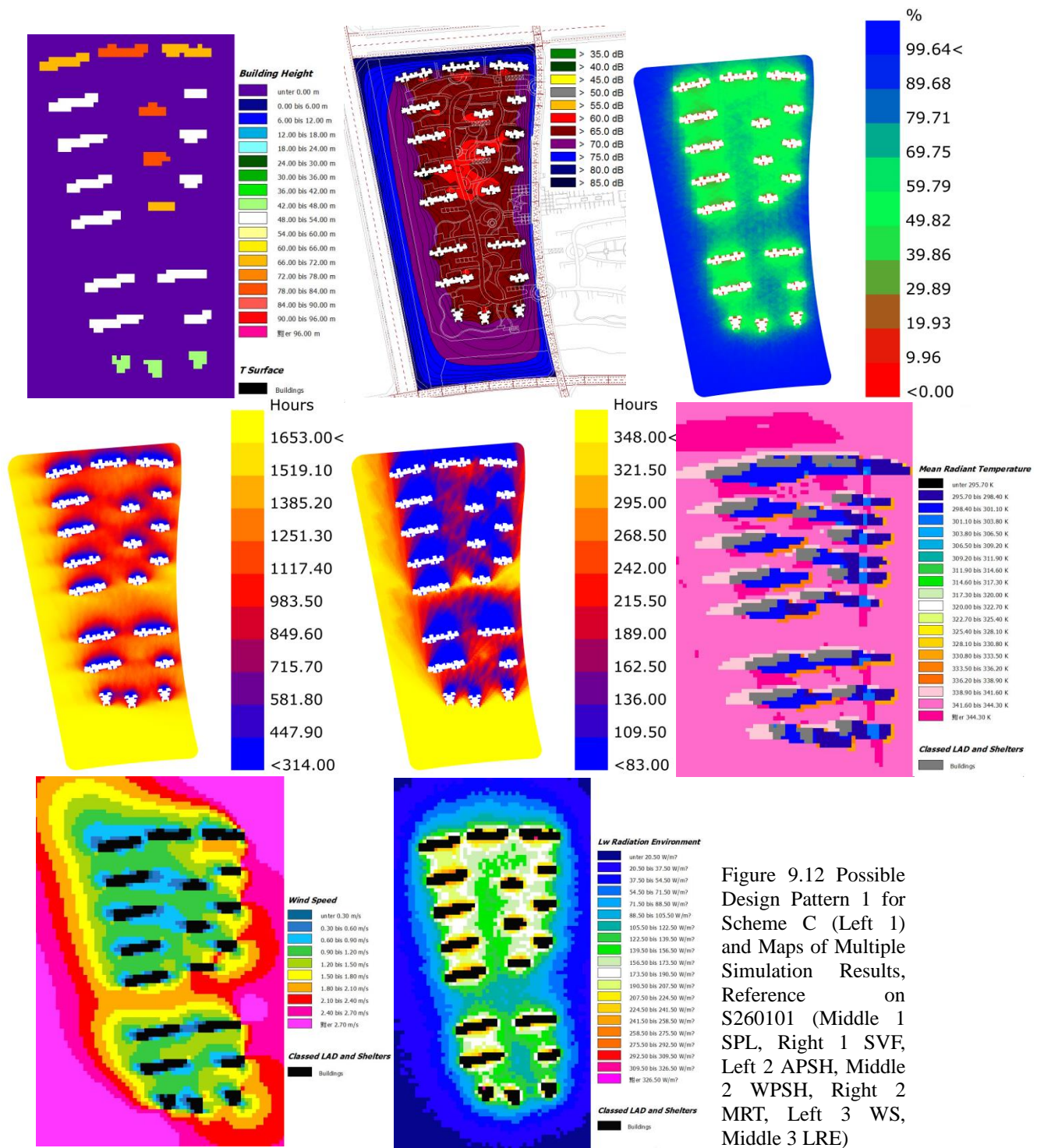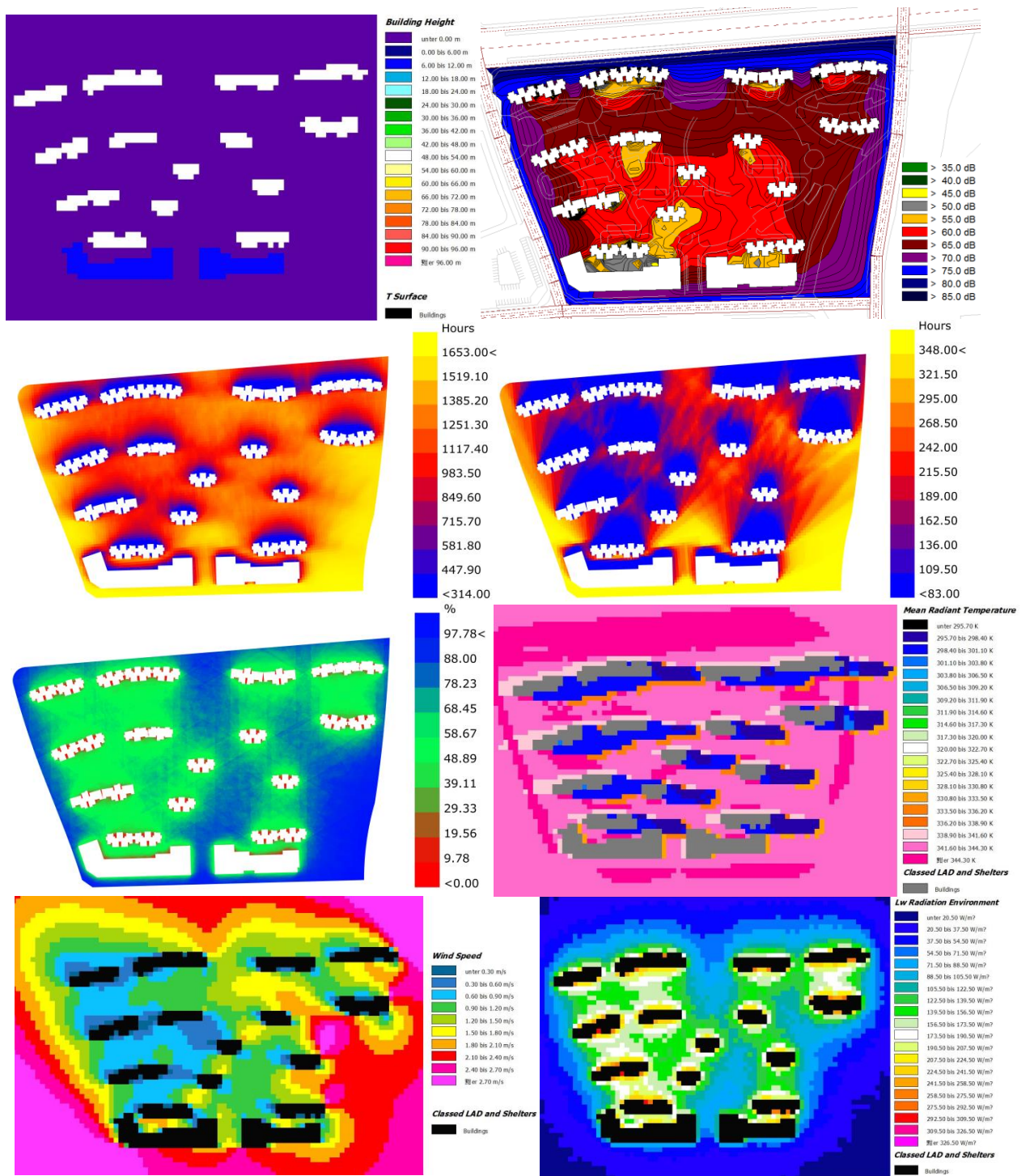
Figure 9.13 Possible Design Pattern 2 for Scheme C (Left 1) and Maps of Multiple Simulation Results, Reference on S260501 (Right 1 SPL, Left 2 APSH, Right 2 WPSH, Left 3 SVF, Right 4 MRT, Left 5 WS, Right 5 LRE)

9.6   Web-based Interactive Tool for Application of MD-MOO Results

Further application of the global optimisation solution relies on the interactive tool for in-time performance prediction and plotting because the needs of testing new possible design schemes never ends and a so fast as to real-time understanding is expected of the integrated performance from a value combination of design parameters. Therefore, with the intention of improvement in certain domain performance, the combination of 20 dimensional variables needs to be calculated and tested in real-time with the three meta-models.

The feasible Pareo solutions of the global optimisation for integrated performance of three domains (acoustic, sunlight and thermal) are extracted into three types of design schemes with 16 templates in total, as well as the descriptive design guidance and suggested numerical range of design variables.

These conclusions could be combined and referred to as archetypes of new designs. However, a large amount of data of feasible solutions cannot be directly used in future designs. Furthermore, a value combination within the suggested range of design variables cannot yet directly project to corresponding integrated performances of the scheme and provide design straightforward reference information. Therefore, an interactive tool is developed to enable users to collect referable, integrated performance information from an input of 20 design variables.

The basic idea is to insert the meta-model prediction module into the webpage-based platform to run on the server. When design parameter values are input, the corresponding performance index values would be predicted by the meta-model for acoustic, sunlight and thermal domain. The selection of design variable combination would be guided by feasible Pareto solution samples available in data set and point on 3d scatter plot on the platform. For clear view, the coordinate is rotatable to view from all angles (figure 9.14).

The tool contains the domain meta-models which represent the projection relationship from design pattern to the corresponding integrated performance. There are 20 slide

bars assigned to the values of the 20 design variables. By limiting the selectable range of 20 design variables to the suggested range, the outcome from the meta-models would be well performed and the domain balanced outdoor environment evaluated from three aspects.
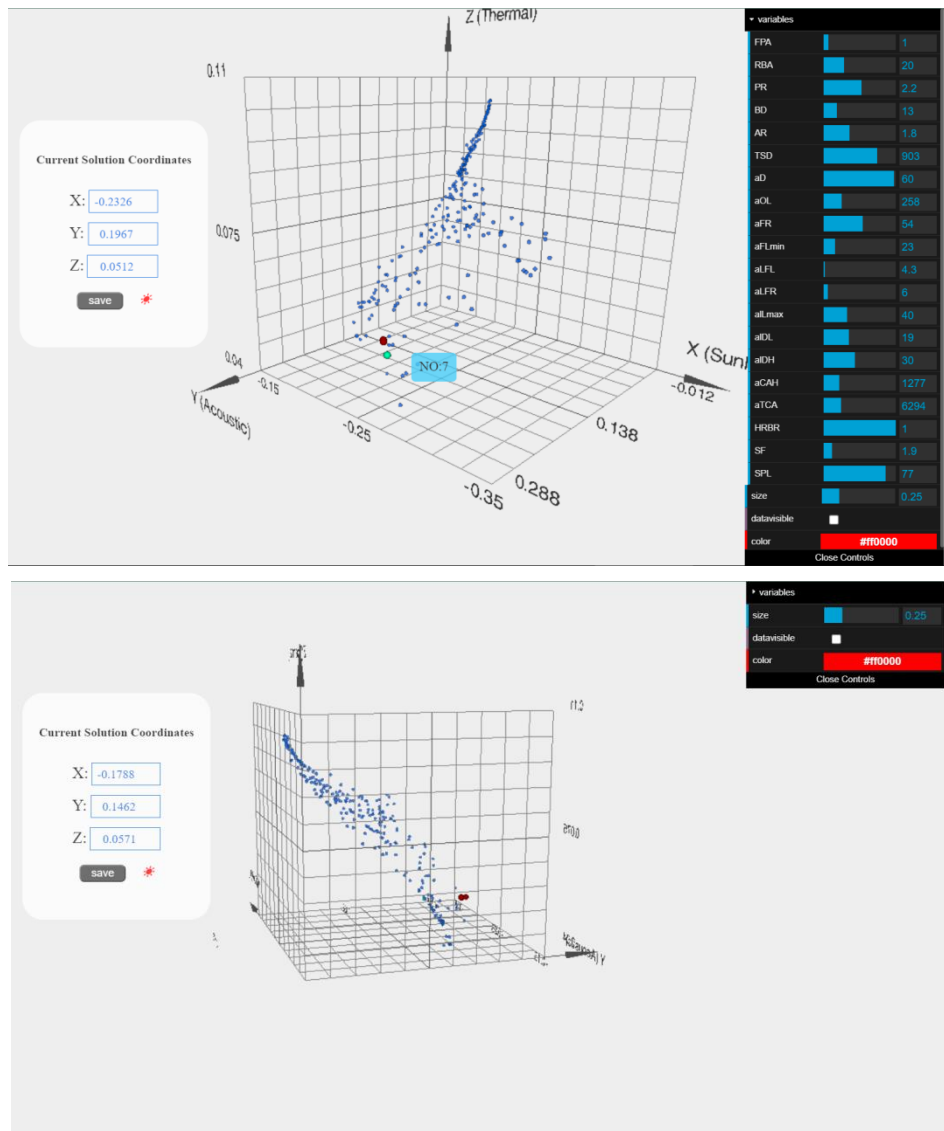

Figure 9.14 User Interface of the Interactive Tool

The normalised domain objectives (NDO) are calculated for each scheme based on the meta-model output as an indicator of holistic performance for each of the three domains. A three-dimensional scatter plot could be generated according to three NDOs of a specific scheme. A scheme would be marked as one point with the values of three coordinates in the chart. The recommended 3D distribution area of feasible Pareto solutions of the global optimisation is also plotted in the scatter chart as a
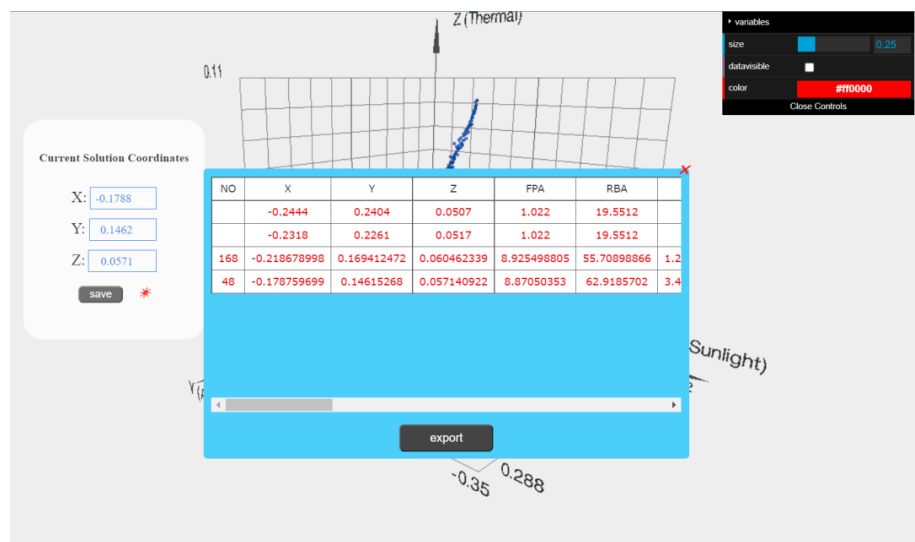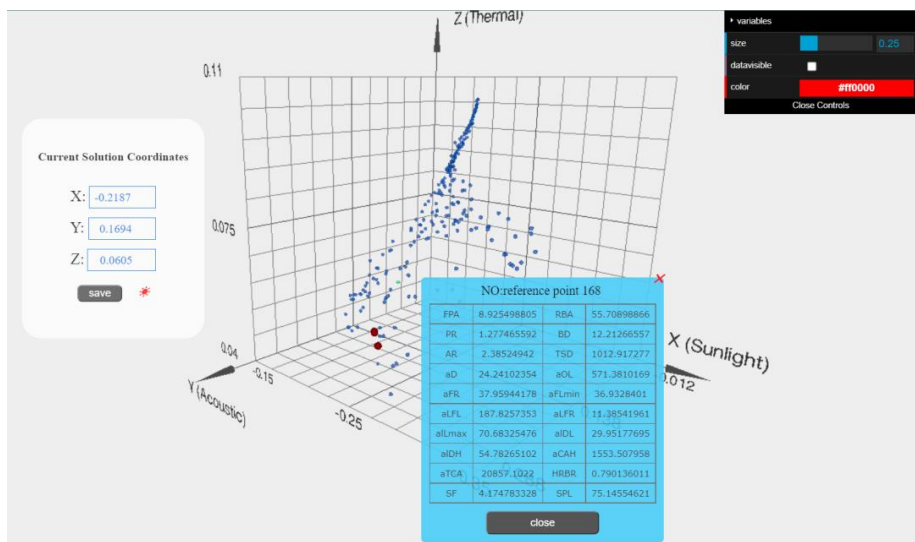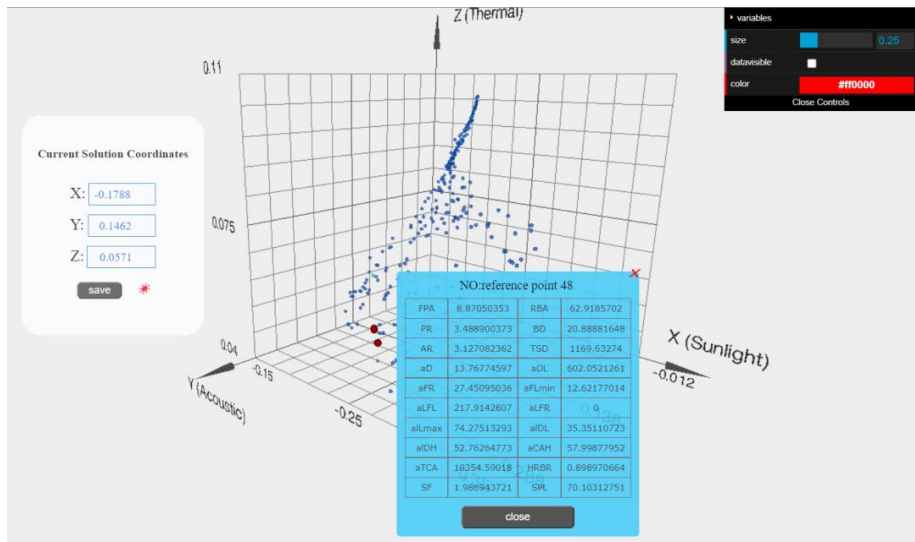
Figure 9.15 Value Combinations Of Reference Cases

and Data Comparison between Scheme and Reference

reference background to allow evaluation of the location of the new scheme point. Only the scheme located within the area of optimised global performance would have a satisfying environmental outcome. In addition to this, the location of the scheme NDOs also represents the domain inclination of the integrated optimisation and possible improvement direction and margin. The custom design scheme could be exported with the interested reference cases including all building distribution parameter values in form of excel file. The value combination of reference cases and data comparisons between scheme and reference are shown in figure 9.15.

This interactive tool would be published on a website once fully developed, so that it could be conveniently accessed by users. Currently, a beta version will be temporarily available online at link http://www.hugoliang.cn:8000/WebForm1.aspx. For future work, a preliminary design assistant application could be developed, which allows a step-by-step, guided selection of possible value combinations for design variables according to submitted design requirements. The numerical amendment suggestions for value combinations based on descriptive requirements, after plotting the scheme on the 3D chart, need further research. With integration of the numerical amendment suggestions, the application would form a closed loop of a value determination system for building distribution indices.
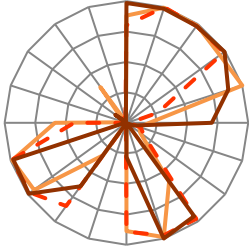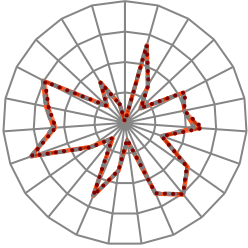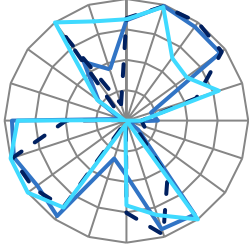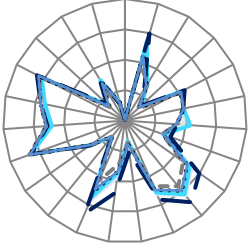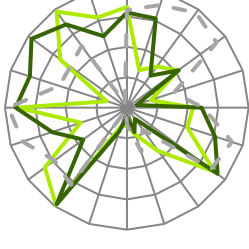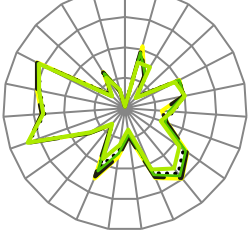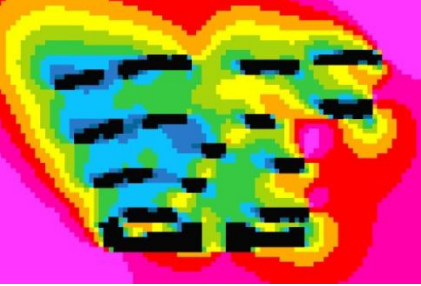
9.7 Summary

This chapter comprises an analysis of the MD-MOO for three integrated environmental performance domains. Firstly, the solution data preparation is introduced for the following analysis, followed by the analysis of multi-domain and single-domain performance improvement achieved by global optimisation. Based on the holistic observation of the feasible Pareto solution set of the MD-MOO over the initial sample designs, three types of design patterns are suggested for a residential ward design to acquire a balanced global performance. The suggested design schemes are validated with representative real-case simulations fitting the suggested design pattern. Finally, a web-based interactive design tool is introduced to carry real-time prediction of global performance for multi-domain (acoustic, sunlight and thermal) studies and design-guiding. It enables designers to directly verify the proposed value

combination of design variables for their global performance.

The analysis of the feasible Pareto solutions of the optimisation concludes three suggested design schemes with 16 templates of design patterns. The applications of the suggested design schemes in manner of descriptive guidance, quantitative selectable range of design variables and an interactive tool to quickly test the schemes are the original results of this research and could contribute to current design progress. The three schemes share optimised and domain balanced performances, and reveal performance inclinations in noise attenuation, sunlight hours and thermal comfort. The characteristics of the design patterns and consequential performances are stated in table 9.3.

Table 9.3 Summary of Suggested Design Scheme

	X Diagram	X arrangement	Y Diagram	Y result	Possible Distribution And Performance
Scheme A		Large Site Scale; Large Building Amount; Medium Building Density; Facade Protection; Cluster Protection;		Most balanced of the three; Acoustic advantages; Extra enhancement in sunlight by reducing building density and increasing edge open space;	
Scheme B		Large Site Scale; Large Building Amount; Facade Protections; Cluster Protection; Medium Opening Encouragement; Large Edge Open Space;		Sunlight advantages; Extra enhancement in acoustic by reducing edge open space; Extra enhancement in thermal by increasing high-rises;	NA
Scheme C		Large Amount of High-Rises; Cluster Protection; High Opening Encouragement; Large Edge Open Space; Various Site Scales.		Thermal advantage; Extra enhancement in sunlight by reducing plot ratio.	

Chapter 10 Conclusion

Following the requirements of holistic design for cross-domain environmental performance and decision making support at early design stage, a three levelled systematic optimisation with integrated simulations are operated. The summary of the contents of the whole thesis are presented in figure 10.1.

10.1   Achievement

Following the research aims and objectives, seven achievements are obtained through this research.

Firstly, statistical analysis methods are utilised in building distribution parametric studies. The statistical measures of building distribution indices including conventional and innovative morphological parameters of single buildings and neighbourhood are consolidated and compared, forming 38 building distribution parameters for further analysis. As a conclusion the grouping of indices based on the characteristics, level grading of indices, interaction rules between indices pairs are achieved.

Secondly, as the provider of data basement, high accuracy simulations over environmental evaluative metrics of each domain based on 44 sample models are initially implemented in three simulation software. Statistical analyses are applied to performance data so that the statistical measurements of performance assessment metrics are generated in three researched domains. They are not only the selected assessment indices of performances but also the design objectives in single domain multi-objective optimisation (SD-MOO) and multi-domain multi-objective optimisation (MD-MOO).
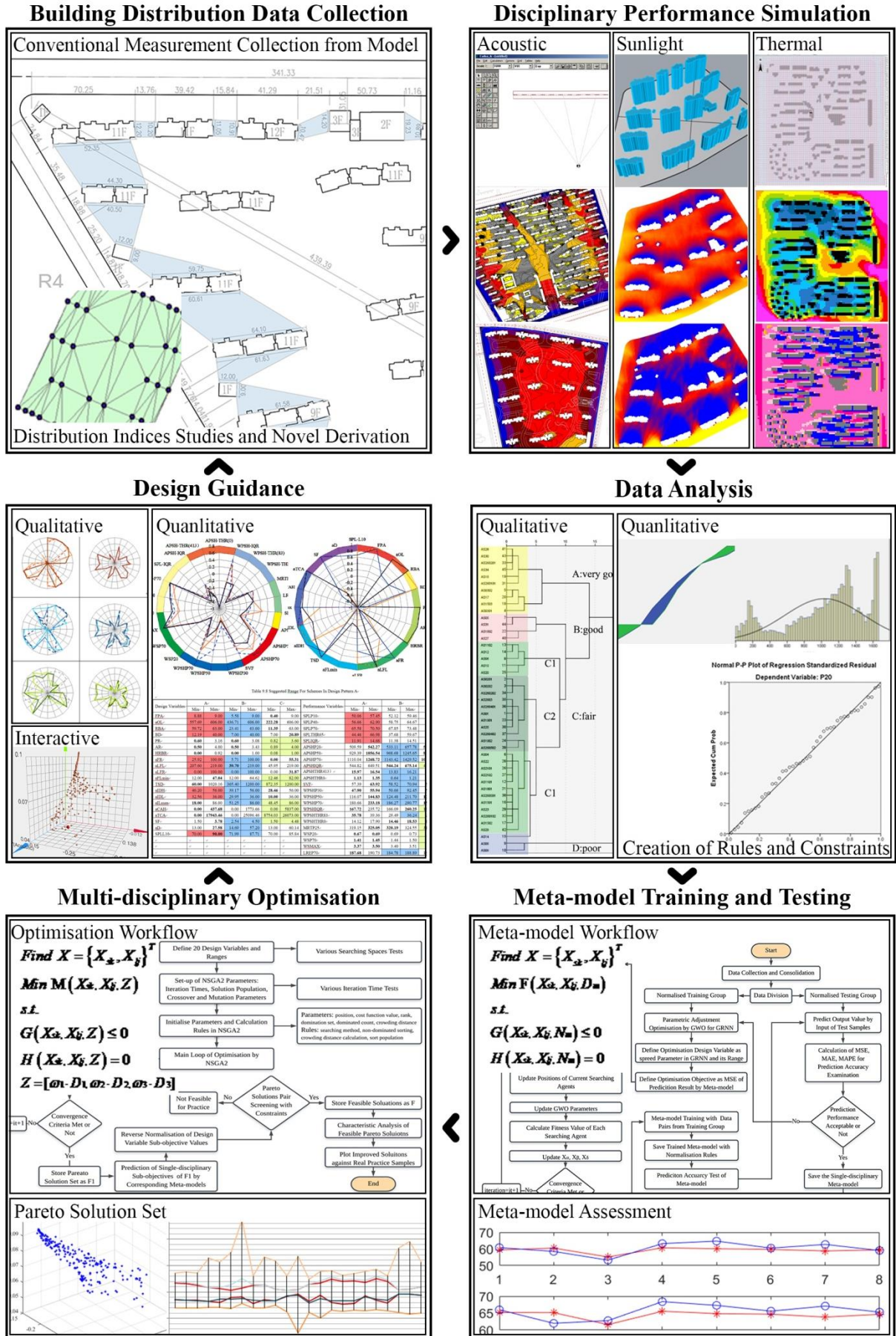
Figure 10.1 summary of the content of this research

Thirdly, qualitative and quantitative analyses are operated for the collected data pairs of building distribution data and environmental performance data. Regarding domain performances, namely acoustic, sunlight possibility and thermal comfort, clusters of the statistical measurements of evaluative metrics, most representative measurements of corresponding performance objective, and underlined rules of environmental performance in relation to building distribution parameters are analysed. Quantitatively, multiple linear regressions (MLR) of representative evaluative variables on building distribution parameters are operated and are applied in following multi-objective optimisation as sensitive analysis of design variables. These qualitative and quantitative results form the descriptive and numeric guidelines for single domain performance design.

Fourthly, multi-domain multi-objective optimisation (MD-MOO) is executed over three environmental domains, noise attenuation, sunlight possibility and thermal comfort, to explore appropriate combination of building distribution to achieve balanced and satisfying performances in multiple and adverse objectives. A hierarchical structure of collaborative optimisation strategy is generated for the optimisation problem of cross-domain environmental performance, which means system level and local level optimisation are executed simultaneously, and local level only exchange data with system level. This structure could be generalised to other cross-domain performance optimisation problem for managements at local simulation, single domain optimisation and global optimisation.

Fifthly, single-domain optimisation and construction of meta-model by Generic Regression Neural Network (GRNN) improved by Grey Wolf Optimiser (GWO) are operated beforehand as basis of multi-domain optimisation. The meta-models of three domains serve not only as single-domain optimisation model, but also as substitutions of the high accuracy simulation models of each domain to reduce time and calculation cost.

Sixthly, for multi-domain multi-objectives optimisation, algorithm of non-dominated genetic algorithm with elitist strategy (NSGAII) is adopted to calculate optimised solution set namely Pareto solution set. At last the Pareto solution set is translated into three dominant building distribution categories grouped by optimised objectives of integrated outdoor environmental performance in residential wards, with various design templates and rules, for future design references.

At last, a graphical interactive tool is achieved for a designer-friendly experience of guided design in early design stage regarding residential ward planning. The optimised solutions are imported and plotted as suggestive design schemes for designers. With reference on the suggestive scheme, the customised scheme could be fast tested regarding its global environmental performance in three domains, and be compared with the suggestive schemes.

10.2   Suggestions for Future Studies

As mentioned in the thesis, this research accomplished an explorative attempt to construct a three-levelled optimisation structure for MD-MOO in cross-domain environmental performance context in residential ward in south-east China. The three layers of local simulation, local optimisation and global optimisation are expatiated and constructed respectively. The emphasis is laid on the global optimisation.

From the viewpoint of application of the interactive tool in early design stage to guide urban pattern forming according to integrated performance expectation, in future work, it would be helpful to involve GIS as an input platform of building distribution parameter. GIS is capable of providing existing residential ward building distribution data as samples for area with digital city database. With tools of urban form visualisation and assessment, it is possible to visual present the pattern based on the building distribution parameter input. Then apply the distribution input into the

interactive tool for fast calculation of multi-domain performance. By compare the input distribution data and pattern and global performance with the optimised solutions, a fast graphical illustrated early design trial-and-test iteration is applicable.

From the viewpoint of enhancing systematic predictive accuracy and reducing impact from predictive errors, the data iterations is essential for a fully validated and generalisation ability tested global optimisation. The data iteration need to work through and loop among the procedures of 1) building distribution parameter (design requirement) 2) transformation to urban form (parametric modelling) 3) input to local simulation (multiple single-domain simulations) 4) input to meta-model training (meta-model training) 5) input to global optimisation (MD-MOO) 6) optimised distribution parameter (Pareto solution analysis) 7) transformantion to urban form (parametric modelling). The iteration will enhance the prediction accuracy of the whole optimisation system via self-learning of the meta-model and optimisation iterations. Furthermore, the re-simulation of optimised solutions in the iteration will valid the whole system against real practice.

Under the limitation of high-dimensional parametric modelling of a meso-scale or neighbourhood scale building cluster model, the optimised solutions from the global optimisation in this research could not be transferred to 3D models currently. Hence the iteration of simulation from the optimised solutions is not operated. Therefore, in future studies to enhance the systematic accuracy, the selection of parameters to describe and to model the neighbourhood is the first step. And the second step is to implement the high-dimensional parametric modelling in a corresponding platform.

Once the parametric modelling is available for meso-scale models, a systematic sampling is essential to expand the model population for further optimisation. The design of computational experiment (DoCE) approach is required for evenly generation of samples to cover the whole possible design space. The Latin hypercube sampling (LHS) approach is recommended, because it is capable of evenly and

systematically generation of sample with respects to existing samples. For architectural and urban design, there exist vast amounts of validated optimal design schemes or trade-off designs with consideration of various constraint factors. These existing schemes should be adopted in the sampling system of the global optimisation, for an honest representation of the real practice. Therefore, LHS is appropriate for this task.

Regarding to the predictive accuracy of the multi-domain optimisation system, the propagation of errors should be conducted in future work. The main focus would be first estimating the ranges of variation of the DV impacted by the predictive errors of the optimisation functions and meta-models. Secondly, the possibility of define weight in the optimisation based on their reliability, which is the weights for single-domain objective variables to form the single-domain optimisation function and for the single-domain optimisation functions to form multi-domain optimisation function.

The integration of simulation for different domains in one platform gains gradually increasing requirements. As seen from this research, the multiple simulation progresses and optimisation progresses could be integrated into one platform, in a run-time manner. This could enhance the productive efficiency of the simulation-optimisation iteration.

Since three performance domains in architecture design and urban design field are considered, it is also possible involving multiple performance fields into consideration, for example, renewable energy (solar photovoltaic potential, urban wind turbine potential) or building energy consumption, etc. Through adjustment on the domain weight and single-domain function weight, the emphasis of the multi-domain multi-objective optimisation could be customised.

The interactive tool in this research is a beta-version with basic functions of the design guiding, scheme comparison and fast prediction of multi-domain environmental performance. A further enhancement of this design guiding tool is required. The idea of implanting fast prediction meta-model in an interactive tool could be applied in various context of design scheme test.

Finally, from a long term view point, the generative architecture or urban form concept is the overall direction of the MDO system applied in these fields as mentioned by Marsault(Marsault, 2017). Under the possibility of generation of topology form following the rules and constraints specified by designers, facilitation of assessive calculation, morphological recognition, mathematical and graphical interpretation, and interaction with opertators into the generation platform, will implement the evolutionary design procegure. The morphology is evolutionary improving itsself to fit for the specified rules. And designer could select from the pool of optimised design schemes.

Appendix A Sample Site Location and Climate Information

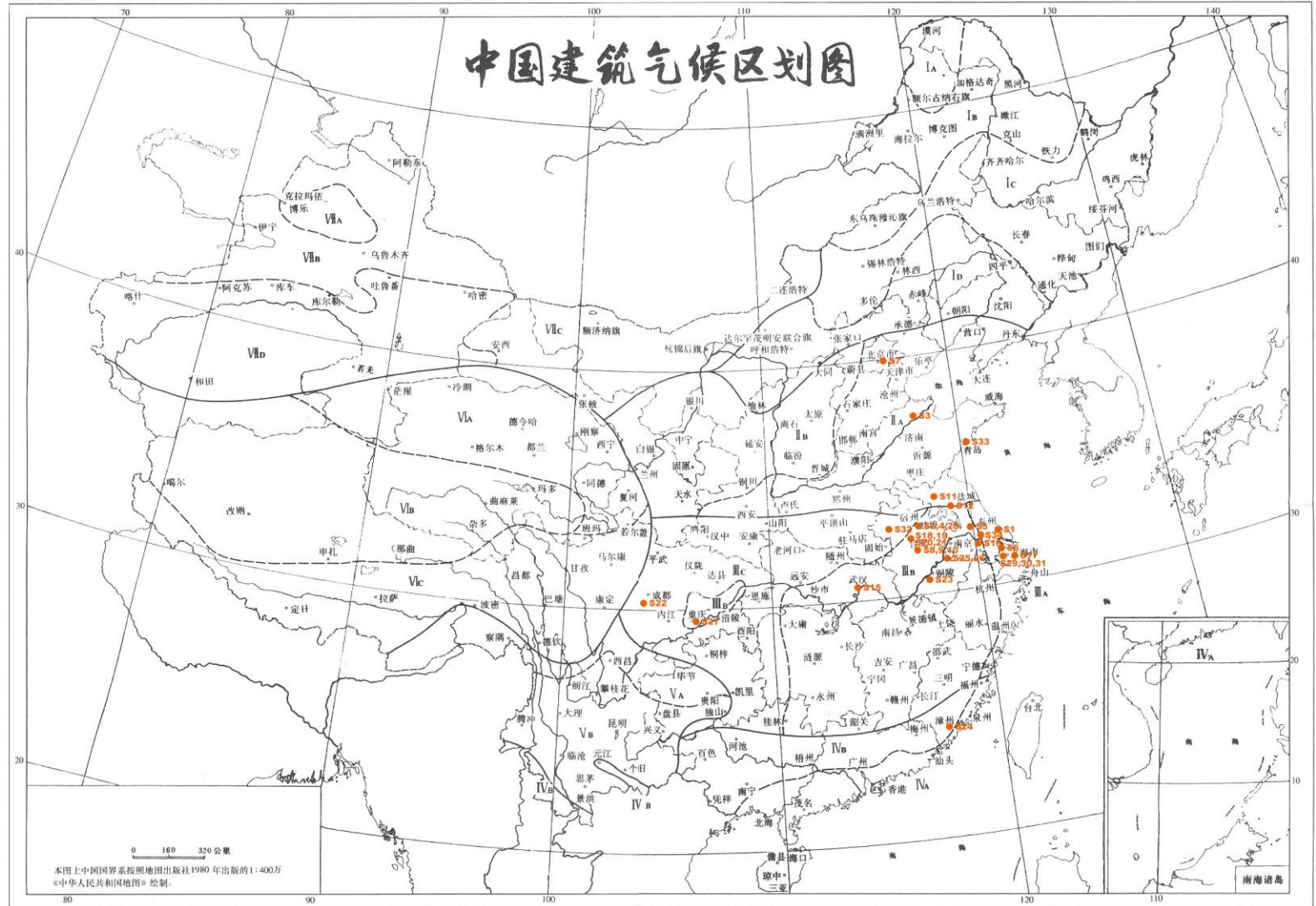Figure A1 Sample Site Location over China Architectural Climate Zoning Map

Table A1 Sample Site Location and Climate Information

Site NO.	Location	Province	City Levels	Traditional Region	Economic Region	Area (km2)	Population (million)	GDP (¥ billion)	Centre Coordinates	Architectural Climate Zoning	Climate Classification
1	DONGTAI	JIANGSU	County-level city	East China	East Coast	3240	1.16	50.67	32°45'N 32.75°N 120°23'E 120.3833°E	III A	Subtropical Monsoon Climate
2	BENGBU	ANHUI	Prefectural-level city	East China	Central China	5952	3.68	8.9	32°55'N 32.9167°N 117°22'E 117.3667°E	III B	Transitional Climate
3	BINGZHOU	SHANDONG	Prefectural-level city	East China	East Coast	9453	3.79	181.76	37°22'N 37.3667°N 117°58'E 117.9667°E	II A	Temperate Monsoon Climate
4	BENGBU	ANHUI	Prefectural-level city	East China	Central China	5952	3.68	8.9	32°55'N 32.9167°N 117°22'E 117.3667°E	III B	Transitional Climate
5	TAIZHOU	JIANGSU	Prefectural-level city	East China	East Coast	5787	5.07	27.01	32°30'N 32.5°N 119°56'E 119.9333°E	III B	Subtropical Monsoon Climate
6	SHANGHAI JIADING	SHANGHAI	Districts of Municipalities	East China	East Coast	463.5	1.47	91.5	31°23'N 31.3833°N 121°15'E 121.25°E	III A	Subtropical Monsoon Climate
7	BEIJING CHAOYANG	BEIJING	Districts of Municipalities	North China	East Coast	470.8	3.54	362.77	39.975°N 39°58'N 116.525° E 116°31'E	II A	Temperate Monsoon Climate
8	HEFEI	ANHUI	Prefectural-level city	East China	Central China	11408.4	7.52	416.43	31°52'N 31.8667°N 117°17'E 117.2833°E	III B	Subtropical Monsoon Climate
9	HEFEI	ANHUI	Prefectural-level city	East China	Central China	11408.4	7.52	416.43	31°52'N 31.8667°N 117°17'E 117.2833°E	III B	Subtropical Monsoon Climate
10	HEFEI	ANHUI	Prefectural-level city	East China	Central China	11408.4	7.52	416.43	31°52'N 31.8667°N 117°17'E 117.2833°E	III B	Subtropical Monsoon Climate
11	SUQIAN	JIANGSU	Prefectural-level city	East China	East Coast	8555	5.55	130.4	33°57'N 33.95°N 118°16'E 118.2667°E	II A	Transitional Climate

12	HUAI'AN	JIANGSU	Prefectural-level city	East China	East Coast	10072	5.39	192.09	32°22'N 32.37°N 116°35'E 116.59°E	II A-III B	Transitional Climate
15	WUHAN	HUBEI	Sub-provincial city	South Central	Central China	8494	10.02	67.62	30°35'N 30.5833°N 114°18'E 114.3°E	III B	Subtropical Monsoon Climate
16	JIANGYIN	JIANGSU	County-level cities	East China	East Coast	987.53	1.62	233.58	31°55'N 31.9167°N 120°17'E 120.2833°E	III B	Subtropical Monsoon Climate
17	SHANGHAI PUDONG	SHANGHAI	Districts of Municipalities	East China	East Coast	1210	5.26	400.1	31°13'N 31.2167°N 121°32'E 121.5333°E	III A	Subtropical Monsoon Climate
18	HUAINAN	ANHUI	Prefectural-level city	East China	Central China	2585	2.45	70.95	32°37'N 32.6167°N 116°59'E 116.9833°E	III B	Subtropical Monsoon Climate
19	HUAINAN	ANHUI	Prefectural-level city	East China	Central China	2585	2.45	70.95	32°37'N 32.6167°N 116°59'E 116.9833°E	III B	Subtropical Monsoon Climate
20	HUAI'AN	JIANGSU	Prefectural-level city	East China	East Coast	10072	5.39	192.09	32°22'N 32.37°N 116°35'E 116.59°E	II A-III B	Transitional Climate
21	HUAI'AN	JIANGSU	Prefectural-level city	East China	East Coast	10072	5.39	192.09	32°22'N 32.37°N 116°35'E 116.59°E	II A-III B	Transitional Climate
22	MEISHAN	SICHUAN	Prefectural-level city	Southwest China	Western China	7186	3.49		30°04'N 30.0667°N 103°50'E 103.8333°E	III C	Subtropical Monsoon Climate
23	CHIZHOU	ANHUI	Prefectural-level city	East China	Central China	8271.7	1.6	30.08	30°39'N 30.65°N 117°29'E 117.4833°E	III B	Subtropical Monsoon Climate
24	ZHANGZHOU	FUJIAN	Prefectural-level city	East China	East Coast	12600	4.84	176.82	23°34'N 25°15'N 116°54'E 118°08'E	IV B	Subtropical Monsoon Climate
25	WUHU	ANHUI	Prefectural-level city	East China	Central China	5988	3.84	187.36	31°19'N 31.33°N 118°22'E 118.38°E	III B	Subtropical Monsoon Climate
26	WUHU	ANHUI	Prefectural-level city	East China	Central China	5988	3.84	187.36	31°19'N 31.33°N 118°22'E 118.38°E	III B	Subtropical Monsoon Climate

27	CHONGQING	CHONGQING	Municipalities	Southwest China	Western China	82402.95	29.36	1145.9	29°33'N 29.55°N 106°33'E 106.55°E	III B	Subtropical Monsoon Climate
28	BENGBU	ANHUI	Prefectural-level city	East China	Central China	5952	3.68	8.9	32°55'N 32.9167°N 117°22'E 117.3667°E	III B	Transitional Climate
29	SHANGHAI MINHANG	SHANGHAI	Districts of Municipalities	East China	East Coast	372	2.48	148.3	31°3'N 31.05°N 121°15'E 121.25°E	III A	Subtropical Monsoon Climate
30	SHANGHAI MINHANG	SHANGHAI	Districts of Municipalities	East China	East Coast	372	2.48	148.3	31°3'N 31.05°N 121°15'E 121.25°E	III A	Subtropical Monsoon Climate
31	SHANGHAI MINHANG	SHANGHAI	Districts of Municipalities	East China	East Coast	372	2.48	148.3	31°3'N 31.05°N 121°15'E 121.25°E	III A	Subtropical Monsoon Climate
32	FUYANG	ANHUI	Prefectural-level city	East China	Central China	9975	10.25	96.25	32°53'N 32.8833°N 115°48'E 115.8°E	III B	Transitional Climate
33	QINGDAO	SHANDONG		East China	East Coast	10654	7.66	661.56	36°04'N 36.0667°N 120°22'E 120.3667°E	II A	Temperate Monsoon Climate
34	JINGJIANG	JIANGSU	County-level cities	East China	East Coast	665.04	0.66	60.08	31°58'N 31.9667°N 120°16'E 120.2667°E	III B	Subtropical Monsoon Climate

Table A2 Sample Site Climate Details

Site NO.	Location	Avg Temperature			Relative Humidity%		Avg Wind Speed (m/s)			Annual Most/Least Frequent Wind Direction/Frequency		Summer Predominant Wind Direction/Frequency			Daylighting Hours(h)				Percentage of Daylighting				Angle of Incidence	
		Hottest Month	Coldest Month	Annual	Coldest Month	Hottest Month	Annual	Summer	Winter	Most Frequent	Least Frequent	Jun	Jul	Aug	Annual	Dec	Jan	Feb	Annual	Dec	Jan	Feb	Winter Solstice	Day of Great Cold
1	DONGTAI	27	0.7	14.2	84	74	3.4	3.3	3.4	ESE 10	WSW,W 3	ESE 17	ESE 13	ESE 13	2118.8	150.4	1445.1	136.8	48	49	46	44	33.6	36.9
2	BENGBU	28	1	15.1	80	71	2.5	2.3	2.5	C18ENE 11	N,WSW,W,NW,NNW 3	C24SSE 12	C25ENE 10	C26ENE 17	2118.8	150.4	1445.1	136.8	48	49	46	44	33.6	36.9
3	BINGZHOU	27.4	-1.4	14.2	73	53	3.2	2.8	3.1	SSW 16	ESE,SE 1	SSW 19	C17SSW 15	C20ENE 15	2118.8	150.4	1445.1	136.8	48	49	46	44	33.6	36.9
4	BENGBU	28	1	15.1	80	71	2.5	2.3	2.5	C18ENE 11	N,WSW,W,NW,NNW 3	C24SSE 12	C25ENE 10	C26ENE 17	2716.6	185.6	188.8	183.4	62	62	62	59	29.8	33.1
5	TAIZHOU	27.4	1.5	14.7	85	76	3.4	3.3	3.5	SE 10	SW,WSW,W 3	SE 16	SE 15	SE 15	1990.9	141.1	130.2	116.7	45	45	41	37	35.5	38.8
6	SHANGHAI JIADING	27.8	3.5	15.7	83	75	3.1	3.2	3	ESE 10	SW,WSW 2	ESE,SE 16	SSE 19	ESE 17	1212.6	33.4	39.1	46.3	27	11	12	14	36.9	40.2
7	BEIJING CHAOYANG	25.9	-4.5	11.6	77	44	2.5	1.9	2.8	C20N 10	W 1	C17S 9	C25S 9	C30N 10	2776	192.5	204.7	196.8	63	66	68	65	26.7	30
8	HEFEI	28.2	2	15.7	81	75	2.7	2.7	2.6	C18ENE 9	SW,WSW,W 2	C15SE 13	S 17	C17ENE 9	2309	172	167.3	156.9	52	56	53	60	33.1	36.4
9	HEFEI	28.2	2	15.7	81	75	2.7	2.7	2.6	C18ENE 9	SW,WSW,W 2	C15SE 13	S 17	C17ENE 9	2118.8	150.4	1445.1	136.8	48	49	46	44	33.6	36.9
10	HEFEI	28.2	2	15.7	81	75	2.7	2.7	2.6	C18ENE 9	SW,WSW,W 2	C15SE 13	S 17	C17ENE 9	2127	152.6	142.4	129.8	48	49	45	41	34.6	37.9
11	SUQIAN	27.3	-0.2	14.4	81	68	2.6	2.5	2.7	ENE 12	N,WSW,W,WNW 3	E,ESE,SE 10	C13ENE 10	C15ENE 16	2127	152.6	142.4	129.8	48	49	45	41	34.6	37.9
12	HUAI'AN	27	0.7	14.2	84	74	3.4	3.3	3.4	ESE 10	WSW,W 3	ESE 17	ESE 13	ESE 13	2127	152.6	142.4	129.8	48	49	45	41	34.6	37.9
15	WUHAN	28.7	3	16.3	79	76	2.6	2.5	2.6	NNE 14	WSW,W,WNW 2	C13SE 9	C12SSW 10	NNE 14	2309	172	167.3	156.9	52	56	53	60	33.1	36.4
16	JIANGYIN	27.4	1.5	14.7	85	76	3.4	3.3	3.5	SE 10	SW,WSW,W 3	SE 16	SE 15	SE 15	2118.8	150.4	1445.1	136.8	48	49	46	44	33.6	36.9
17	SHANGHAI PUDONG	27.8	3.5	15.7	83	75	3.1	3.2	3	ESE 10	SW,WSW 2	ESE,SE 16	SSE 19	ESE 17	2118.8	150.4	1445.1	136.8	48	49	46	44	33.6	36.9
18	HUAINAN	28	1	15.1	80	71	2.5	2.3	2.5	C18ENE 11	N,WSW,W,NW,NNW 3	C24SSE 12	C25ENE 10	C26ENE 17	2241.4	170.8	163.3	151.3	51	55	52	49	34	37.3
19	HUAINAN	28	1	15.1	80	71	2.5	2.3	2.5	C18ENE 11	N,WSW,W,NW,NNW 3	C24SSE 12	C25ENE 10	C26ENE 17	2241.4	170.8	163.3	151.3	51	55	52	49	34	37.3
20	HUAI'AN	27	0.7	14.2	84	74	3.4	3.3	3.4	ESE 10	WSW,W 3	ESE 17	ESE 13	ESE 13	1200.4	62.4	68.7	61.5	27	20	21	20	35.8	39.1

#	City																								
21	HUAI'AN	27	0.7	14.2	84	74	3.4	3.3	3.4	ESE 10	WSW,W 3	ESE 17	ESE 13	ESE 13	1200.4	62.4	68.7	61.5	27	20	21	20	35.8	39.1
22	MEISHAN	25.5	5.4	16.1	85	81	1.1	1.1	0.9	C42NNE 11	E,ESE 1	C40N NE 7	C41N NE 9	C44N 9										
23	CHIZHOU	28.6	3.2	16.2	79	75	3	2.9	3.1	NE 20	SSE 0	C18SW 17	SW 23	NE 17	2508.6	188	190.4	180.6	56	62	61	59	30.4	33.7
24	ZHANGZHOU	28.7	12.7	21	80	76	1.6	1.6	1.6	C36ESE 17	NNE,NE,SSW,SW,WSW 1	C38ESE 15	C34S 10	C36ESE 10	1989.9	147.2	138.3	117.5	44	46	43	38	35.3	38.6
25	WUHU	28.6	3.2	16.2	79	75	3	2.9	3.1	NE 20	SSE 0	C18SW 17	SW 23	NE 17	1989.9	147.2	138.3	117.5	44	46	43	38	35.3	38.6
26	WUHU	28.6	3.2	16.2	79	75	3	2.9	3.1	NE 20	SSE 0	C18SW 17	SW 23	NE 17	1989.9	147.2	138.3	117.5	44	46	43	38	35.3	38.6
27	CHONGQING	28.5	7.5	18.2	75	82	1.3	1.4	1.2	C33N 11	ESE 1	C37N 10	C29NE 8	C30NE 8	1989.9	147.2	138.3	117.5	44	46	43	38	35.3	38.6
28	BENGBU	28	1	15.1	80	71	2.5	2.3	2.5	C18ENE 11	N,WSW,W,NW,NNW 3	C24SSE 12	C25ENE 10	C26ENE 17	1989.9	147.2	138.3	117.5	44	46	43	38	35.3	38.6
29	SHANGHAI MINHANG	27.8	3.5	15.7	83	75	3.1	3.2	3	ESE 10	SW,WSW 2	ESE,SE 16	SSE 19	ESE 17	2346.3	166.5	161	152.7	53	54	51	49	32.9	36.2
30	SHANGHAI MINHANG	27.8	3.5	15.7	83	75	3.1	3.2	3	ESE 10	SW,WSW 2	ESE,SE 16	SSE 19	ESE 17	2241.4	170.8	163.3	151.3	51	55	52	49	34	37.3
31	SHANGHAI MINHANG	27.8	3.5	15.7	83	75	3.1	3.2	3	ESE 10	SW,WSW 2	ESE,SE 16	SSE 19	ESE 17	2045.9	138.7	123.7	108.4	46	37	34	32	35.9	39.2
32	FUYANG	28	1	15.1	80	71	2.5	2.3	2.5	C18ENE 11	N,WSW,W,NW,NNW 3	C24SSE 12	C25ENE 10	C26ENE 17	1990.9	141.1	130.2	116.7	45	45	41	37	35.5	38.8
33	QINGDAO	25.2	-1.2	12.2	85	63	5.4	4.9	5.6	SSE 16	NE,ENE,WSW 1	SSE 30	SSE 29	SSE 20	1990.9	141.1	130.2	116.7	45	45	41	37	35.5	38.8
34	JINGJIANG	27.4	1.5	14.7	85	76	3.4	3.3	3.5	SE 10	SW,WSW,W 3	SE 16	SE 15	SE 15	2019.4	171.7	145.9	99.9	16	52	44	31	42	45.5

Appendix B Matlab Program Script of Meta-model Training via GWO-GRNN

I. Meta-Model Training For Acoustic Domain

%% clear all variable value
clc;
clear all
close all
nntwarn off;

%% load acoustic data of input P and output T
load data_sheng;
%% load sunlight data of input P and output T in two groups for separate utilisation for two modes' training
%load data3;
%load data4;
%% %% load sunlight data of input P and output T in two groups for separate utilisation for two modes' training
%load data5;
%load data6;

%% load data and separate data for training set and testing set
train_input=p(:,1:35);
train_output=t(:,1:35);
test_input=p(:,36:43);
test_output=t(:,36:43);

%%normalisation of input and output
[p_train,inputs]=mapminmax(train_input);
[t_train,outputps]=mapminmax(train_output);
p_test=mapminmax('apply',test_input,inputs);
t_test=mapminmax('apply',test_output,outputps);

%% applying GWO for optimisation of GRNN parameter
SearchAgents_no=10; %number of wolf packs
Max_iteration=20; %maximum iteration number
dim=1; %two parameters c and g, need to be optimised
lb=0.01; %lower bound of parameter
ub=3; %upper bound of parameter

Alpha_pos=zeros(1,dim); %initialise the position of alpha wolf

Alpha_score=inf; %initialise value of objective function for alpha wolf, change this to -inf for maximization problems

Beta_pos=zeros(1,dim); % initialise the position of beta wolf
Beta_score=inf; %initialise value of objective function for beta wolf, change this to -inf for maximization problems

Delta_pos=zeros(1,dim); % initialise the position of delta wolf
Delta_score=inf; %initialise value of objective function for delta wolf, change this to -inf for maximization problems

Positions=initialization(SearchAgents_no,dim,ub,lb);

Convergence_curve=zeros(1,Max_iteration);

l=0; % loop counter

while l<Max_iteration   % Loop over the number of iterations
     for i=1:size(Positions,1)   %loop over every wolf

          % If the location exceeds the search space,need to return to the search space
          Flag4ub=Positions(i,:)>ub;
          Flag4lb=Positions(i,:)<lb;
          % If the position of Wolf is between the maximum and the minimum value, no need to adjust the position. If the position exceeds the maximum value, return to the maximum boundary
          % If the position exceeds the minimum value, return to the minimum boundary

          Positions(i,:)=(Positions(i,:).*(~(Flag4ub+Flag4lb)))+ub.*Flag4ub+lb.*Flag4lb;
          %~refers to invert

          % Calculate the fitness function value
          cmd = Positions(i,1);
          net=newgrnn(p_train,t_train,cmd);
          %predict output
          prediction_result=sim(net,p_test);
          %predict error
          grnn_error=t_test-prediction_result;

          fitness=mse(grnn_error);
          % MSE was taken as the objective function value of optimisation

          if fitness<Alpha_score
          % If the objective function is less than the alpha wolf's objective function
               Alpha_score=fitness;

```
            % Then the alpha wolf objective function value is updated to the optimal objective
function value
            Alpha_pos=Positions(i,:);
            % meanwhile update alpha wolf location to optimal location
        end


        if fitness>Alpha_score && fitness<Beta_score
        % If the value of the objective function is between alpha and beta wolves
            Beta_score=fitness;
            % Then update the Beta Wolf objective function value to the optimal objective
            function value
            Beta_pos=Positions(i,:);
            % meanwhile update beta wolf location to optimal location
        end


        if fitness>Alpha_score && fitness>Beta_score && fitness<Delta_score
        %if the objective function value is between the beta wolf and delta wolf objective
        function value
            Delta_score=fitness;
            % Then update the delta wolf objective function value to the optimal objective
            function value
            Delta_pos=Positions(i,:); % meanwhile update delta wolf location to optimal
            location
        end
    end


    a=2-l*((2)/Max_iteration);
    %for each iteration, the corresponding a value is calculated,a decreases linearly fron 2 to 0


    for i=1:size(Positions,1) % loop over every wolf
        for j=1:size(Positions,2) % loop over every dimension


            % Surround prey and update position


            r1=rand(); % r1 is a random number in [0,1]
            r2=rand(); % r2 is a random number in [0,1]


            A1=2*a*r1-a; %calculated coefficient A, Equation (3.3)
            C1=2*r2; % calculated coefficient C, Equation (3.4)


            % alpha wolf position update
            D_alpha=abs(C1*Alpha_pos(j)-Positions(i,j)); % Equation (3.5)-part 1
            X1=Alpha_pos(j)-A1*D_alpha; % Equation (3.6)-part 1
```

```matlab
            r1=rand();
            r2=rand();

            A2=2*a*r1-a; % calculated coefficient A, Equation (3.3)
            C2=2*r2; % calculated coefficient C, Equation (3.4)

            % Beta wolf position update
            D_beta=abs(C2*Beta_pos(j)-Positions(i,j)); % Equation (3.5)-part 2
            X2=Beta_pos(j)-A2*D_beta; % Equation (3.6)-part 2

            r1=rand();
            r2=rand();

            A3=2*a*r1-a; % calculated coefficient A, Equation (3.3)
            C3=2*r2; % calculated coefficient C, Equation (3.4)

            % Delta wolf position update
            D_delta=abs(C3*Delta_pos(j)-Positions(i,j)); % Equation (3.5)-part 3
            X3=Delta_pos(j)-A3*D_delta; % Equation (3.5)-part 3

            % position update
            Positions(i,j)=(X1+X2+X3)/3;% Equation (3.7)

        end
    end
    l=l+1;
    Convergence_curve(l)=Alpha_score;
end
% bestc=Alpha_pos(1,1);
% bestg=Alpha_pos(1,2);
bestp=Alpha_pos(1,1);
%% Print the result of parameter selection
disp(' Print the result of parameter selection');
str=sprintf('best parameter = %g',bestp);
disp(str)


%%adopt the best method to establish GRNN network
net=newgrnn(p_train,t_train,bestp);
% Perform predictive output on the test set
output_test_pre=sim(net,p_test);
% prediction error before reverse normalisation
eorry2=mse(t_test-output_test_pre);
% reverse normalise the predicted results
test_pre=mapminmax('reverse',output_test_pre,outputps);
```

% prediction error after reverse normalisation
err_pre=test_output-test_pre;

%% The predicted results are compared with the actual results
figure('Name','Acoustic Metamodel Prediction Accuracy Tested by 8 Samples')
subplot(5,1,1);
plot(test_pre(1,:),'*r-');hold on;plot(test_output(1,:),'bo-');
legend('Predicted SPLP10','Original SPLP10')

subplot(5,1,2);
plot(test_pre(2,:),'*r-');hold on;plot(test_output(2,:),'bo-');
legend('Predicted SPLP40','Original SPLP40')

subplot(5,1,3);
plot(test_pre(3,:),'*r-');hold on;plot(test_output(3,:),'bo-');
legend('Predicted SPLP70','Original SPLP70')

subplot(5,1,4);
plot(test_pre(4,:),'*r-');hold on;plot(test_output(4,:),'bo-');
legend('Predicted 100-SPLTHR(65)','Original 100-SPLTHR(65)')

subplot(5,1,5);
plot(test_pre(5,:),'*r-');hold on;plot(test_output(5,:),'bo-');
legend('Predicted SPLIQR','Original SPLIQR')

result=[test_output',test_pre'];
%% evaluation indicators for predictive data
MAE=mymae(test_output',test_pre');
MSE=mymse(test_output',test_pre');
MAPE=mymape(test_output',test_pre');

Meta-model training and testing for sunlight and thermal domain are the same as in acoustic field, expect for data loading and ploting arrangement. Will not repeat here.

II.   Initialisation of Grey Wolf Optimizer (GWO)

```
%_____
%    Grey Wolf Optimizer (GWO) source codes version 1.0
%
%    Developed in MATLAB R2011b(7.13)
%
%    Author and programmer: Seyedali Mirjalili
%
%    e-Mail: ali.mirjalili@gmail.com
%            seyedali.mirjalili@griffithuni.edu.au
%    Homepage: http://www.alimirjalili.com
%
%    Main paper: S. Mirjalili, S. M. Mirjalili, A. Lewis
%                Grey Wolf Optimizer, Advances in Engineering
%                Software , in press,
%                DOI: 10.1016/j.advengsoft.2013.12.007
%
%_____


% This function initialize the first population of search agents
function Positions=initialization(SearchAgents_no,dim,ub,lb)

Boundary_no= size(ub,2); % numnber of boundaries

% If the boundaries of all variables are equal and user enter a signle
% number for both ub and lb
if Boundary_no==1
    Positions=rand(SearchAgents_no,dim).*(ub-lb)+lb;
end

% If each variable has a different lb and ub
if Boundary_no>1
    for i=1:dim
        ub_i=ub(i);
        lb_i=lb(i);
        Positions(:,i)=rand(SearchAgents_no,1).*(ub_i-lb_i)+lb_i;
    end
end
```

Appendix C Matlab Program Script of Multi-Domain Multi-Objective Optimisation via NSGAII

I.   Multi-Domain Multi-Objective Optimisation via NSGAII

```
%
% Copyright (c) 2015, Yarpiz (www.yarpiz.com)
% All rights reserved. Please read the "license.txt" for license terms.
%
% Project Code: YPEA120
% Project Title: Non-dominated Sorting Genetic Algorithm II (NSGA-II)
% Publisher: Yarpiz (www.yarpiz.com)
%
% Developer: S. Mostapha Kalami Heris (Member of Yarpiz Team)
%
% Contact Info: sm.kalami@gmail.com, info@yarpiz.com
%


clc;
clear;
close all;
```

%% Problem Definition

```
CostFunction=@(x) MOP6(x);          % Cost Function

nVar =20;       % Number of Decision Variables
% Size of Decision Variables Matrix
VarMin1=-1;VarMin2=-1;VarMin3=-1;VarMin4=-1;VarMin5=-1;
VarMin6=-1;VarMin7=-1;VarMin8=-1;VarMin9=-1;VarMin10=-1;
VarMin11=-1;VarMin12=-1;VarMin13=-1;VarMin14=-1; VarMin15=-1;
VarMin16=-1;VarMin17=-1;VarMin18=-1;VarMin19=-1;VarMin20=0.555555555555556;

VarMax1=1;VarMax2=1;VarMax3=1;VarMax4=1;VarMax5=1
VarMax6=1;VarMax7=1;VarMax8=1;VarMax9=1;VarMax10=1;
VarMax11=1;VarMax12=1;VarMax13=1;VarMax14=1;VarMax15=1;
VarMax16=1;VarMax17=1;VarMax18=1;VarMax19=1;VarMax20=1;

VarMin=[VarMin1,VarMin2,VarMin3,VarMin4,VarMin5,VarMin6,VarMin7,VarMin8,VarMin9,Va
rMin10,VarMin11,VarMin12,VarMin13,VarMin14,VarMin15,VarMin16,VarMin17,VarMin18,Var
Min19,VarMin20];
VarMax=[VarMax1,VarMax2,VarMax3,VarMax4,VarMax5,VarMax6,VarMax7,VarMax8,VarMax
```

9,VarMax10,VarMax11,VarMax12,VarMax13,VarMax14,VarMax15,VarMax16,VarMax17,VarMax18,VarMax19,VarMax20];

%% NSGA-II Parameters

MaxIt=400;   % Maximum Number of Iterations

nPop=400;   % Population Size

pCrossover=0.9;   % Crossover Percentage
nCrossover=2*round(pCrossover*nPop/2);   % Number of Parents (Offsprings)

pMutation=0.4;   % Mutation Percentage
nMutation=round(pMutation*nPop);   % Number of Mutants

mu=0.02;   % Mutation Rate

sigma=1;   % Mutation Step Size

%% Initialization

```
empty_individual.Position=[];
empty_individual.Cost=[];
empty_individual.Rank=[];
empty_individual.DominationSet=[];
empty_individual.DominatedCount=[];
empty_individual.CrowdingDistance=[];

pop=repmat(empty_individual,nPop,1);

for i=1:nPop

    pop(i).Position=(VarMax-VarMin)*rand+VarMin;

    pop(i).Cost=CostFunction(pop(i).Position);

end

% Non-Dominated Sorting
[pop, F]=NonDominatedSorting(pop);

% Calculate Crowding Distance
pop=CalcCrowdingDistance(pop,F);
```

% Sort Population
[pop, F]=SortPopulation(pop);

%% NSGA-II Main Loop

for it=1:MaxIt

    % Crossover
    popc=repmat(empty_individual,nCrossover/2,2);
    for k=1:nCrossover/2

        i1=randi([1 nPop]);
        p1=pop(i1);

        i2=randi([1 nPop]);
        p2=pop(i2);

        [popc(k,1).Position,
popc(k,2).Position]=Crossover(p1.Position,p2.Position,VarMin,VarMax);

        popc(k,1).Cost=CostFunction(popc(k,1).Position);
        popc(k,2).Cost=CostFunction(popc(k,2).Position);

    end
    popc=popc(:);

    % Mutation
    popm=repmat(empty_individual,nMutation,1);
    for k=1:nMutation

        i=randi([1 nPop]);
        p=pop(i);

        popm(k).Position=Mutate(p.Position,mu,sigma,VarMin,VarMax);

        popm(k).Cost=CostFunction(popm(k).Position);

    end

    % Merge
    pop=[pop
        popc

```matlab
        popm]; %#ok

    % Non-Dominated Sorting
    [pop, F]=NonDominatedSorting(pop);

    % Calculate Crowding Distance
    pop=CalcCrowdingDistance(pop,F);

    % Sort Population
    pop=SortPopulation(pop);

    % Truncate
    pop=pop(1:nPop);

    % Non-Dominated Sorting
    [pop, F]=NonDominatedSorting(pop);

    % Calculate Crowding Distance
    pop=CalcCrowdingDistance(pop,F);

    % Sort Population
    [pop, F]=SortPopulation(pop);

    % Store F1
    F1=pop(F{1});

    % Show Iteration Information
    disp(['Iteration ' num2str(it) ': Number of F1 Members = ' num2str(numel(F1))]);

    % Plot F1 Costs
    figure(1);
    PlotCosts(F1);
    pause(0.01);

end

%% Results
```

II.   Cost Function

```matlab
function z=MOP6(x)

    z1=fun5(x);
    z2=fun6(x);
    z3=fun7(x);
    z=[z1 z2 z3]';

end
```

%%% definition of fun5(x)

```matlab
function fitness = fun5(x)
% calculation of fitness value of individual
% x              input       individual
% fitness        output      fitness value of individual

%load acoustic meta-model
load netg_sheng

%prediction output via network
x=x';
%inputn_test=mapminmax('apply',x,inputs);
%x=inputn_test;
an=sim(net,x);
a2=0.2*an(4,1)+0.2*an(1,1)+0.2*an(2,1)+0.2*an(3,1)+0.2*an(5,1);
fitness=a2;
```

%%% definition of fun6(x)

```matlab
function fitness = fun6(x)
% calculation of fitness value of individual
% x              input       individual
% fitness        output      fitness value of individual

% load sunlight meta-model
load netg_guang

%prediction output via network
x=x';
%inputn_test=mapminmax('apply',x,inputs);
%x=inputn_test;
an=sim(net,x);
```

a3=(an(1,1)+an(2,1)+an(3,1)+an(4,1)+an(5,1)+an(6,1)+an(7,1)+an(8,1)+an(9,1)+an(10,1)+an(11,1)+an(12,1)+an(13,1))/13;
fitness=a3;

%% definition of fun7(x)
function fitness = fun7(x)
% calculation of fitness value of individual
% x   input   individual
% fitness   output   fitness value of individual

% load thermal meta-model
load netg_re

%prediction output via network
x=x';
%inputn_test=mapminmax('apply',x,inputs);
%x=inputn_test;
an=sim(net,x);
a4=0.2*an(1,1)+0.2*an(2,1)+0.2*an(3,1)+0.2*an(4,1)+0.2*an(5,1);
fitness=a4;

Appendix D

Qualitative Analysis Process of Acoustic Simulation Results

I.   Rule and Reason Lying in Noise Maps and Curves of Very Good Performance Cluster

As seen in the dendrogram of clustering analysis, cluster D is the best performed group, with SPL attenuation over 30-40 dBA dropping from P90 around 75 dBA to P10 around 45 dBA.

The curves of cases from cluster D appear in a gradually increasing tendency from minimum to maximum with homogeneous gradient (Figure D1). The turning points of slightly gradient changes are at P30 and P70-P75 where indicating the critical data dividing quiet zone, acceptable zone and noisy zone on site grid.

Figure D1 Histogram (up-left), Curve (up-right), and Noise Map (down-left) of Sample from Very Good Performance Group (Site 1801)

The histograms of cluster D appear in bimodal shape of low peak locating around 45-55 dBA, and valley bottom at 65-75 dBA and high peak at 75-85 dBA (Figure D1). The height of low peak is below 9%, taller than which of high peak of below 6%. The slopes on either side of the two peaks are flat and wide. This echoes with the gradually increase curve of homogenous changing gradient. The thick tails on both sides of the low peak indicate the high proportion of low noise level zone. The clear and deep valley bottom reflects the noise barrier effects by the street facing buildings. The height of valley bottom is as low as 1.5%, compared to other clusters' histogram.

Analysis of the noise map and building distribution pattern could explain the noise data pattern aforementioned. The street facing buildings are arranged close to the street, namely having small separation distance from sound sources. This exposes the front facade of street facing buildings in high level traffic noise which may require extra acoustic seal and barrier to assure indoor acoustic performance. However, if the small distance of front facade and sound source combined with closed site boundary of long building facades, the acoustic performance inside a site dramatically improved due to the protection from sacrificed facades. This drop from 75dBA to 70dBA is shown by the thick left tail of the high peak.

Sites in cluster D share high building density and relatively long facades. These characteristics enhance the noise attenuation at inner area. Noise level drop due to this reason is from 60dBA to 40dBA, represented by the two thick tails of the low peak in histogram.

It could be said that the best performance cluster is contributed by 1) highly enclosed site boundary and small separation distance to road for dramatic noise attenuation from 75dBA to 70 dBA; 2) high building density and facade length for potent noise attenuation at inner area of the site.

II.   Rule and Reason Lying in Noise Maps and Curves of Good Performance Cluster

As seen in dendrogram of clustering analysis, cluster B is the group of good acoustic performance. This group presents SPL attenuation of 20-30 dBA and ranges in 55-75 dBA. Comparingly, cluster B provides less noise attenuation over the site to cluster D, yet present quieter (5dBA) acoustic condition to cluster A under same noise attenuation level.

The curve of cluster B appears less steep than the curve of cluster D. The turning points are P40 and P70. Seen from the curve there are two sub-clusters of sample exist in cluster B. The sub-cluster B1 is relative noisy site (maximum sound source level at 80dBA) with considerable noise attenuation ability (as low to 40-50dBA) (e.g. S10, S24). These sites show noise barrier effect at P70 value, and have noise attenuation from P40 to P10 due to building arrangement. The sub-cluster B2 is relatively less noisy site (sound source level at 75dBA) with satisfactory amount of noise attenuation (e.g. S1501, S20, S260401) and high proportion of area with satisfactory noise level (55-65dBA). These site has some ability of attenuation below P75 due to the enclosure at boundary, but not as strong as the highly enclosed boundary shows in cluster D. However, noise attenuation from P40 to P10 is very limited for these sites due to the building configuration of low-medium density and large separation distance between buildings.

The histogram of cluster B appears obviously narrower than the spread appearance of cluster D's histogram. For good performance cluster B, the low peak locates at 55-65dBA, higher than 50dBA in cluster D (very good performance). Furthermore, the tails on either side of low peak is thin and short, which refers to a concentration of noise level between 55-65dBA shown as orange area in noise map. This is especially obvious in type B2 site with lower density and larger separation, that the low peak is especially pointy with very thin tails compared to type B1.
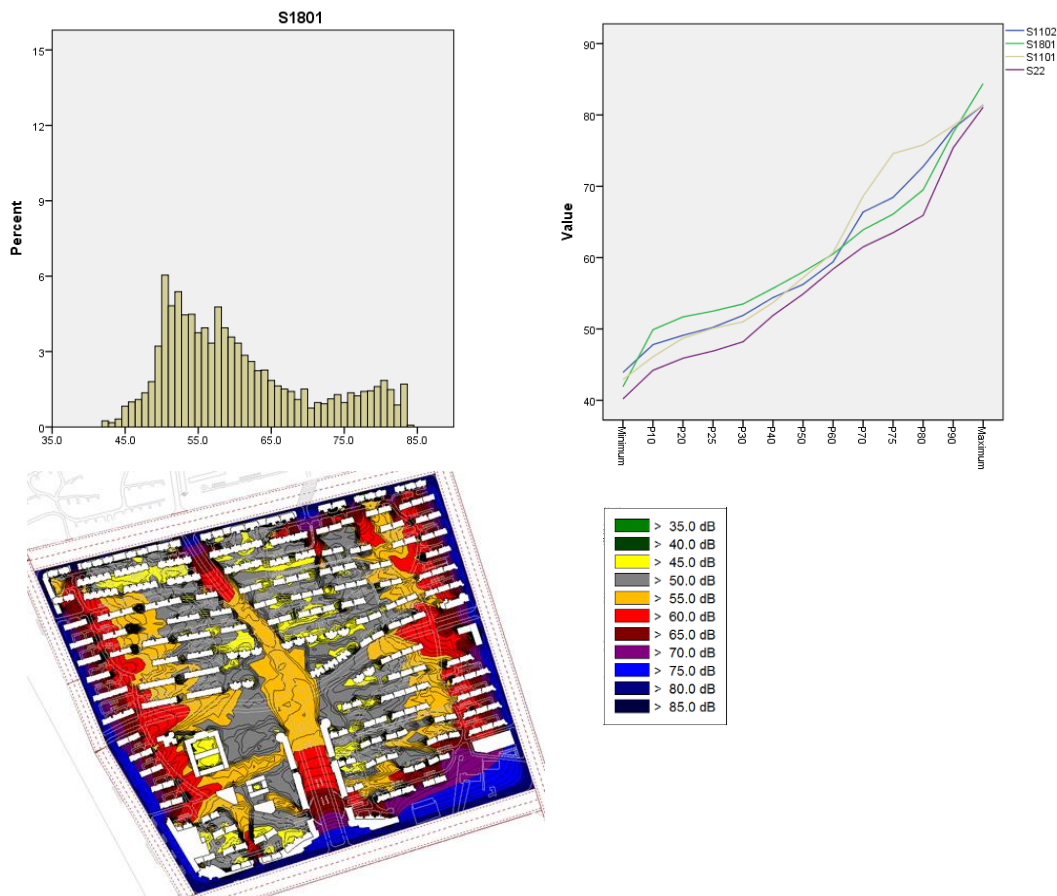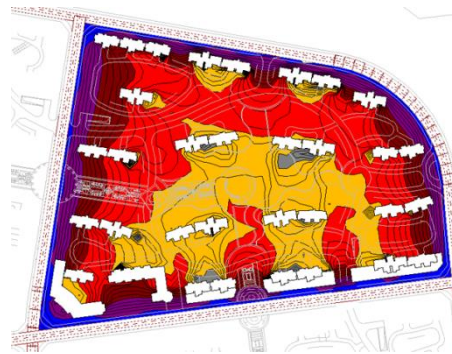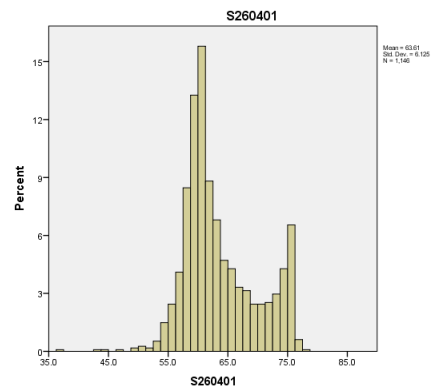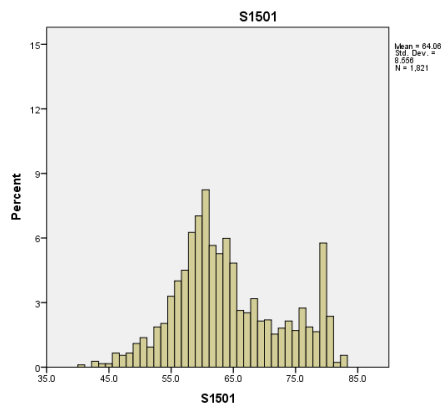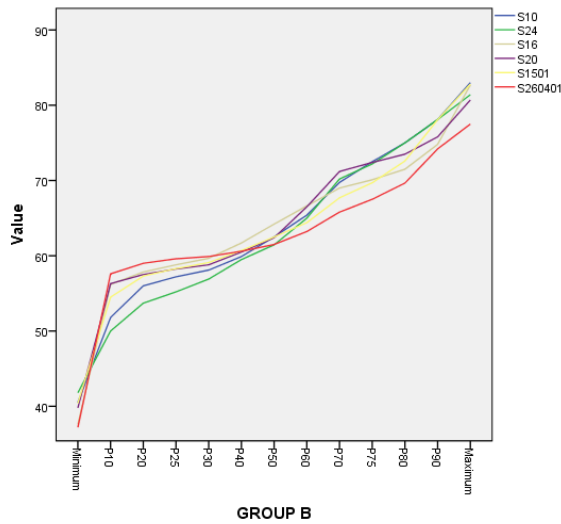
Figure D2 Histogram (up-left), Curve (up-right), and Noise Map (down-left) of Sample from Good

Performance Group (Site 1501 and Site 260401)

Noise maps of cluster B presents large proportion of area in orange (over 55dBA) and red (over 60dBA), but less area with noise level below 55dBA. This echoes with the narrow low value peak on histogram. Relatively long continuous facade, large building separation and low-medium density is the overall building distribution characteristics in inner area of the site. These characters lead to limited power of noise

attenuation in inner area, therefore the few areas have noise level below 55dBA and there exists high proportion of area in orange to red (55-65dBA).

The valley bottom of histogram locates at 65-75dBA, same as for cluster D. However, the height of the bottom is approximately 3%, almost twice as which for cluster D. It could be explained by the medium level boundary enclosure. The wide intervals at boundary allows noise propagating inside the site without barrier, generating inner-forward noise contour in dark red (65-75dBA) area on the map and high peak centre around 70dBA on the histogram.

The high value peak locates at 75-80dBA with thin tails on both sides. Compared to sites in cluster D, the sound source level of sites from cluster D could even be higher (high peak at 80dBA) than which of cluster B. In other words, the inner acoustic performance could be strongly improved with building arrangement, although sound source is still highly influential.

To summarise the analysis based on cluster B, it is said that:

1) P70 represents noise drop due to barrier effect of the street facing buildings, P10-P40 represents the noise attenuation ability of inner building distribution.

2) large proportion of area in satisfactory noise level (55-65dBA) could also achieve general good acoustic performance if quiet level (below 55dBA) not achievable in certain building distribution condition.

3) for inner noise attenuation, long continuous building facade is positive factor while large building separation and low-medium density is negative factor. When combined together they lead to satisfactory inner noise attenuation result.

4) less boundary enclosure results in taller valley bottom appeared on histogram of grid SPL data.

5) the inner acoustic performance could be strongly improved with building arrangement, although sound source is still highly influential.

III. **Rule and Reason Lying in Noise Maps and Curves of Fair Performance**

Cluster

Cluster A includes the majority of cases, with a fair acoustic performance. Its capable noise attenuation is around 20-30 dBA, and overall noise level ranges between 60s-80s dBA. Compared to cluster D and B, cluster A clearly lack of noise blocking ability, that under similar sound source intensity P10 value representing quiet area barely achieves noise level below 60dBA. Cluster A data distributes on histogram as narrow spread and pointy peaks.

There exist four sub-groups in cluster A, divided in cluster procedure based on slight differences of attenuation ability and data range.
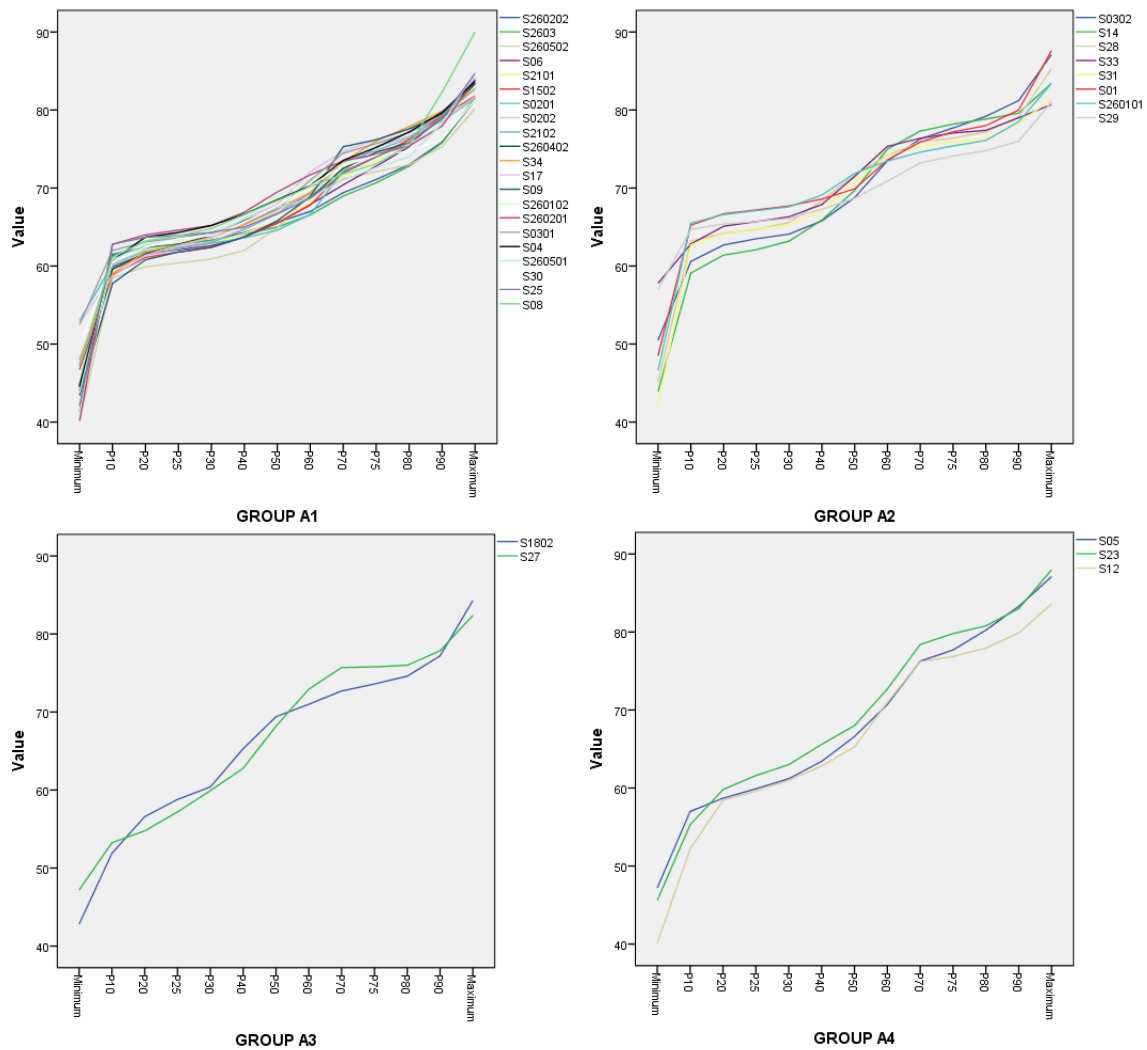
Figure D3 Curve Comparisons of Four Sub-Groups of Fair Performance Cluster A

Sub-cluster A1

Sub-cluster A1 includes the majority of samples and could be divided into four types of conditions. The ranges of various types are summarised as: type 1 of attenuation from 80 to 60dBA (normal), type 3 from 83 to 58dBA (strengthened), type 2 from 80 to 65dBA (weakened and noisy), and type 4 from 75 to 60dBA (weakened and less noisy).
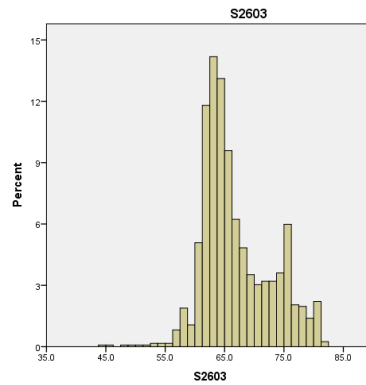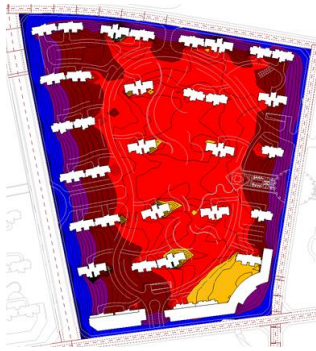
Type 1 (S2603)
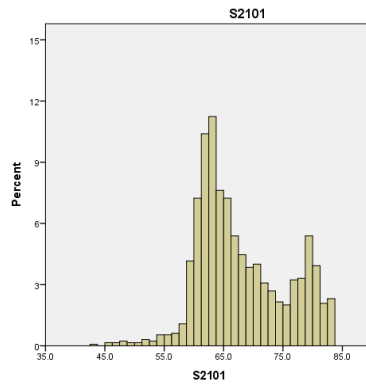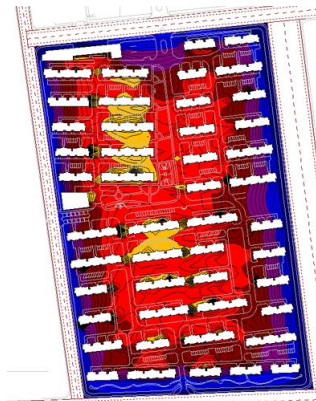
Type 2 (S2101)

Type1 shares key building distribution characteristics of very small distance to road, normal source level, but lack of noise barrier ability on-edge and in-site due to open boundary and low density. These configurations appear on histograms as short high peak (located at 75dBA), considerable height of valley bottom (at 70-75dBA), but very tall and narrow low peak (around 60dBA). Seen on noise maps, small dark blue area hinted less noisy sound source and its immediate encountering of street facing facades result in short high peak on histogram. However, the lack of enclosure at boundary especially on east-west sides allows noise to entre inner site, causing valley

bottom as high as 3% on histogram. Low inner density or large building separation distance results in almost nil noise attenuation in-site, leading to high proportion of area in 60-65dBA marked as red in noise map and extremely dramatic drop on left tail of low peak.

Type2 is similar to type 1 on histogram only has shorter valley bottom height, namely slight higher noise barrier effect at site boundary. Its building configurations are open boundary and medium density of medium rise buildings. It shows that slightly increase in density or reduction in building separation distance could lead to improvement in noise attenuation both on-edge and in-site. However the power is still very limited that the low peak locates still around 60dBA with lower height at 10% and slightly thicker left tail of low peak.

Type3 has similar histogram to type1 and type 2, while the differences are its enclosed boundary and very small site. It indicates that although enclosure could help in noise attenuation but would be limited by site size. The valley bottom height is around 1-1.5%, shorter than which of type 1 and 2 due to effect of enclosed boundary.

Type4 shows quite different distribution characteristics on noise map: at least one large distance to road, open boundary, large separation distance between buildings, and low-medium density. This distribution means lack of noise attenuation ability both on-edge and in-side of the site. Therefore, this shows on histogram very differently compared to other types. A taller high peak locates at 75dBA, where locates valley bottom in other types. This is because the large distance between road and street facing buildings allowing noise to propagate freely in before meeting front row buildings, where noise level attenuate to 70-75dBA show as purple on the map. Valley bottom appears at 60-65dBA in height of 5% which is significantly taller than all other clusters and types. It is the result of large area available for noise to propagate and gradually reduced from 65 to 60dBA, due to wide distance between building façade and road and wide intervals between building facades. Pointy low

peak indicates a concentration in red area (60-65dBA). There barely exists area with value below 60dBA due to lack of inner noise attenuation ability. Worth mentioning, S0301 performs perfectly in thermal and sunlight performance is listed in fair performance for acoustic performance. The reasons are its large separation to road, low density, short continuous façade length and relative open boundary, namely very lack of noise attenuation ability both on-edge and in-site.
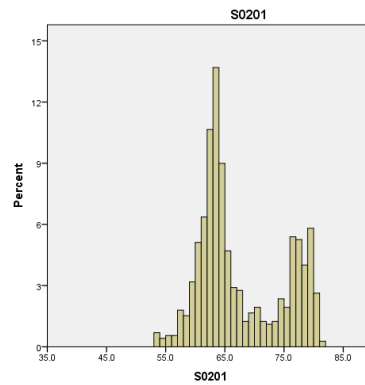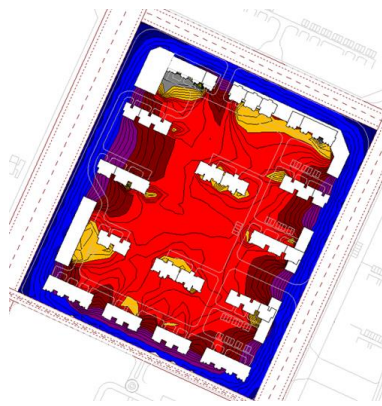
Type 3 (S0201)
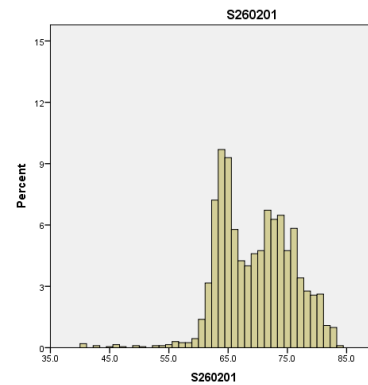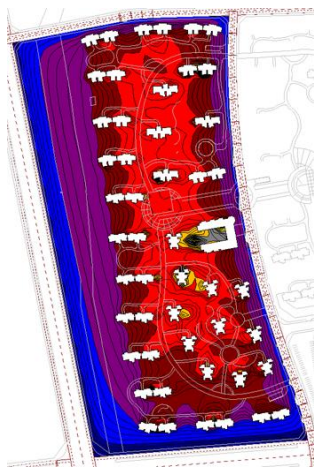
Type 4 (S260201)

Figure D4 Histogram (Up-Left) and Noise Map (Down-Left) of Four Types of Samples from

Sub-Cluster A1 of Fair Performance Group

To conclude the analysis about four types of condition in sub-group A1, it could be said:

1) Type 2 compare to type 1, high density, shorter building separation distance would

be good for inner noise attenuation ability;

2) Enclosure works well for noise attenuation, but small size would weaken its power;

3) Set-back distance from street facing buildings to road may not be contributive for inner area acoustic performance; it only avoids noise stress of street facing facades. For better inner acoustic performance, enclosed site boundary with buildings however sacrificing street facing facade being exposed in strong noise, could be more effective than approach of setting-back all buildings from road.

4) short in height of valley bottom in histogram indicates thorough noise barrier effect at street facing buildings.

Sub-cluster A2

The curve and histogram of sub-cluster A2 is very close to sub-cluster A1, especially type 4 in sub A1, however the differences are generally noisier of A2 with large proportion of area in 65-70dBA compared to 60-65dBA for A1-type 2.

Building distribution of sub-cluster A2 leads to lack of noise barrier effect both at site boundary and inside the site. The characteristics are listed as: strong sound source level, low density, large distance between buildings, medium-large set-back distance from road and relative open boundary.

To conclude the result of sub-cluster A2, it could be said:

1) sound source is very influential on overall noise level if boundary not highly enclosed;

2) set-back distance is not contributive for inner acoustic performance if no front façade barrier existed
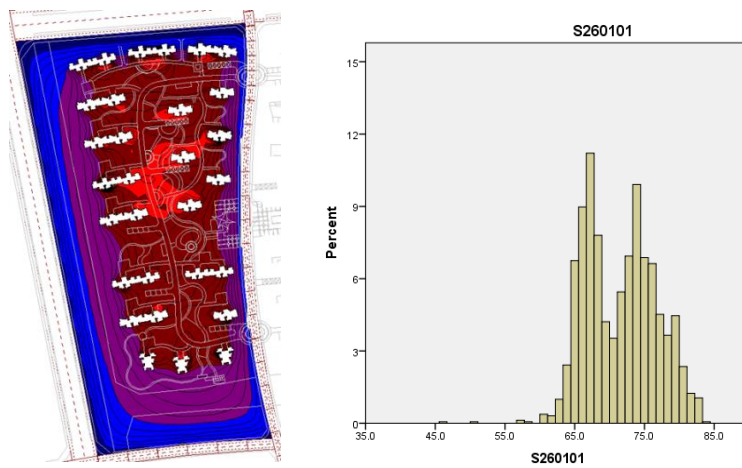
Figure D5 Histogram and Noise Map of Samples from Sub-Cluster A2

of Fair Performance Group (Site 260101)

Sub-cluster A3

Sub-cluster A3 is actually performed well at inner acoustic environment, with low level below 40dBA and considerable amount of area below 65dBA. However, wide area of 70-75dBA (marked as purple) between front row buildings and road decreases the average SPL overall site. Therefore, sub-cluster A3 is considered as fair level.
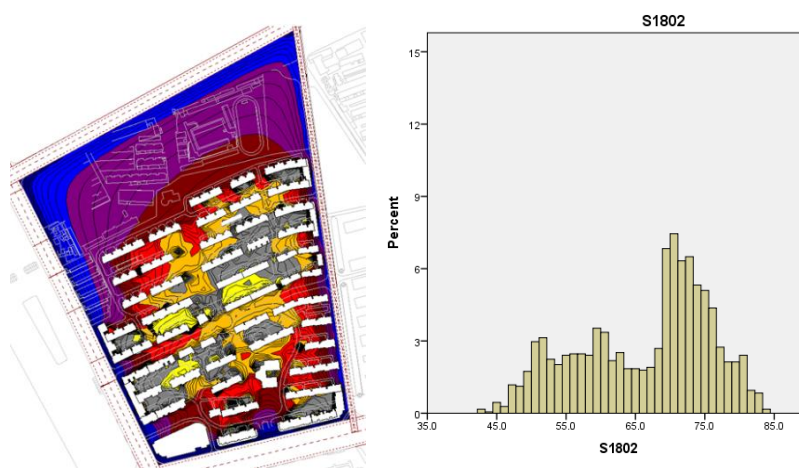
Figure D6 Histogram and Noise Map of Samples from Sub-Cluster A3

of Fair Performance Group (Site 1802)

As seen in the histogram of A3, the large distance between front row buildings and road provides spaces for noise propagating freely to decrease from 80dBA to 70dBA,

which forms the thick and long right tail of the high peak. The location of high peak is at 70dBA where traffic noise meets front row facades and suddenly drops to 65dBA. This appears as extremely thin left tail or dramatic drop on the left of high peak, as well as lack of dark red area (65-70dBA) on noise map.

The valley bottom moves left to 656dBA compared to other sub-clusters and has a low height at 3%, appearing on curve as P50 becoming the turning points instead of P70. It is as a result of the wide distance to road by which the noise meeting front row façade has already attenuated to 70dBA.

The low peak in histogram has a flat and spread appearance as in cluster D. Enclosed boundary, long continuous building façade both on edge and in-side of the site and high density contribute to this quiet in-site environment.

To conclude:

1) Large distance to road to allow noise level to drop to 65-70 when meeting front row building facades

2) when combining large distance to road with enclosed boundary, it would lead to very good inner attenuation without exposure front façade in strong noise, which could be the best approach of noise attenuation without considering economic reasons.

3) for site with small and large separation distance between buildings and roads, the turning point on curve indicating noise barrier effect at street facing façade is P70 and P50 respectively.

Sub-cluster A4

Sub-cluster A4 shows the protection effect of half enclosed boundary on a building

cluster lack of inner noise attenuation ability. Compared to sub-cluster A2, A4 has normal separation distance to road, but stronger source level. Both sub-clusters has building group lack of inner noise attenuation ability, but A4 performs better shown as flatter left tail, and left skewed location of low peak. It shows on curve of A4 as larger gradient of curve below P70, due to front façade effect. It could be concluded that enclosed boundary is very helpful for inner acoustic performance of medium density site lack of inner noise blocking ability.
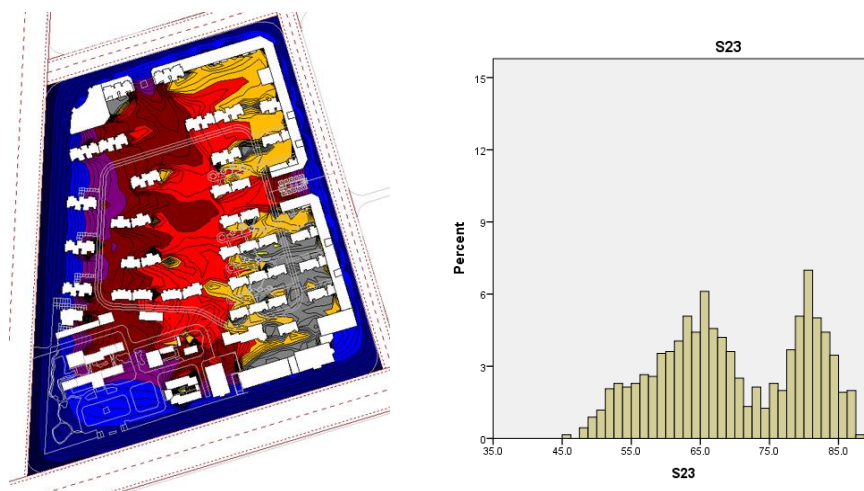
Figure D7 Histogram and Noise Map of Samples from Sub-Cluster A4

of Fair Performance Group (Site 23)

In a word, the conclusions of this section based on analysis on fair performance cluster are:

1) short in height of valley bottom in histogram indicates thorough noise barrier effect at street facing buildings.

2) for site with small and large separation distance between buildings and roads, the turning point on curve indicating noise barrier effect at street facing façade is P70 and P50 respectively.

3) sound source is very influential on overall noise level if boundary not highly enclosed;

4) to improve overall attenuation, enclosure at boundary is especially powerful, but

small site size would weaken its power.

5) to improve inner acoustic performance, high density, short separation distance between buildings, long continuous facade are contributive.

6) Set-back distance from street facing buildings to road may not be contributive for inner area acoustic performance, if no front façade barrier existed.

7) best approach for overall quietness: combining large distance to road with enclosed boundary, it would lead to very good inner attenuation without exposure front façade in strong noise, which could be the best approach of noise attenuation without considering economic reasons.

8) best approach for inner quietness: enclosed site boundary by buildings however sacrificing street facing facade being exposed in strong noise, could be significantly effective than approach of setting-back all buildings from road.

IV. Rule and Reason Lying in Noise Maps and Curves of Poor Performance Cluster

Cluster C performs worst in noise attenuation which is less than 20 dBA. The range is 70-85dBA, as seen on its curve. The gradient of the curve is flat and turning points of P40 and P70 are not as obvious as in other clusters.
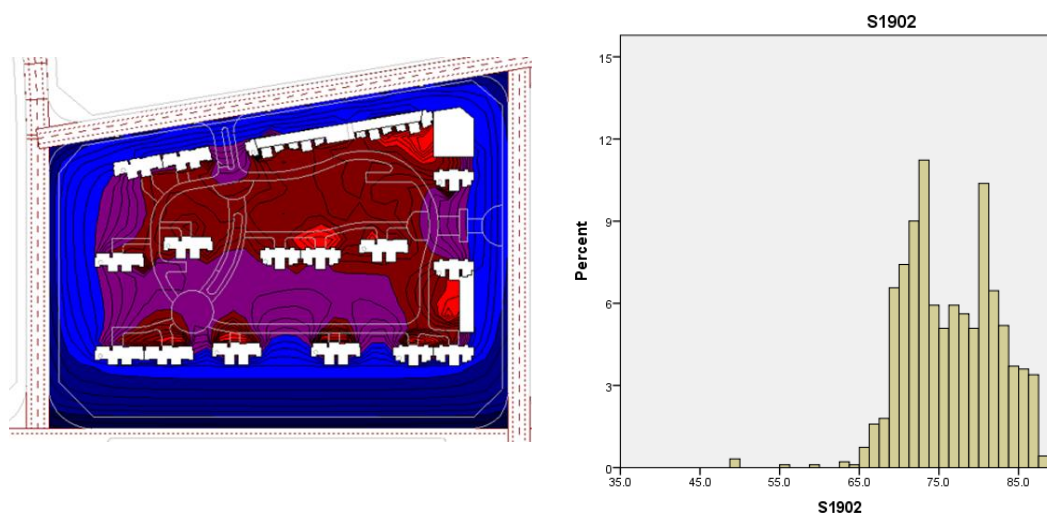
Figure D8 Histogram and Noise Map of Samples from Poor Performance Group (Site 1902)

The histogram of cluster C shows right skewed pattern in general: low peak located around 70dBA, bottom valley at 75dBA and high peak at 80dBA. This could be interpreted as noisy over the site. Echoed on the noise map, the sound source level is relatively stronger than in other clusters which result in thick right tail of high peak. The less enclosed boundary contributes to the tall valley bottom. And low density and short continuous building facades as well as been small site, altogether lead to extreme lack of inner noise attenuation ability.

It could be said that the distribution pattern in cluster C is worth avoiding from acoustic performance point of view: small site of less enclosed boundary and low density, especially when exposed to strong sound source.

Appendix E

Qualitative Analysis Process of Sunlight Simulation Results

The cluster result of APSH and WPSH, and comparison and explanation of difference between two cluster results is expounded.

I.   Analysis of Clustering of Sites and Discovery of Rules by APSH

In this section, cluster output based on statistical measure (P10-P90) of APSH is analysed. Characteristics of each cluster group is discussed and explained based on sunlight hour maps and sunlight hour data of sites in a group. After that, Qualitative tendencies and rules between cluster characteristics and corresponding building distribution is discovered and discussed.

Based on dendrogram of clustering by P10-P90 of APSH and physical meaning of each cluster by characteristics shown in sunlight maps, four clusters are ascertained: cluster A of very good sunlight possibility, B of good performance, C of fair performance and D of poor condition.

i.   Rules and Reasons in APSH Maps and Histograms of Very Good Performance Cluster A

Seen from APSH maps, cluster A has best sunlight possibility with majority area covered in yellow. There are two sub-groups in the cluster.

Compared to other clusters, histogram of cluster A has no minimum and left peak, even barely has a hump in the left tail at position of left peak (300h). This is a reflection of very few shadow area shown as blue area in APSH maps (Figure E1).
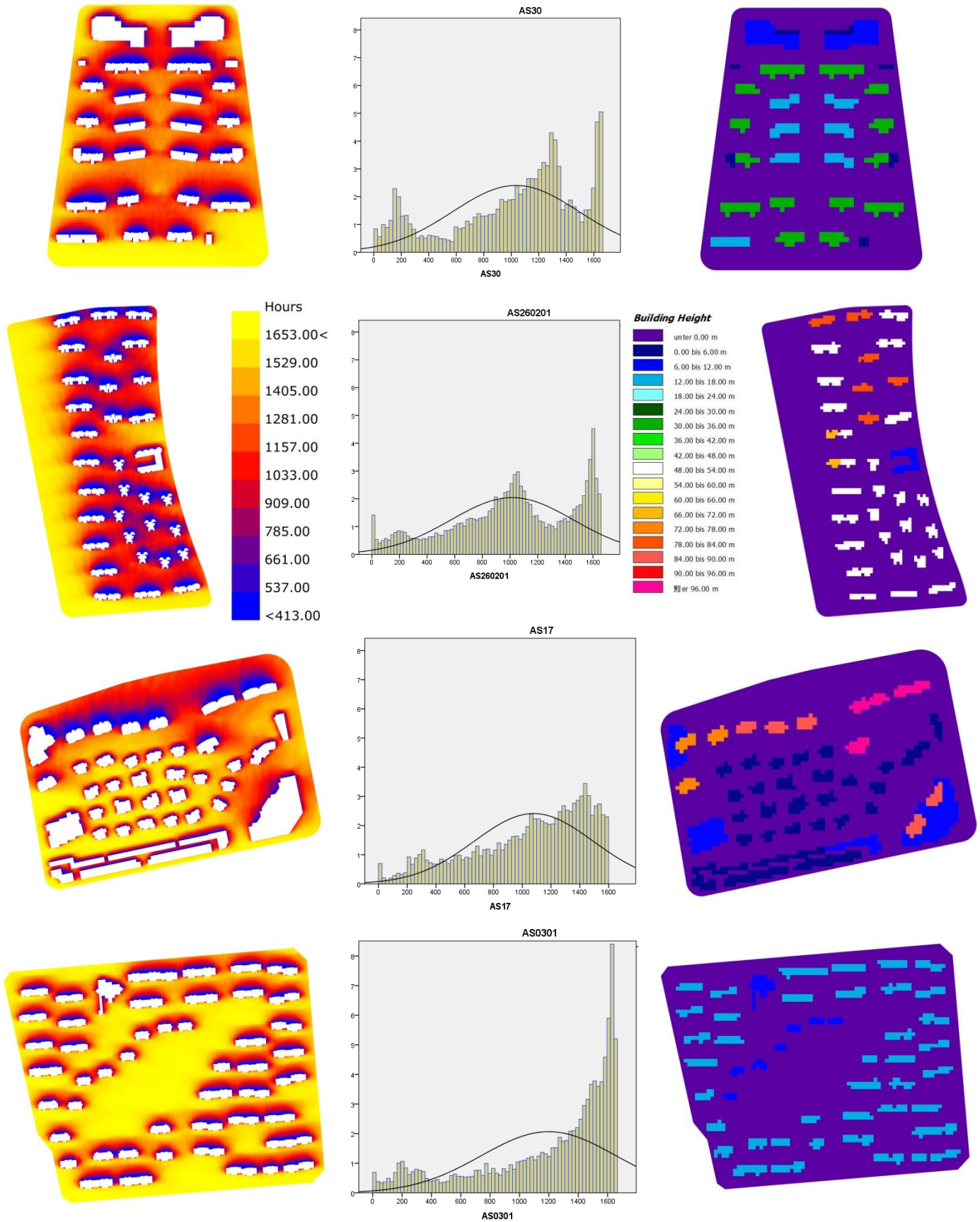
Figure E1 APSH maps and histograms of example sites from cluster A: S30, S260201 and S17 from A1, S0301 from A2

A1 shows 3 peaks in their histograms: the minimum peak reflecting area covered by heavy shadow, maximum peak reflecting unblocked area and a right-skewed peak reflecting relatively long sunlit area. As a comparison, A2 has only one maximum peak, and a long and thin left tail, indicating the large ratio of unblocked and very small ratio of partly blocked and shaded area. A2 is the sub-group of highest level sunlight hour.

Within sub-group A1, three types of building distribution exist, seen from APSH maps in figure E1: 1. low and mid-rise community of medium density with some open space on the edge of the site; 2. high-rise community of low density with large open space on the edge of the site; 3. mix of loose distributed low-rise and high-rise buildings. Three example cases S30, S260201 and S17 are used as representatives respectively for 3 types of distribution.

Comparing S30 and S260201, namely comparing low/mid-rise community and high-rise community, they both have tall maximum peak which is taller than the other two peaks. This is an indication of the existence of open spaces in the sites. The open space locates on the south of site 30 and on the west of site 260201, shown in their APSH maps in figure E1. The right skewed peak in the middle indicates the sunlight condition of partly shaded area. The location of the peak axis is 1300h for S30 while 1050h for S260201, which means in the large partly shaded area S30 generally has higher sunlight hour than S260201 (Table E1). The location and height of the minimum peak is similar between S30 and S260201. The valley between middle peak and minimum peak locates around 400-600h. It can be said that based on comparison of representative sites of low/mid-rise and high-rise community with certain open spaces on the edge of site, the low/mid-rise community provides higher overall sunlight hour than high-rise community, especially obvious in partly shaded area.

Table E1 histogram analysis of sample site in cluster A

Cluster	A1	A1	A1	A2
Example	S30	S260201	S17	S0301
Explanation	Low-Rise, Medium Density, Open Space	High-Rise, Medium Density, Open Space	Mix-Rise, Low Density	Low/Mid-Rise, Low Density
Max Peak Location	1653	1600	1420	1653
Middle Peak Location	1300	1050	N/A	N/A
Min Peak Location	150	200	N/A	N/A
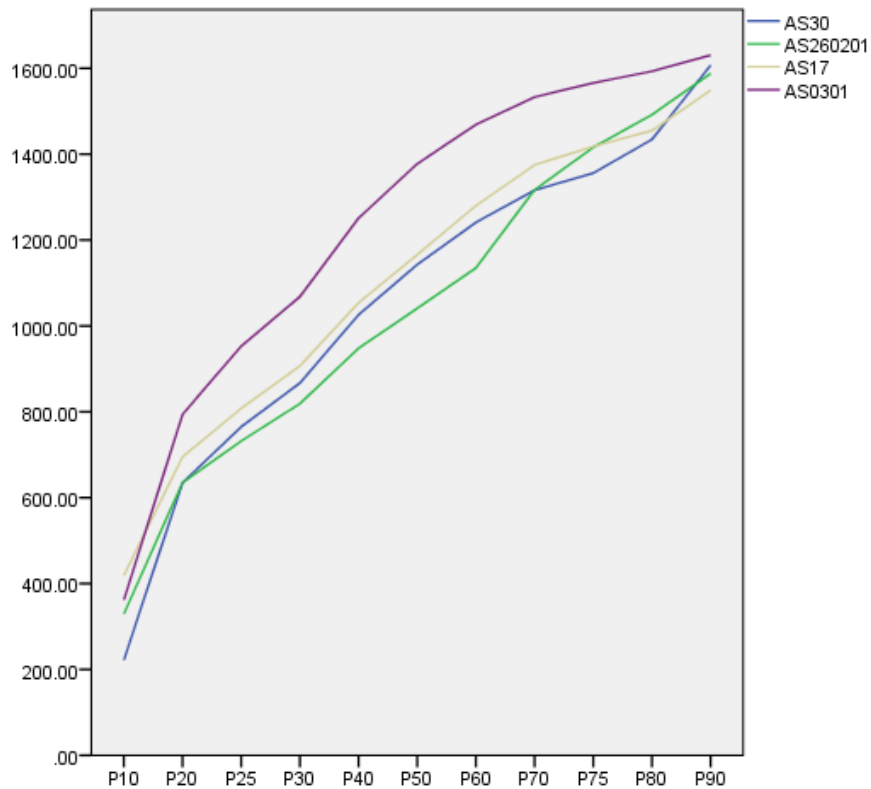
S17 is a mix of low/mid-rise and high-rise community. Seen in building height map, it has 78-90m tall buildings on the north and southeast corner of the site. The middle of the site is filled with very low density low-rise houses. This distribution provides opportunity of even and extended sunlight hour from the south to the middle of the site, but will have heavy shadow on the north due sky-scrapers on the north edge. The large ratio of low-rise community on the south makes the histogram of S17 to appear in single peak as in S0301. The only maximum peak locates at 1420h, left moved compared to 1653h in S0301(Table E1). Under the impact of sky-scrapers on the north, the left tail of S17 histogram is thicker than the left tail of S0301 histogram. The percentage of frequency between 400-600h in S17 histogram is twice as much as

which of S0301 histogram. This indicates that compared to low-rise community represented by S0301, mix-rise community still has less sunlight hours due to impact of high-rised part.

However, comparing S17 with S30 and S260201, S17's single peak is actually a merge of middle and maximum peak, meaning majority area has very long sunlight hour not far from maximum value. Also few areas are covered in heavy shadow due to large proportion of 4F houses and low density, S17 has no minimum peak. To summarise, compared to common low/mid-rise and high-rise community, extreme combination of very low and super high buildings in a low density approach with concern of tall buildings on the north, is still possible to generate very good sunlight environment. But the economic limit makes this combination difficult to generalise. It may only be appropriate for site condition of specific surroundings like natural landscape.

A2 is the sub-group of best sunlight condition. S0301represents cluster A2 of the low/mid-rise community with low density and large in-site open space performs longest overall sunlight hours. Majority of its APSH map is covered in yellow. Correspondingly, S0301 has only maximum peak and a very thin left tail. This indicates that the combination of low/mid-rise community in low density with open in-site space, namely low development intensity, is the most ideal attribution set for even and long accumulative sunlight hour year-round. However, in practice balance between economic requirement and low development intensity is still needed.

Based on the above analysis, some rules could be stated as below:
1. low development intensity including low/mid-rise community, low density and in-site open space, is the most ideal design combination for most extended APSH condition.

2. for low/mid-rise and high-rise community with open spaces on the edge of site, the low/mid-rise community provides higher overall sunlight hour than high-rise community, especially obvious in partly shaded area, shown as right skewed middle peak.

3. compared to common low/mid-rise and high-rise community, extreme combination of mix-rise community in a low density approach with concern of tall buildings on the north, is still possible to generate very good sunlight environment similar to but not as long APSH as low development intensity combination.

ii.   **Phenomenon and Reason in APSH Map and Histogram of Good Performance Cluster B**

Cluster B is defined as group of relatively good sunlight condition. Shown in its histograms (Figure E2), they have two peaks: the maximum peak and minimum peak. The valley bottom between two peaks is located at 400-600h. The two peaks are of similar height and width, but the maximum peak is still slightly higher.

The maps of S05 and S31 show that they are low/mid-rise site of low and medium density. Longer facade and no open space can also be observed from the maps, which lead to the increase of shaded area covered in blue in APSH maps and shown as taller minimum peak in histograms compared to site in A2. The tall and wide maximum peak shows the influence of large building separation because of low/mid-rise buildings and medium density, where unblocked area locates.

The pattern of histogram of cluster B is actually very similar to which of cluster A2, but has one more minimum peak. Site in cluster B also has low-rise community as in cluster A2, but medium density, long facade and no open space, compared to low density, short facade and in-site open space in cluster A2. This indicates that increasing building density, facade length and decreasing open space ratio will cause

dramatic increase of shaded area, expressed as a minimum peak of similar size to maximum peak.
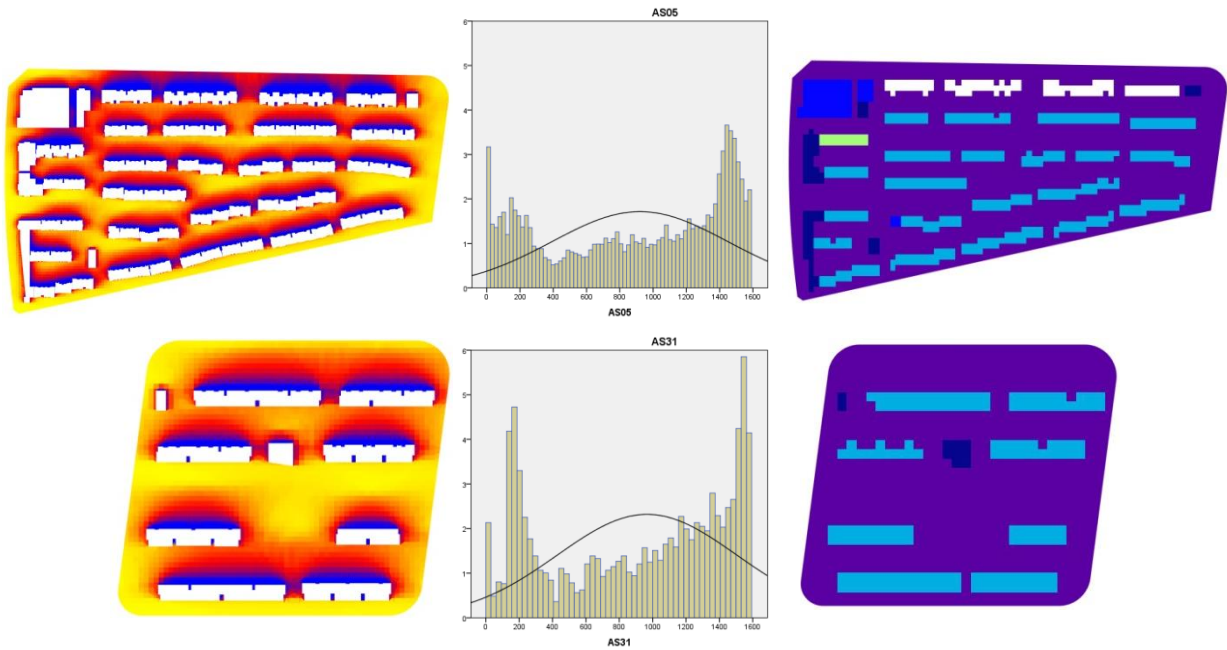
Figure E2 APSH maps, building height maps and histograms of S05 (first row) and S31 (second row) from cluster B

Based on analysis on cluster B, it could be said that compared to site with attributions of low/mid-rise, low density, short facade length, and proper amount of open space, sites with medium density, long facade and no open space will cause dramatic increase of shaded area, expressed as a tall minimum peak with similar size to the maximum peak.

Table E2 histogram analysis of sample site in cluster B

Cluster	B	B
Example	S05	S31
Distribution Charcteristics	Low-rise, Medium Density, Less Enclosed Boundary, No Open Space	Low-Rise, Medium Density, Open Boundary, In-site Open Space
Max Peak Height	3.7%	5.9%
Max Peak Location	1440h	1600h
Middle Peak Height	N/A	N/A
Middle Peak Location	N/A	N/A
Min Peak Height	3.1%	4.8%
Min Peak Location	180h	200h
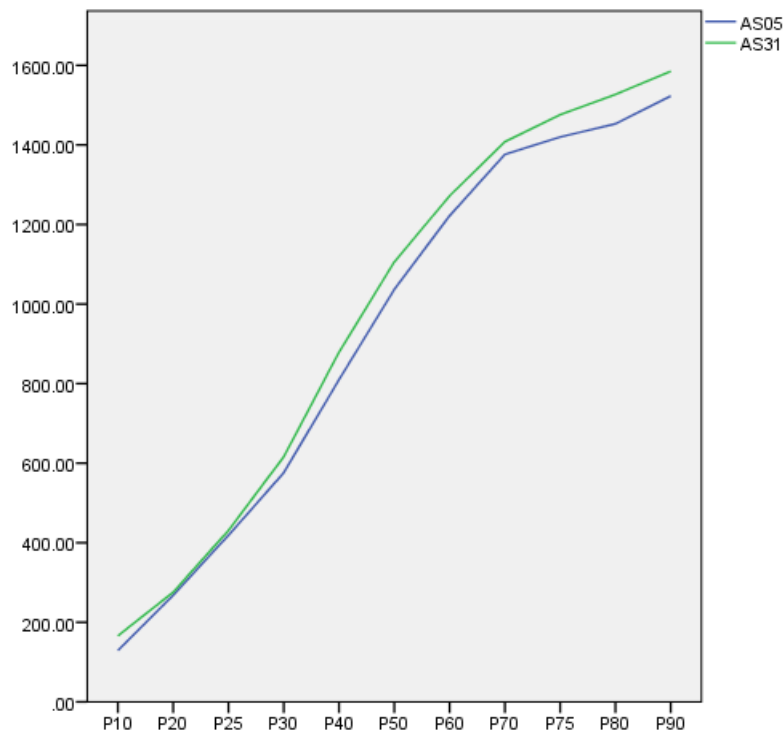
iii. Phenomenon and Reason in APSH and Histogram of Fair Performance Cluster (C)

Cluster C is considered the group with fair condition of APSH, because the APSH maps and histograms both show considerate amounts of high sunlight hour area as well as heavily shaded area. This cluster is categorised into three sub-groups: C1 refers to low/mid-rise site in high building density, with and without open spaces; C2 refers to high-rise site in medium density, with and without open spaces.

Sub-group C1 consists of sites in low/mid-rise community in high density. Majority of sites in this sub-group shares building distribution characteristics and APSH performance with S2102: 1. Low/mid-rise community in high density, with some in-site open space; 2. three peaks shown in histogram, short minimum and maximum peaks and tall right-skewed middle peak(Figure 6.5.4).

The histogram of S2102 has a middle peak with axis at 1300h, closed to maximum peak as seen in table E3. Area of 1300h is where between building long facades,

shown as orangey red in APSH map. The small in-site open space is also under influence of building shadow, so that it contributes to the accumulation of orangery red area in map as well as yellow unblocked zone. The contribution from the in-site open space on yellow unblocked area leads to the small maximum peak in histogram. Blocked area behind the long facade of low-rise buildings gives the small minimum peak in the histogram.

Some sites in C1 have open spaces on the edge of the sites, for instance S12. S12 also consists of low/mid-rise buildings in high density and even half enclosed on north-east side. But it has a relatively large open space compared to its own size, to the south and east. In the histogram of S12, its axis of middle peak locates at 1050h, with height of 3.1%. This middle peak is much less right-skewed compared to the middle peak of S2102 and also significantly decreased in height. This is the result of influence of long facades, enclosure and high density. Although the low and middle range APSH of S12 is much lower than which of S2102, S12 has a boost in maximum peak, due to influence of the on-edge open space. The height of maximum peak is 5.5% for S12 compared to 2% for S2102.

Table E3 histogram analysis of sample site in cluster C1

Cluster	C1	C1	C1	C1
Example	S2102	S1101	S12	S1501
Explanation	Low-rise, High Density, Less Enclosed Boundary, In-site Open Space	Low-rise, High Density, Enclosed Boundary, No Open Space	Low-rise, High Density, Less Enclosed Boundary, On-edge Open Space	Mid-Rise, Low Density, Open Boundary, In-site Open Space
Max Peak Height	2%	1.8%	5.5%	3.7%
Max Peak Location	1600h	1650h	1650h	1650h
Middle Peak Height	4%	3.8%	3.1%	2.5%
Middle Peak Location	1300h	1250h	1050h	1000h
Min Peak Height	1.5%	3.1%	5%	1%
Min Peak Location	200h	200h	190h	190h
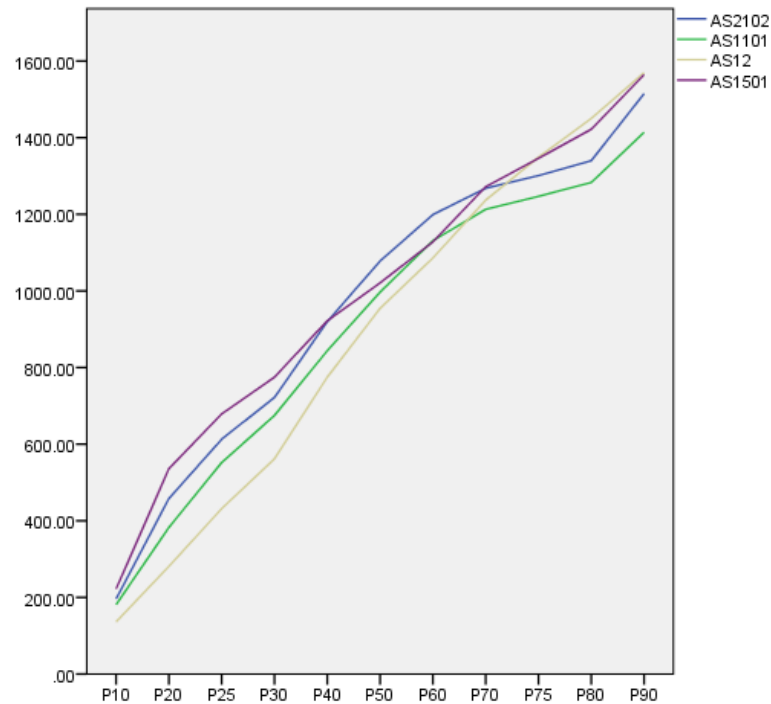
S1501 is of similar histogram pattern as S12. But S1501 has large in-site open space, than on-edge open space. Although there are long facades clustered causing heavy shadow behind buildings, because of the low density, S1501 has generally good APSH performance. Therefore, it could be said that open space on-edge or very large in-site open space, will boost maximum peak. But it cannot be considered a sign of good overall sunlight possibility. The overall APSH condition relies on the combination of distribution attributes.

S1101 represents site of low/mid-rise community in high density with small in-site open space, but highly enclosed boundary. As seen in its histogram, compared to S2102, the location of the middle peak axis is not changed. However, there is a protruding hump between 600-1000h on the left tail of the middle peak, indicating the increased partly shaded area due to the enclosure of the site. This is similar to the hump in histogram of S0201 due to enclosure, although S0201 is in high-rise condition. The minimum peak is also increased in height, 3.1% for S1101 compared to 1.5% for 2102. Therefore it could be said that enclosure at boundary for low/mid-rise site in high density, will lead to expand of frequency of lower tail of middle peak, in range of 600-1000h.
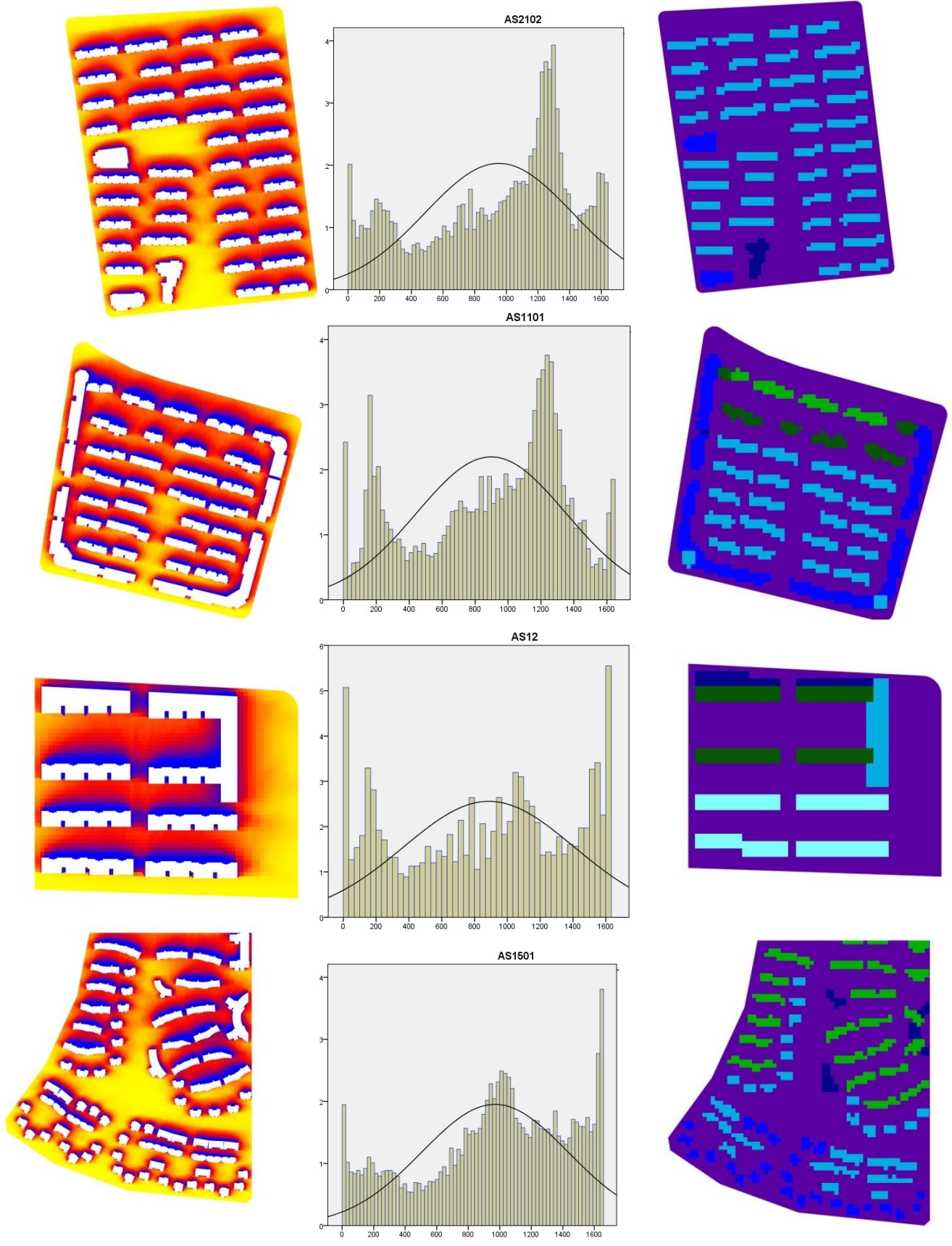
Figure E3 APSH maps, building height maps and histograms of S2102, S1101, S12 and S1501 from cluster C1

Cluster C2 has most clear pattern of histogram compared to all other clusters: small minimum peak and maximum peak, and extremely tall, right-skewed middle peak with considerable width (Figure E4). The tiny maximum peak is an indication of lack of open space no matter in-site or on the edge of site where sunlight hour can reach high value around maximum. The small minimum peak is a reflection of a relative low ratio of the blocked area behind high-rise buildings. That is because of the short facade of high-rise buildings and relative large building separations.

As seen in figure 6.5.5, S2603 as a representative of site consisted of high-rise buildings in a medium density and of no open space. Because of lack in unblocked area, the histogram of S2603 has tiny maximum peak. Under the impact of high-rise buildings and a long facade to the south, a rounded minimum peak exists between 0-400h.
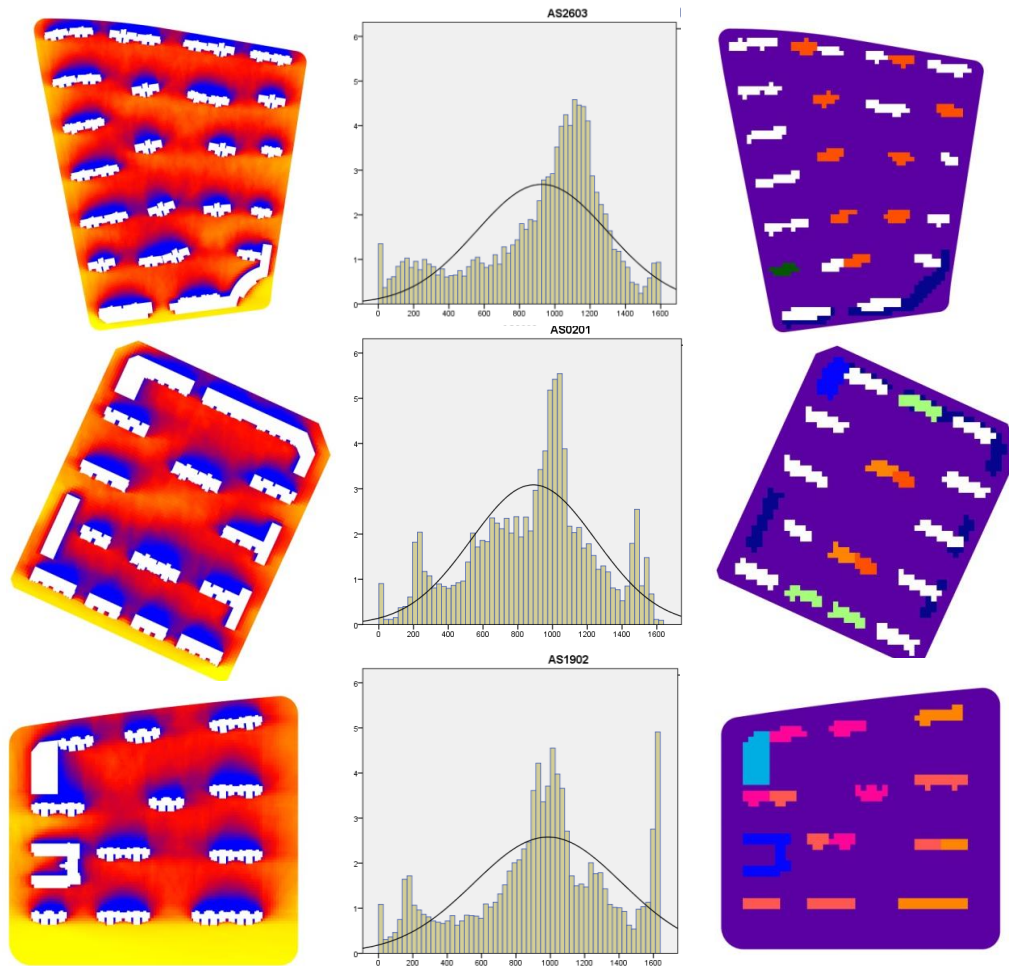
Figure E4 APSH Maps and Histogram of S2603, S0201 and S1902 from Cluster C2

Similar to S2603, map of S0201 is also covered in orange and red namely1000-1400h of sunlight. Obviously, S0201 has high degree of enclosure on the boundary of the building cluster. This leads to a left move of the middle peak axis at 1000h, compared to axis at 1180h of S2603 (Table E4). It also leads to an increase of height of minimum peak, at 2% compared to 1% for S2603. Between 600-1000h in histogram of S0201, a protruding hump between 500-900h shows in the left tail of the right-skewed middle peak. The hump is an indication of the boost of low sunlight hour area due to the enclosed boundary of high-rise community in S0201. This is similar to the appearance of the hump in S1101, due to enclosure in low/mid-rise community. Therefore, it could be summarise that high-rise community in medium density if the enclosure at boundary increased, will have higher frequency on the low to medium range in histogram, especially on at minimum peak and between 500-900h.

Table E4 histogram analysis of sample site in cluster C2

Cluster	C2	C2	C2
Example	S2603	S0201	S1902
Explanation	High-Rise, Medium Density, Open Boundary, No Open Space	High-Rise, Medium Density , Enclosed Boundary, No Open Space	High-Rise, Medium Density , Less Enclosed Boundary, Open Space
Max Peak Height	1	2.5	5
Middle Peak Location	1180	1000	1000
Min Peak Height	1	2	1.8
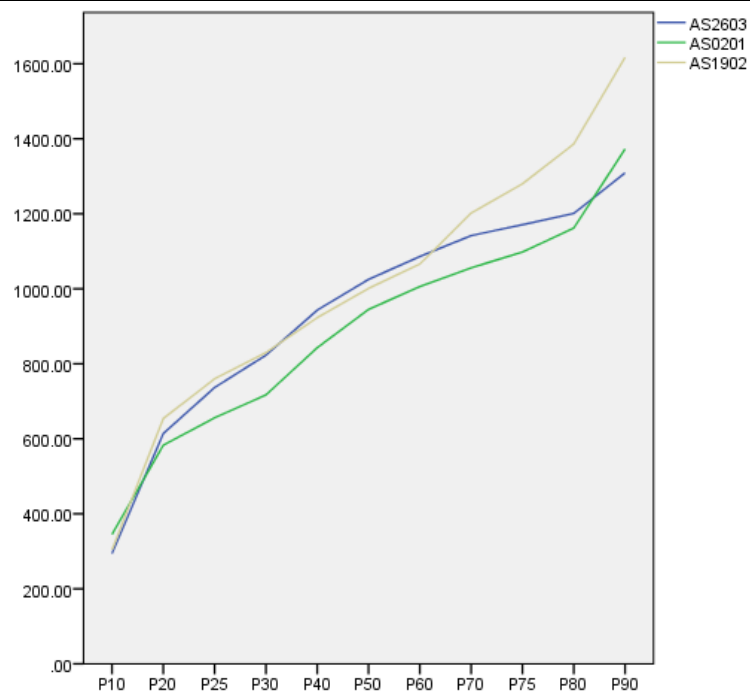
S1902 is a high-rise site in medium density with open space on the south edge of the site. Building height of S1902 is also higher than which of S2603 and S0201. The histogram of S1902 shows a tall maximum peak due to the existence of the open space. The height of maximum peak boosts to 5% compared to 1% in S2603 histogram. Under similar condition of boundary enclosure, S1902 has higher building which causes heavier shadow from the building compared to S2603. So that the axis of middle peak in S1902 histogram is at 1000h to the left of which in S2603 at 1180h. As well as the location of middle peak, height of minimum peak is also influenced by increase of building height. The height of minimum peak in S1902 histogram is 1.8% compared to 1% in S2603 histogram. As a result, it could be said that for high-rise building of medium density, to have open space will boost sunlight hour in high level range. However, it cannot be defined as a sign of good sunlight possibility as the sunlight hour between buildings is not improved. But for existent site and newly-designed site need improvement in sunlight possibility, adding in open space is still an approach worth considering. Also, for high-rise site of medium density, increase building height will cause increase in minimum peak and less right skewness of the middle peak in histogram, corresponding to decreasing of sunlight hour in low and middle level ranges.

Based on the analysis of cluster C, conclusions could be consolidated as below:

1. the majority of practices have fair APSH condition, and have three peaks in histogram of APSH grid data.

2. for low/mid-rise high density and high-rise medium density sites, open space on the edge of site will boost maximum peak in histogram. However, tall maximum peak should not be considered as a sign of good APSH performance.

3. for low/mid-rise site in high density, increase of enclosure at boundary will cause increase of frequency of lower tail of middle peak, in range of 600-1000h.

4. high-rise medium density site, if having short facade length, will have small ratio blocked area, shown as small minimum peak in histogram.

5. for high-rise medium density site, when enclosure at boundary increases, the height of minimum peak and between 500-900h will increases. Namely, area of low and medium level APSH value increases when enclosure increases.

6. for high-rise site of medium density, increase building height will cause increase in minimum peak and less right skewness of the middle peak in histogram, corresponding to decreasing of sunlight hour in low and middle level ranges.

iv. **Phenomenon and Reason in APSH and Histogram of Poor Performance Cluster (D)**

Cluster D is the group of poor sunlight possibility. Seen from the map, sites in cluster D are of high-rise buildings and high density with some enclosure at boundary. The maps are considerably covered in purple and blue. The histogram also shows three peaks, but the middle peak is left skewed or with axis at the median.

As shown in figure E5 and table E5, S09 is more enclosed than S08 is. The minimum peak of S09 has a boost on height of 9.2%, compared to 2% for S08. In addition, S14 is less building height than S08 and S09, yet due to the high level of enclosure, the middle peak is even more left skewed at 650h than 800h of S08 and 900h of S09. This indicates that high level of enclosure will lead to further left skew on middle peak and boost on height of minimum peak in histogram.

S14 has a large on-edge open space relative to its own size on the south of the site. Its effect is a supper tall maximum peak in its histogram. This also echoes to the fact discussed under other condition that on-edge open space will lead to boost of maximum peak.

Based on analysis of cluster D, it could be summarised that

1. for poor APSH condition, the middle peak in histogram is left skewed.

2. for high-rise high density sites, increasing enclosure at boundary will cause further left skew on middle peak and boost on height of minimum peak in histogram.

3. for high-rise high density site, on-edge open space will lead to boost of maximum peak.
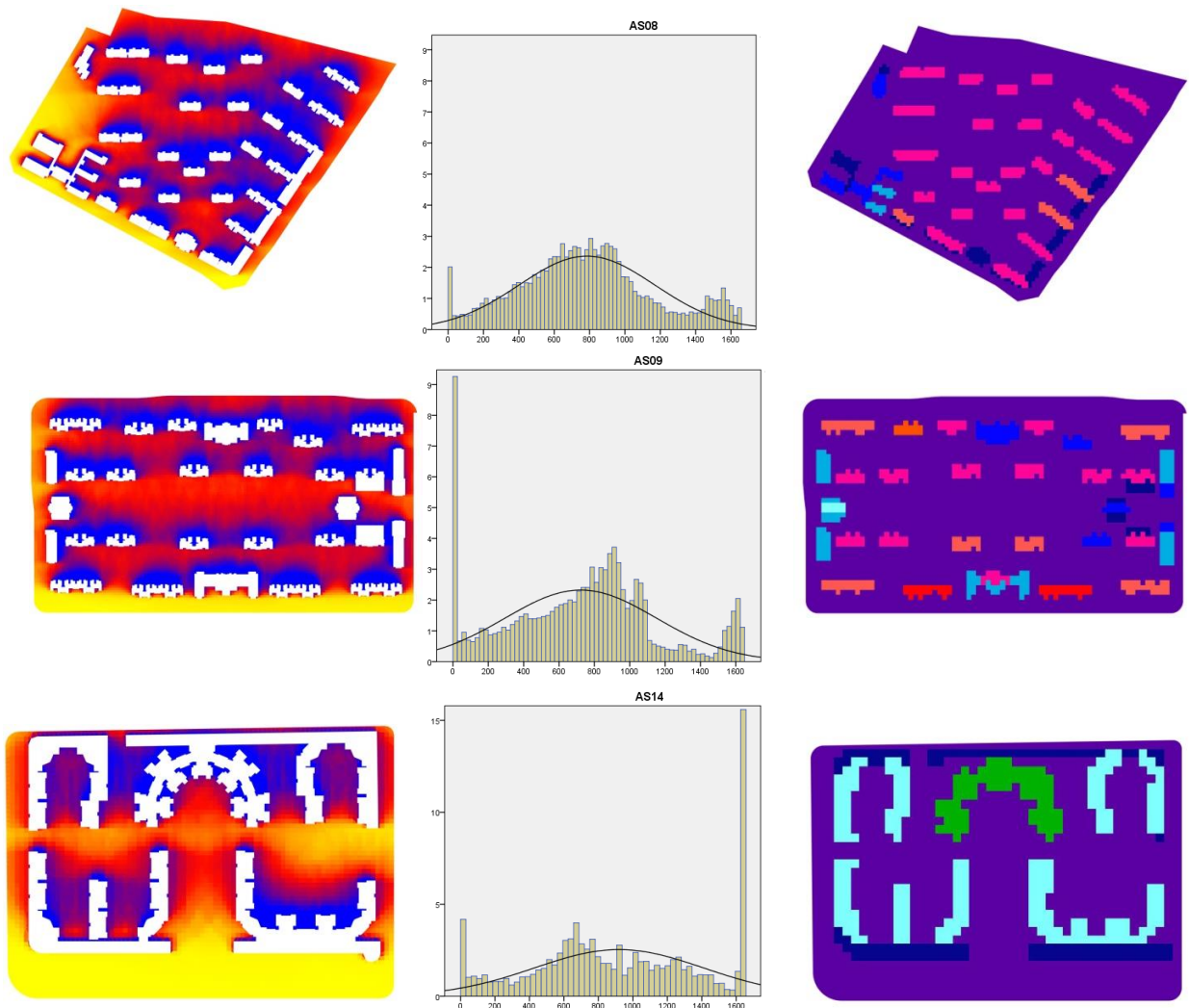
Figure E5 APSH Maps and Histogram of S2603, S0201 and S1902 from Cluster D

Table E5 histogram analysis of sample site in cluster D

Cluster	D	D	D
Example	S08	S09	S14
Explanation	High-rise, High Density, Enclosed Boundary, No Open Space	High-rise, High Density, Enclosed Boundary, No Open Space	High-rise, High Density, Enclosed Boundary, On-edge Open Space
Max Peak Height	1.5%	2%	15.1%
Max Peak Location	1570h	1600h	1650h
Middle Peak Height	3%	3.5%	3.1%
Middle Peak Location	800h	900h	650h
Min Peak Height	2%	9.2%	5%
Min Peak Location	0h	0h	0h
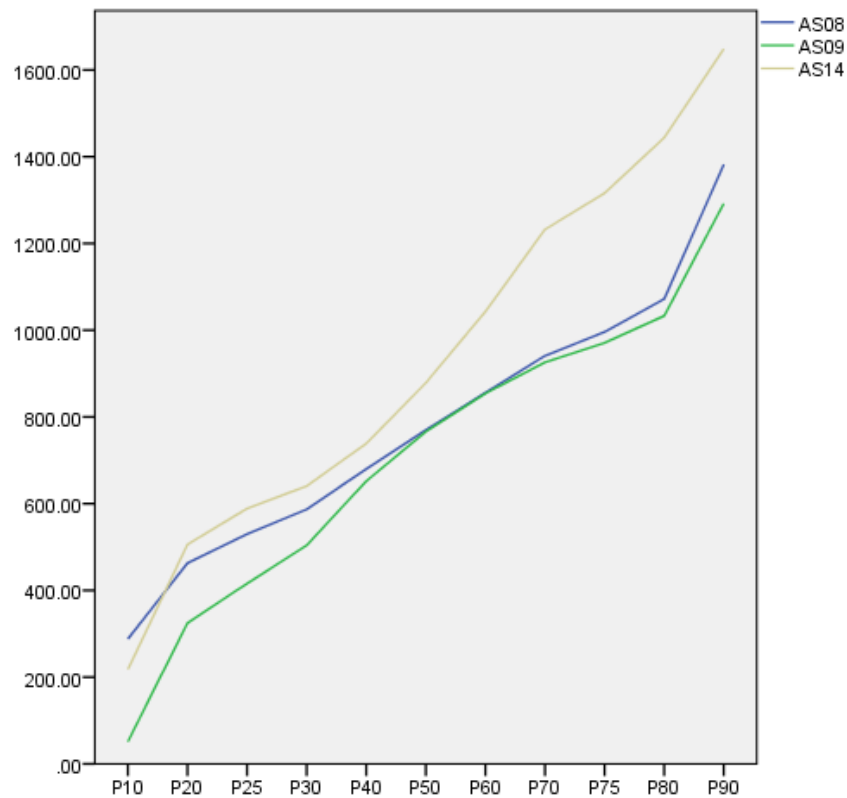
II. Analysis of Clustering by WPSH and Discovery of Rules

i. Rules and Reasons in WPSH Maps and Histograms of Very Good Performance Cluster A

Cluster A is the group of site with very good WPSH performance. Their WPSH maps shows large area covered in yellow, and few blue area of blocked zone. Seen from figure E6, sites in cluster A consists of low-rise building in low density, evenly distributed and have in-site open space. The histograms show two peaks, which is actually the middle peak merges into maximum peak. Location and height of maximum, middle, minimum peak are listed in table E6.
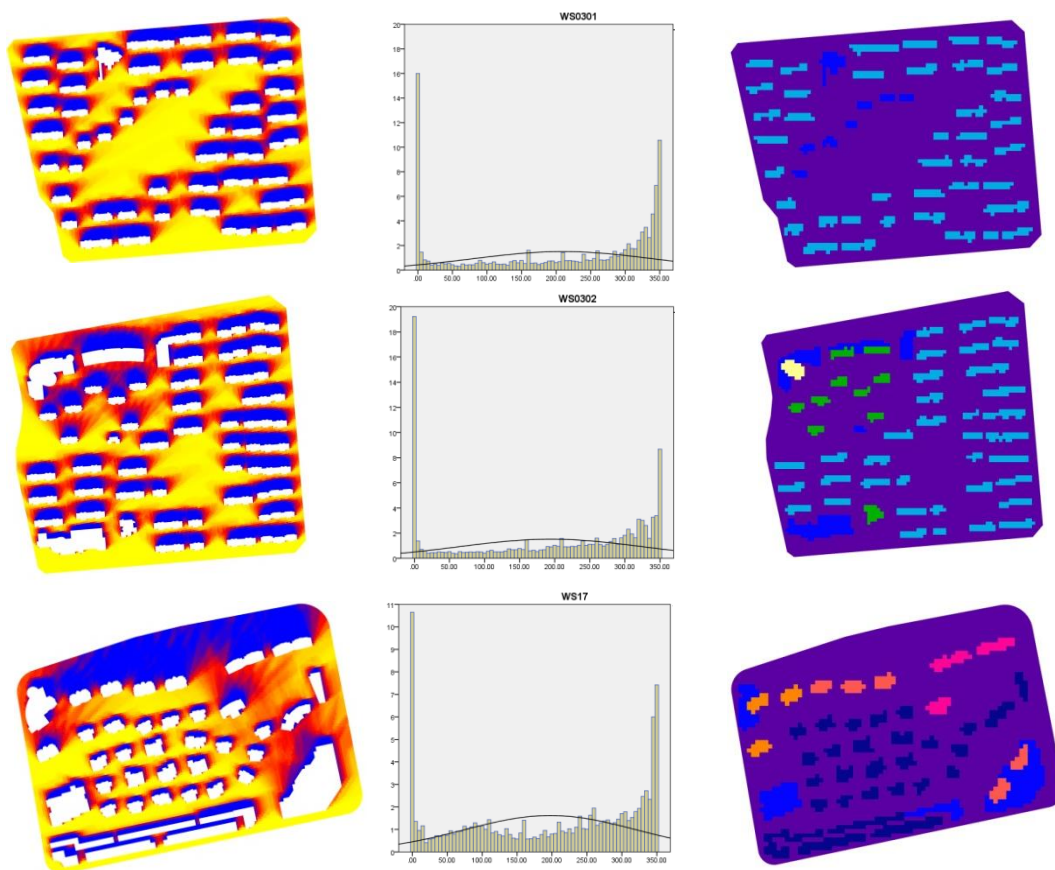
Figure E6 WPSH Maps and Histograms of S0301, S0302 and S17 from Cluster A

From the analysis of differences in table 6.6.1, it could be said that in low/mid-rise

site of low density,

1. increasing building height will increase minimum peak height.

2. increasing in-site open space will increase maximum peak height.

3. mix high-rise buildings in low/mid-rise density site, will cause significant decreasing in minimum peak, but generate a left skewed middle peak.

4. low-rise buildings , low density, in-site open space, uneven is best combination of distribution attributes for extended overall WPSH.

Table E6 WPSH Maps and Histograms of S0301, S0302 and S17 from Cluster A

Cluster	A	A	A
Example	S0301	S0302	S17
Description	Low/mid-rise, Low Density, Uneven, In-site Open Space	Low/mid-rise, Low Density, Uneven, In-site Open Space	Low/mid and High Rise Low Density, Even
Max Peak Height	11%	9%	7.3%
Max Peak Location	350h	350h	350h
Middle Peak Height	N/A	N/A	1%
Middle Peak Location	N/A	N/A	90h
Min Peak Height	16%	19%	10.7%
Min Peak Location	0h	0h	0h

ii.   Phenomenon and Reason in WPSH Map and Histogram of Fair Performance Cluster B

Cluster B is the group of site with fair WPSH performance. Three sub-groups are divided: B1 for low/mid-rise, mix density sites, B2 for low/mid-rise, medium density site, and B3 for high-rise, medium density sites.

Sub-group B1 shares attributes of low-rise buildings, mix density of low and high, unevenness, and in-site open space. The sub-group B1 is a special case of B2, which shares similar attributes but in medium density and evenly distributed. B1 has medium density in average, but consists of high density building clusters and low density building clusters.

WPSH maps of B1 shows that open space is covered in yellow, while area in between

long facades of buildings is in blue. The histogram appears three peaks. Parameters of histograms are shown in table E7. The relatively tall minimum peak of 15-20% is under influence of long mid-rise facades in the high density building clusters. The short maximum peak of 5% is because the small in-site open space cannot contribute much to the height of maximum peak than the large open space do. The middle peak locates at 100h, with thick tails on both sides, especially on the right.

Sub-group B2 consists of sites with attributes of medium density, evenness, on-edge open space on the south, and low/mid-rise buildings (S29, S30) or mix-rise buildings (S260501, S23, S28, S16). Their WPSH maps shows yellow in open space, red between buildings and blue behind buildings. Also three peaks appear in histograms. Details could be seen in table E7.

Compared to B1, B2 shares similar height at minimum peak, mean while, maximum peak of B2 histogram is much taller than which of B1. This is because the on-edge open space on the south in sites of B2 contributes more on height of maximum peak, than small in-site open space does in B1. The location of middle peak for B2 is 200h, on the right of 100h of B1. It shows evenly distributed, medium density site could provide higher middle peak location than uneven, mix density sites. Uneven, mix-density sites tends to boost extreme condition of large shadow area and large unblocked area.

A special case S14 in B2 has taller maximum peak (18%) and smaller middle peak axis (110h) compared to majority sites in B2. The possible reasons are larger on-edge open space on south relative to the site's own size leading to tall maximum peak, and the enclosure at boundary leading to further left skew of middle peak.
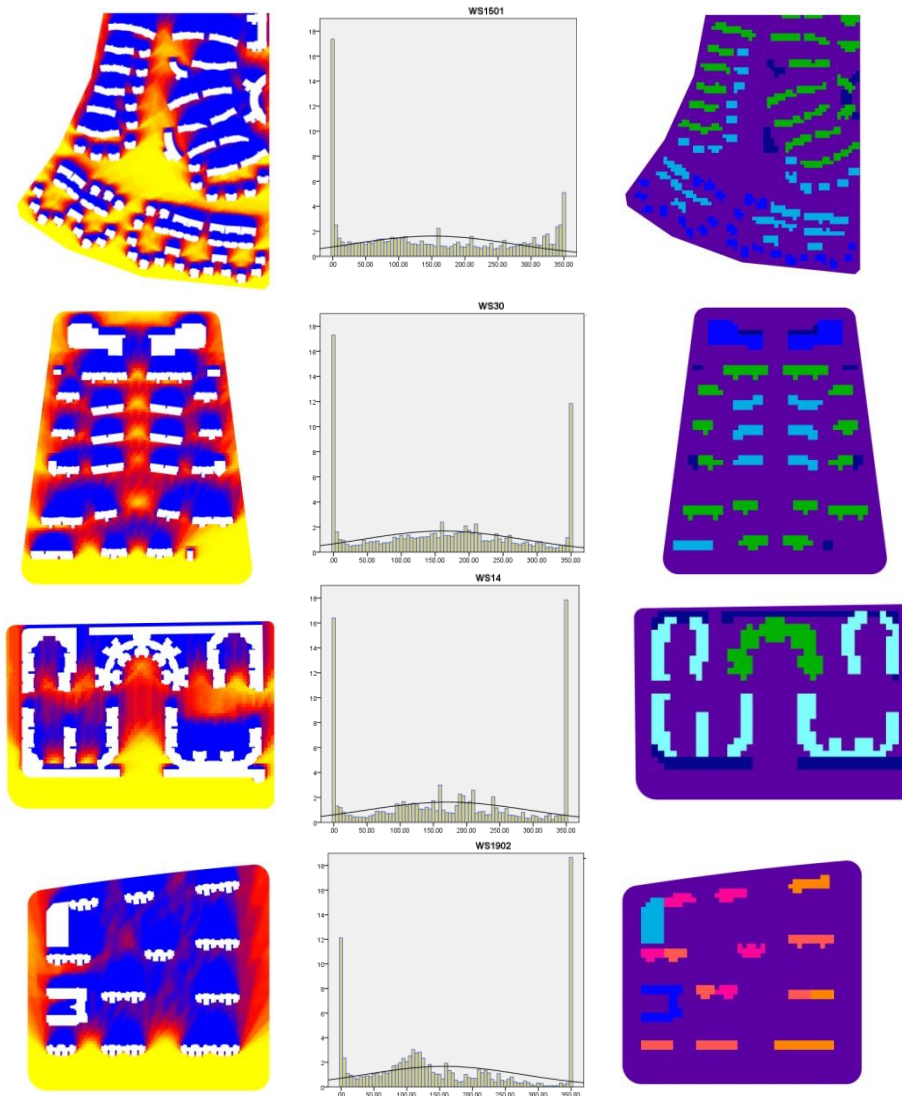
Figure E7 WPSH Maps and Histograms of B1 (S1501), B2 (S30, S14) and B3 (S1902) from Cluster B

Sub-group B3 clusters sites with attributes of high-rise buildings, medium density, unevenness, south on-edge open space. The WPSH maps shows yellow in open space, blue between buildings and red on east and west edges. The three peak histogram details are shown in table E7. It is worth to mention that the middle peak height of B3 is taller and the minimum peak is lower than B1 and B2. This indicates that high-rise building in medium density leads to increased ratio of shaded and partly shaded area, shown as taller and more left skewed middle peak, but less totally blocked area shown as shorter minimum peak. Example site S1902 of B3 has tall maximum peak than majority sites in B2, due to significant contribution from south on-edge open space. However, a special case of S260201 in B3 only has maximum peak at height of 7%,

although with large on-edge open space. That is because the space is on the west edge of the site, which contributes less to the height of maximum peak than space on the south.

Table E7 WPSH Maps and Histograms of S1501, S30, S14 and S1902 from Cluster B

Cluster	B1	B2	B2	B3
Example	S1501	S30	S14	S1902
Description	Low/mid-rise, Low and High Density, Uneven, In-site Open Space	Low/mid-rise, medium Density, Even, On-edge Open Space on South	Low/mid-rise, medium Density, Even, Enclosed, On-edge Open Space on South	High-rise, Medium Density, Uneven, On-edge Open Space on South
Max Peak Height	5%	12%	18%	19%
Max Peak Location	350h	350h	350h	350h
Middle Peak Height	1.5%	2%	1.5%	3%
Middle Peak Location	100h	200h	110h	110h
Min Peak Height	17.3%	17.5%	16.2%	12%
Min Peak Location	0h	0h	0h	0h

The curve chart of P10-P90 comparison between sites in cluster B also echoes with the above discussions. Low/mid-rise site in mix density represented by S1501, compared to medium density site (S30), has lower value at P10-P70, showing the impact from high density part; while has higher value at P70-P90, showing the boosting effect on maximum peak by low-rise building in low density section. High-rise medium density site compared to low/mid-rise site, has higher value at P10-P30 and P75-P90, and lower value at P30-P75. This agrees with the conclusion of high-rise site has more left skewed middle peak and shorter minimum peak in histogram under medium density condition.

From the comparison between sub-groups in cluster B, conclusions are drawn as below:
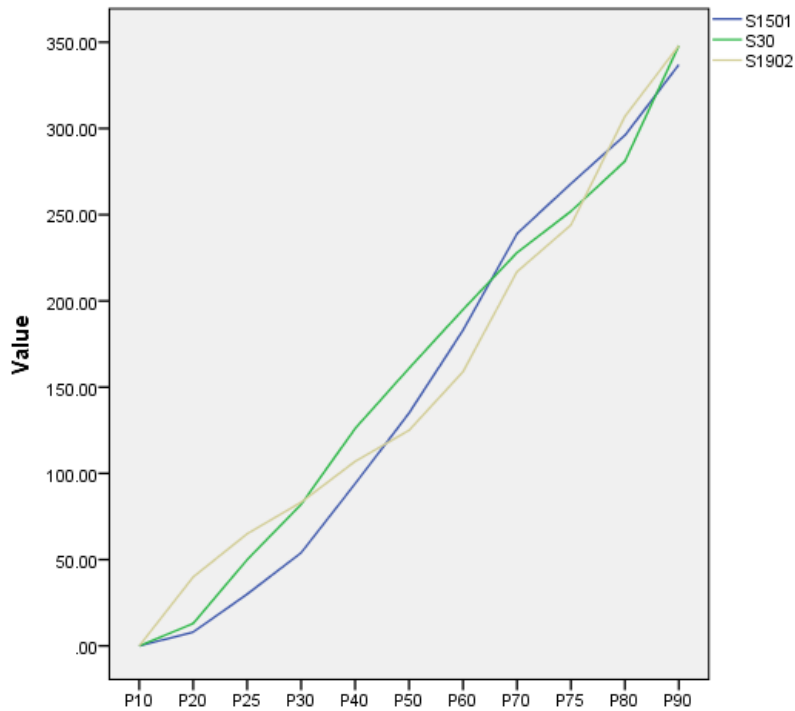
Figure E8 Curve Chart of Cluster B: Low/mid-rise, Mix-density (S1501), Low/mid-rise, Medium-density (S30), High-rise, Medium-density (S1902)

1. Evenly distributed, medium density site could provide higher middle peak location than uneven, mix density sites.

2. for low/mid-rise site and high-rise site in medium density, larger on-edge open space on south leading to taller maximum peak. The sequence of contribution to the height of maximum peak from various types of open spaces is (high to low): on-edge on south, on-edge on other direction, in-site open space.

3. for low/mid-rise site in medium density, the enclosure at boundary leading to further left skew of middle peak.

4. high-rise building in medium density leads to taller and more left skewed middle peak, but shorter minimum peak, compared to low/mid-rise buildings.

iii. Phenomenon and Reason in WPSH Map and Histogram of Poor Performance Cluster C

Cluster C contains sites with generally poor WPSH conditions. They are defined poor due to the significantly left skewed middle peak and tall minimum peak in histograms. The majority of areas in WPSH maps in cluster C are covered in blue; slices of yellow exist at the south edge of the site. The histograms of cluster C have three peaks, with a left skewed middle peak.

Sub-group C1 consists of sites of low/mid-rise buildings in high density, evenly distributed and with small open spaces. Their histograms shows minimum peak at height around 20-27%, maximum peak at height of 5%, and middle peak location at 150-200h similar to cluster B2 of low-rise, medium density sites. S1101 is a special case in sub-group C1. Compared to majority case in C1, S1101 is highly enclosed, which is reflected in its histogram as more left skewed middle peak at 160h. This could be interpret as for low/mid-rise high density site, increasing enclosure leads to further left skew in middle peak.

Figure E9 WPSH Maps and Histograms of C1 (S2101, S1101), C2 (S06)

Sub-group C2 is a special condition of C1, with attributes of mid-rise buildings, high density, evenness, enclosure and in-site open space. Similarly, its histogram has maximum height of 6% closed to which of C1, middle peak axis at 160h closed to S1101 from C1 but with enclosed boundary. However, the height of minimum peak of C2 is especially high at 30-40%. This is as a result of combination of high density, mid-rise building and enclosure. Comparing C1 and C2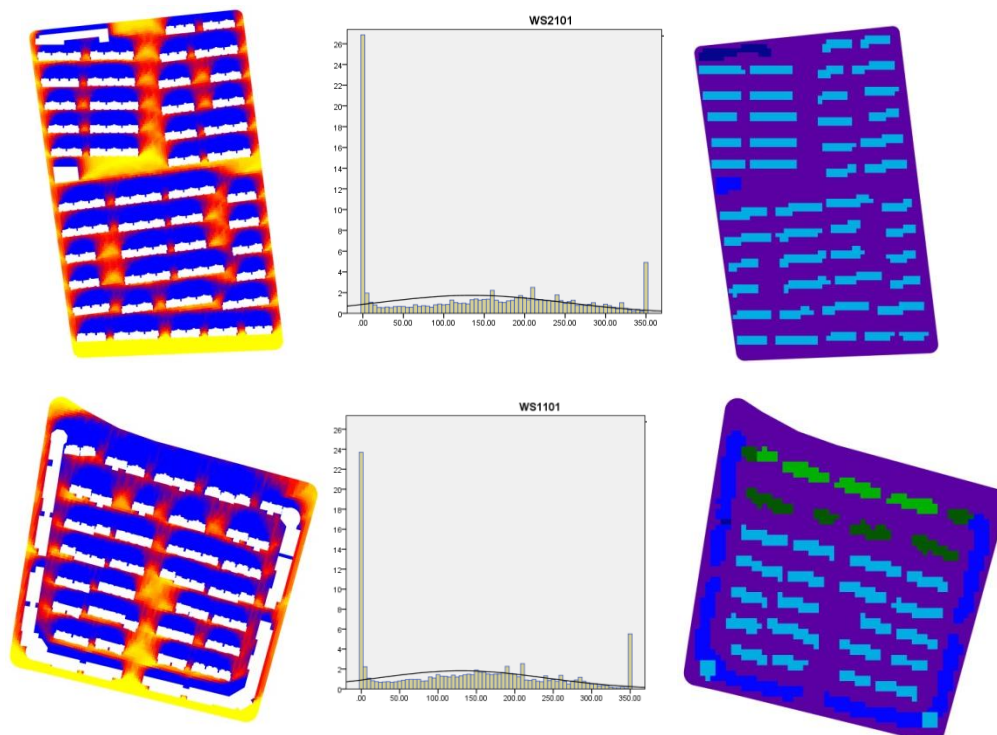, it could be said that for low/mid-rise high density enclosed site, it tends to have tall minimum peak and very left skewed middle peak; and increasing building height leads to significant increase in minimum peak, namely increase in totally blocked area.

Different from C1 and C2, C3 refers to site with attributes of high-rise buildings, high density, evenly distributed. The histograms of C3 shows obvious shorter minimum peak of 10-15% than C1 and C2. Correspondingly, a higher middle peak of 3.5% shows at location of 110h, lower than the locations of C1 and C2 (Table E8). It indicates that under medium density condition, compared to low/mid-rise buildings, high-rise sites leads to shorter minimum peak, but lower location of middle peak axis. This could be interpret as high-rise sites tend to have less ratio of totally blocked area, but higher ratio of partly shaded area. S0201 and S260402 both have on-edge open space on the south compared to S260401 in C3. The maximum peak difference shows that in high-rise high density site, on-edge open space on the south boosts the height of maximum peak.

The curve chart of P10-P90 of WPSH from sites in cluster C accords with the above discuss variation trend. P10-P25 in C1 (blue and green line), and even till P30 in C2 (grey line) are zero, means totally blocked area, which echoes with the extremely tall minimum peak in their histograms. Meanwhile, curves of sites in C3 (purple, yellow and red line) have lower value than curves of C1 and C2 since P40. These echoes with high-rise sites have lower middle peak location, namely higher ratio of low value range. Therefore, it could be summarised that in medium density, high-rise sites provide less zero sunlight area behind the building than low/mid-rise, however the partly shaded area has obviously less sunlight hour than low/mid-rise sites have.

There is a hidden condition of this conclusion that generally high-rise buildings do not have as long facade as low/mid-rise buildings. The large zero sunlight area behind low/mid-rise buildings are heavily influenced by their long facade. So, shorter facades are suggested for high-rise buildings for purpose of less zero sunlight area.

Based on abovementioned analysis, conclusions are drawn as below:
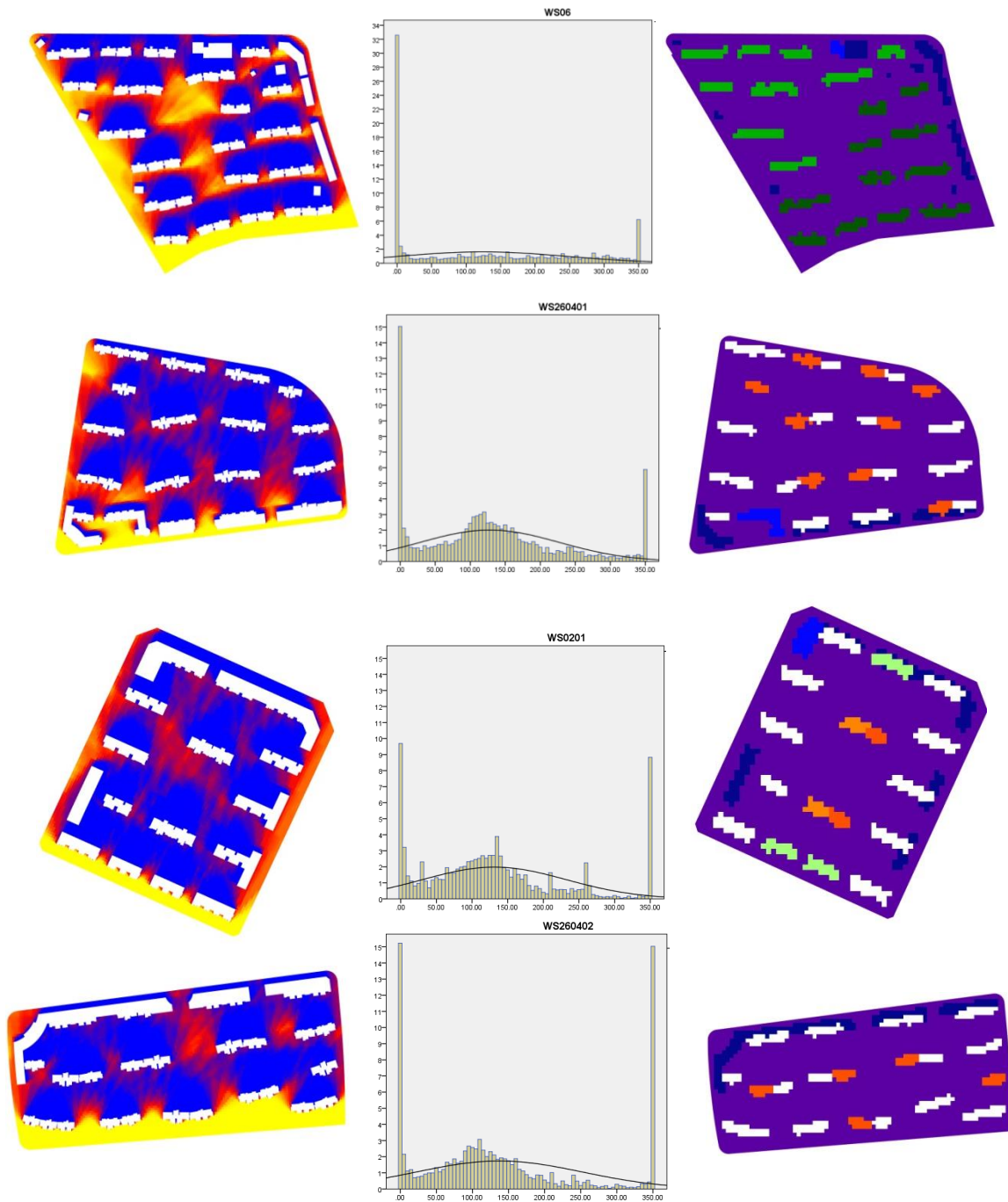
Figure E10 WPSH Maps and Histograms of C3 (S260401, S0201, S260402)

Table E8 WPSH Histograms Details of C1 (S2101, 1101), C2 (S06) and C3 (S260401, S0201, S260402)

Cluster	C1	C1	C2	C3	C3	C3
Example	S2101	S1101	S06	S260401	S0201	S260402
Description	Low/mid-rise, High Density, Even, In-site and On-edge Open Space	Low/mid-rise, High Density, Even, Enclosed, On-edge Open Space	mid-rise, Medium Density, Even, Enclosed, In-site Open Space	High-rise Medium Density, Even	High-rise Medium Density, Even, Enclosed On-edge Open Space on South	High-rise Medium Density, Even, On-edge Open Space on South
Max Peak Height	5%	5%	6%	6%	9%	15%
Max Peak Location	350h	350h	350h	350h	350h	350h
Middle Peak Height	2%	1.7%	1.3%	3.5%	4%	3%
Middle Peak Location	200h	160h	140h	110h	140h	110h
Min Peak Height	27%	24%	33%	15%	9.8%	15.5%
Min Peak Location	0h	0h	0h	0h	0h	0h
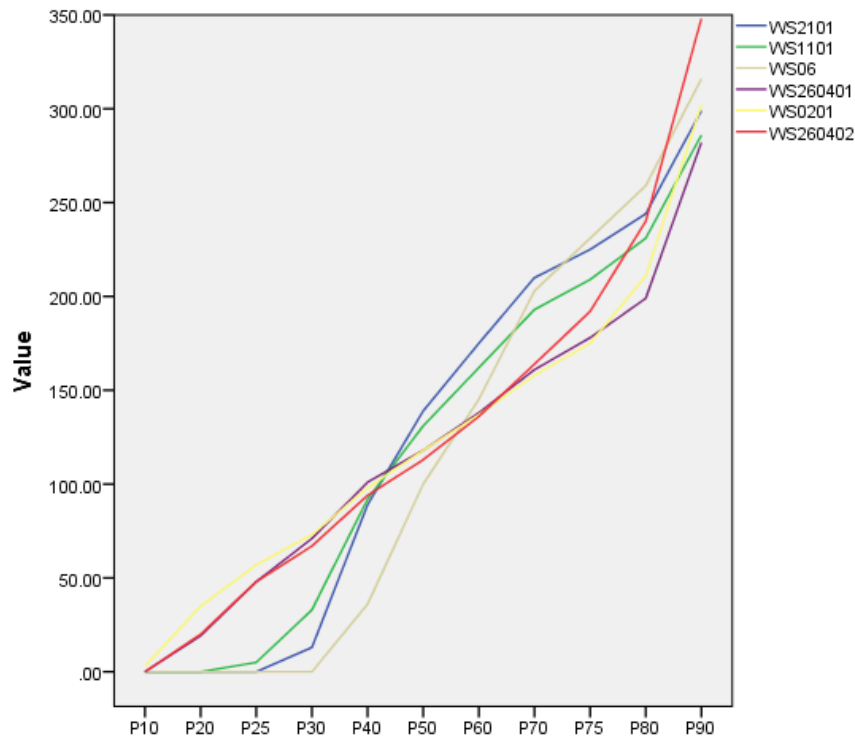
Figure E11 Curve Chart of Cluster C1: Low/mid-rise, High-density (S2101, S1101), C2: Mid-rise, Medium-density (S06), C3:High-rise, Medium-density (S260401, S0201, S260402)

1. for low/mid-rise high density enclosed site, it tends to have tall minimum peak and very left skewed middle peak

2. for low/mid-rise high density site, increasing enclosure leads to further left skew in middle peak.

3. for low/mid-rise high density site, increasing building height leads to significant increase in minimum peak.

4. in high-rise high density site, on-edge open space on the south boosts the height of maximum peak.

5. under medium density condition, compared to low/mid-rise buildings, high-rise sites leads to shorter minimum peak, but lower location of middle peak axis. Namely, high-rise sites tend to have less ratio of totally blocked area, but higher ratio of partly shaded area

iv. **Phenomenon and Reason in APSH Map and Histogram of Extremely Poor Performance Cluster D**

Cluster D is defined as the group of extremely poor WPSH performance. There are two sub-groups: D1 for low-rise, high density, enclosed, even site with on-edge open space; D2 for high-rise, high density, enclosed, even site with no open space. The WPSH maps of cluster D is almost all covered in blue except the edges of the site (Figure E12).

Seen in table E9, in histograms of cluster D, the minimum peak has height around 20-30%, and maximum peak at height of 10%. The height of minimum peak is similar to the level of C2 of mid-rise, high density, enclosed site. Comparing S20 to S09, although S20 has larger on-edge open space on the north, its maximum peak is not significantly taller than the peak of S09 with a slice of open space on the south. This is because open space in other direction does not contribute as much as which on the

south to the height of maximum peak.

The location of the middle peak axis is below 100h, even merging into the minimum peak. Namely, the majority of area in site has WPSH close to minimum value. Details of histograms of two example sites S20 and S09 could be seen in table E9.

Curve chart of S09 and S20 P10-P90 to see, high-rise and low-rise which is better.
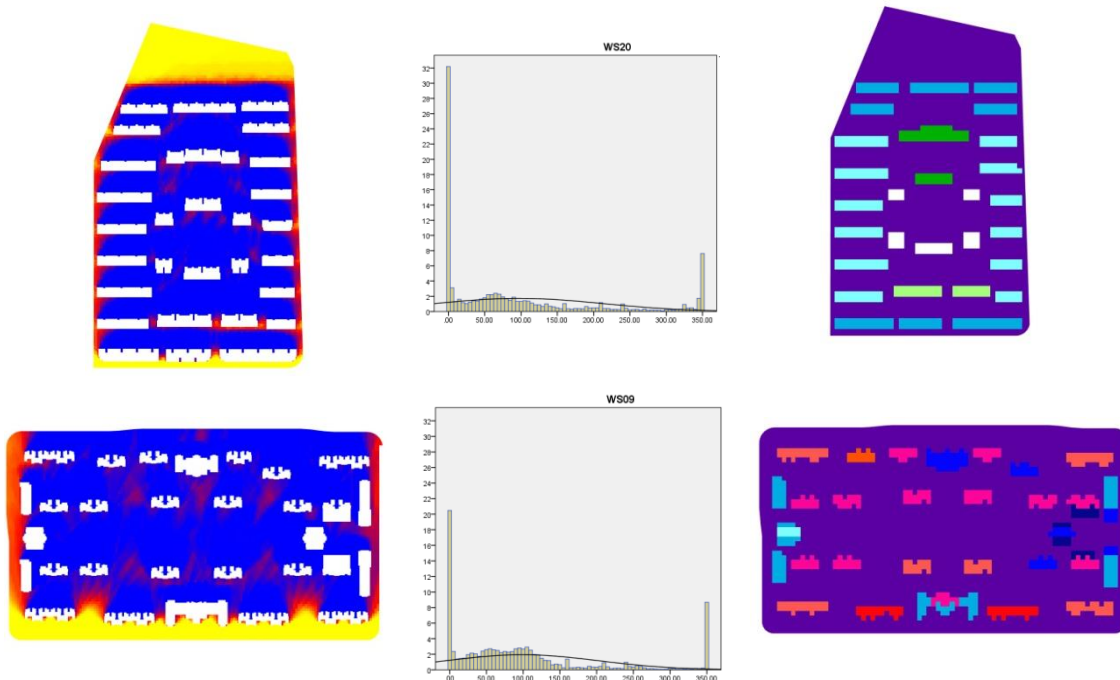
Figure E12 WPSH Maps and Histograms of D (S20, S09)

Table E9 WPSH Histograms Details of D1 (S20), D2 (S09)

Cluster	D	D
Example	S20	S09
Description	Low/mid-rise, High Density, Even, Enclosed, On-edge Open Space	High-rise, High Density, Even, Enclosed
Max Peak Height	7%	9%
Max Peak Location	350h	350h
Middle Peak Height	2.4%	2.8%
Middle Peak Location	70h	115h
Min Peak Height	32%	21%
Min Peak Location	0h	0h

A curve chart is drawn to compare performance differences between low/mid-rise and

high-rise sites under high density condition. Based on these two cases, the high-rise site has less totally blocked area and higher value in lower range, while low/mid-rise site has higher value in higher range. The critical point is at P70. This conclusion needs more sites to validate.
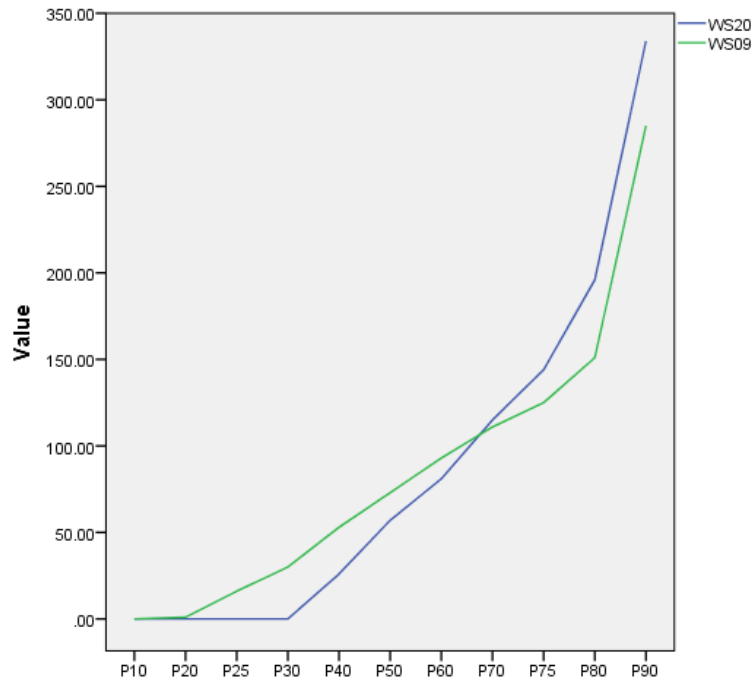
Figure E13 Curve Chart of Cluster D1 (S20); D2 (S09)

Therefore, to summarise, no matter high-rise or low-rise buildings combined with attributes of high density, enclosed, even leads to worst WPSH condition.

Appendix F

Qualitative Analysis Process of Thermal Comfort Simulation Results

I. Analysis of Clustering of Sites by MRT and Discovery of Qualitative Rules of Distribution

i. Rule and Reason Lying in MRT Maps and Curves of Good Performance Cluster

Cluster A indicates the group of sites with good performance in MRT value, as the proportion of MRT value below 300K (26.85 ℃) at least accounts for 30% for group A1 and 40% for group A2. The curves of group A1 and A2 have the leap between P30 to P40 and P40 to P50 respectively (figure F1).

The level of the building distribution indices of sites from group A1 and A2 is defined through matching the building distribution characteristics of sites with the grading of each corresponding distribution indices mentioned in chapter 4.

For group A1, the combination of the following distribution characteristics will lead to high proportion of small MRT: high plot ratio (2.18-3.55), clustered distribution (TSD: 593.27-1054.17), open boundary (aFR: 53.5-64.1) or relative enclosed (aFR: 66.9-85.9), full high-rise ratio (100%). Any enhance of these indices will boost proportion of small MRT. Among all indices, plot ratio has most significant influential power.

A special case S0202 in group A1 shows same level of small MRT proportion in spite of in medium plot ratio and different location of small MRT values on the map. All building shadows in other sites of cluster A locate on the south-east side of the building's south façades, however in S0202, the shadows locate on the north-east side

of the building's north façades. Location of small MRT value is in line with location of building shadows which influenced by the south-west facing orientation of the buildings of S0202. This orientation also leads to the relatively large shadow casting with respect to its building height, namely lower MRT caused by medium plot ratio. Because MRT is a momentary variable related to sunlight direction and the interest time slot in the afternoon, the orientation of building will manifest stronger impact on MRT than on other accumulated performance variables.

Group A2 shares the leap between P40 to P50 in the curve of MRT, namely the area of small MRT accounts for highest proportion of 40%. The only two cases in this group represent one condition of each. S08 represents sites with relatively enclosed boundary, and extremity of very high plot ratio and very large max façade length. S14 represents sites with medium plot ratio and extremity of very high level of density, evenness, enclosure and max façade length. The two types of distribution could be summarised as 1. highly enclosed site with extreme plot ratio; 2. highly enclosed site with extreme density and evenness and medium plot ratio. Both of the distribution types have potential of large shadow casting, therefore they may cause sunlight availability problem in winter or even all year round.

Because small MRT is related to large building shadow area, the distribution requirement of small MRT is reverse to which of long annual and winter possible sunlight hour. To achieve overall good integrated performance, a compromise needs to be made according to the weight of each aspect of environmental performance.

Compared to their APSH and WPSH maps and clustering results, all most all sites in group A1 are clustered into satisfactory cluster B by APSH and WPSH, except for S1902 in cluster A of good performance. For sites in cluster A2 of MRT, S14 and S08 locate in cluster A (good) and B (satisfactory) by APSH and WPSH respectively. These could be interpret as site with high proportion of small MRT does not have to perform worse in sunlight availability, but even have good performance in APSH and

WPSH. Therefore, it could be said that although the requirement on distribution by low MRT and high APSH and WPSH are opposite but not conflictive. They could be balanced and coexist with proper arrangement of distribution characteristics.

Examples of good balance are aforementioned S1902 and S14 (figure F1 and F2). As seen in their WPSH maps, the two sites share characteristic of south on-edge open space, which helps boost the proportion of WPSH peak value. Furthermore, in spite of high-rise ratio only at 35%, S14 has large amount of due east-west orientated buildings rather than north-south orientated buildings. The east-west orientation leads to projection of larger shadow at 16:00 when assessing MRT, than accumulated through winter when assessing WPSH. In other words, the east-west orientation helps on better MRT performance without presenting excessive negative impact on sunlight availability in winter. However, the indoor environmental performance still prefers north-south orientation of residential buildings. Therefore, outdoor integrated performance needs to be considered at second place and compromise to indoor requirements

Future work could be extended on building orientation regarding to thermal comfort and sunlight availability, on details of comparison of south-west and south-east orientation and optimised degree of rotation. Phenomenon of shadow overlapping its own footprint of a building due to orientation change also requires further discussion, which may contribute to reducing solar radiation received on building surface. Except for MRT on horizontal grid, surface temperature of facades and mean radiant temperature from facades on vertical grid would be used as assessment variables.
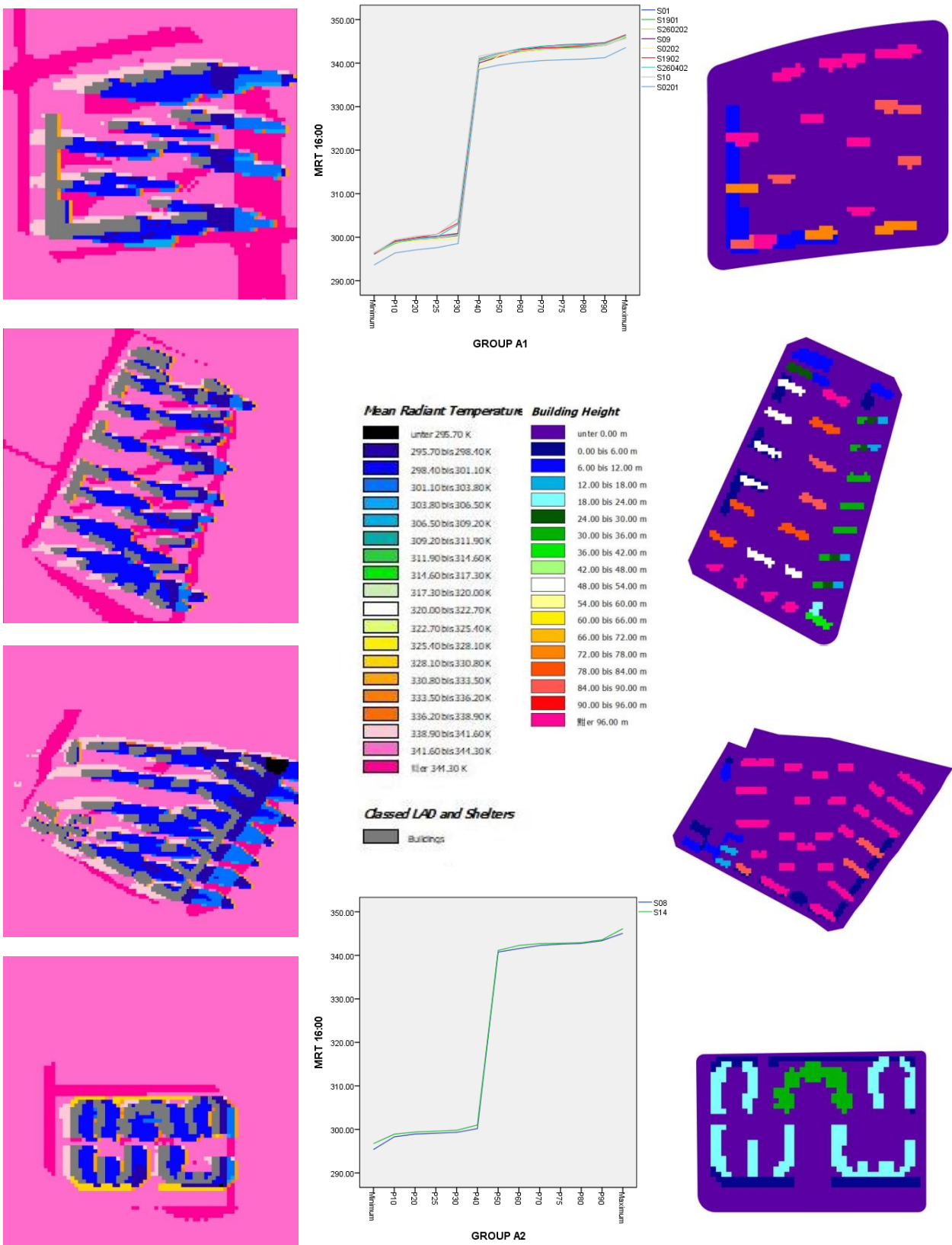
Figure F1 MRT Map, Building Height Map and Curve of Statistical Measure of
Group A1 (Up: S1902, S0202) and A2 (Down: S08, S14)
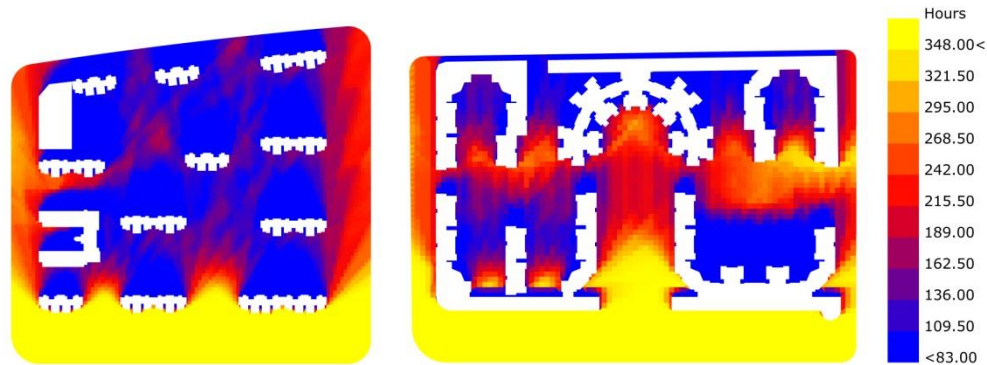
Figure F2 WPSH maps of S1902 and S14

To summarise, the analysis of cluster A of good MRT performance, concludes that:

1. The combination of high plot ratio, clustered distribution, open boundary or relative enclosed, large residential building area, will lead to high proportion (30%) of small MRT value, namely cooler outdoor environment.

2. Highly enclosed site with extreme plot ratio and extreme density, evenness and medium plot ratio will lead to very high proportion (40%) of small MRT value., but may cause sunlight availability problem.

3. Higher plot ratio is most dominant in reducing mean radiant temperature.

4. Orientation of buildings manifest stronger impact on MRT than on accumulated sunlight hour, due to the momentary characteristic of MRT.

5. The requirement on building distribution by low MRT and high APSH and WPSH are opposite but not conflictive, so that win-win is achievable with proper arrangement of distribution characteristics: south on-edge open space and east-west orientation.

6. Building orientation optimisation regarding to outdoor thermal and sunlight comfort, surface temperature on facade, mean radiant temperature from facade on vertical grid are suggested to be supplemented in future work.

ii. Rule and Reason Lying in MRT Maps and Curves of Fair Performance Cluster

The fair cluster B has two sub-groups of B1 and B2. Seen on the curve of the sub-groups, the leaps locate between P25 to P30 and between P20 to P25, respectively for B1 and B2. In other words, group B1 provides higher proportion of small MRT than B2. Except the leap on the polyline, there is another transaction between P30 to P40 and P25 to P40 for B1 and B2 respectively. This section on the polyline refers to the orange slices on the edge of the building shadow shown on MRT map.

Matching the MRT map with building distribution conditions (figure F3), sites in sub-group B1 share characteristics of medium-high plot ratio (1.56-2.37), medium density (0.13-0.20), medium enclosure level (aFR: 51.7-71.8%), medium SVF (56.2-62.8%), and medium-high high-rise ratio (over 69%-100%). Meanwhile, sites in sub-group B2 share characteristics of low plot ratio (0.98-1.86), low-medium distance to road (13.9-39.9), medium-high max facade length (65.5-112.4), and attribute pairs of high density (0.21-0.32) with low high-rise ratio (0-0.35) or low-medium density (0.09-0.16) with high high-rise ratio (0.52-1.0).

A counter example in sub-group B2 is S260401 which has high plot ratio and small building density which has a potential to be clustered into cluster A. However, S260401 is grouped into fair cluster of MRT performance rather than good cluster. Due to the south-west facing orientation of buildings, the shadow at 16:00 at which MRT is assessed overlap greatly with buildings own foot print area. As a result, the shadow area accounts for smaller ratio than other site with orientation other than south-west facing.

A small group of sites in cluster B2 are also worth of noticing because they show fair MRT performance and well balance their sunlight performances. They share distribution characteristics of low plot ratio and low density, namely very spatially open site which should have potential of very good sunlight performance but poor

MRT performance. Therefore, rules of their distribution worth extraction for reference in future design.
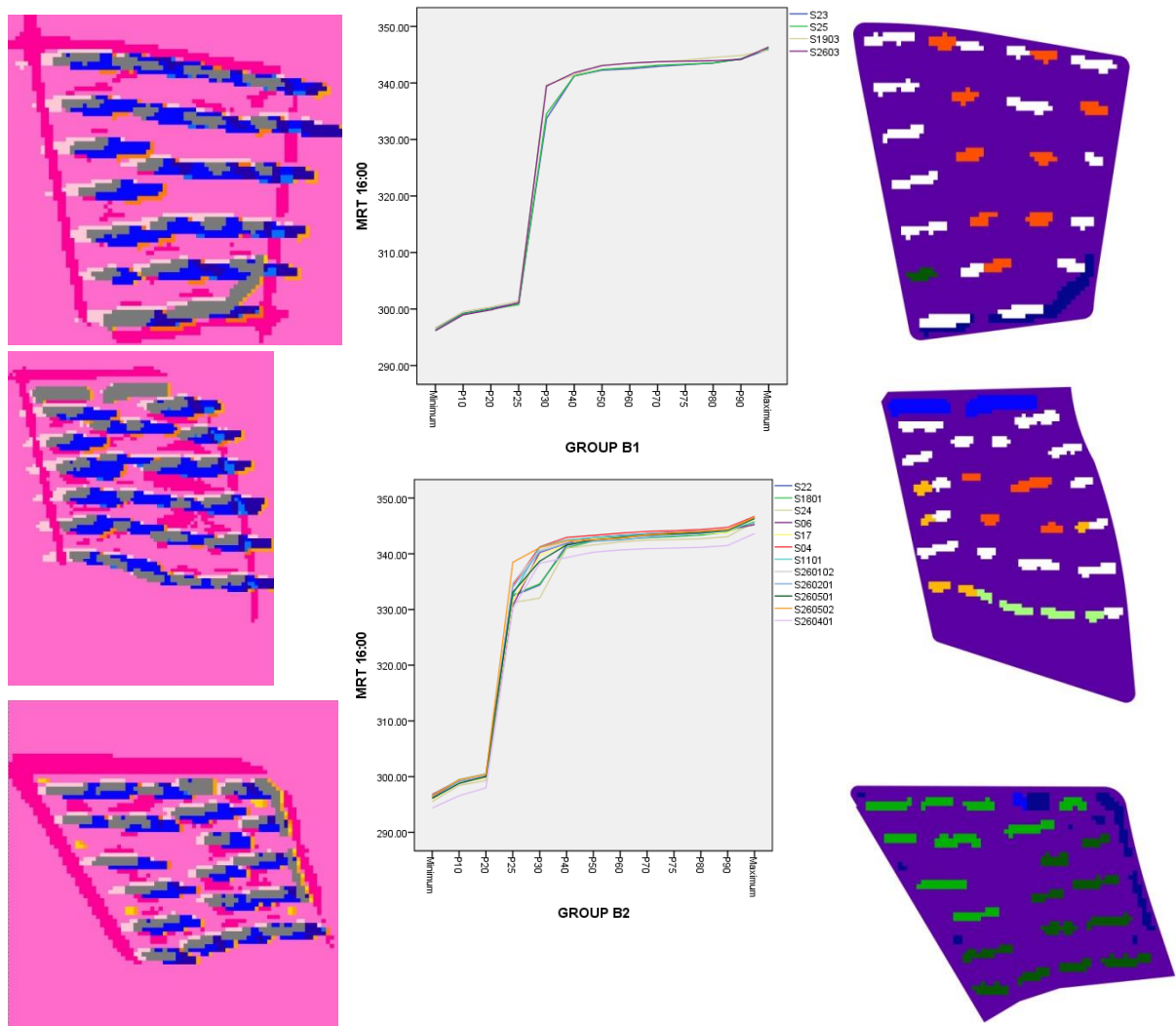
Figure F3 MRT Map, Building Height Map and Curve of Statistical Measure of
Group B1 (Up: S2603) and B2 (Middle: S260502 Down: S06)

Sites 260102, 260201, 260501 and 260502 are all from cluster B2, however with 100% high-rise ratio but low PR and low density (middle of figure F3). They appear fair MRT performance and good performance for both APSH and WPSH (middle in figure F4). By analysis of their building distribution, they are actually high-rise community, but having especially large on-edge open space which makes plot ratio become small. Therefore, their MRT performances show character of high-rise community while large open space provides good sunlight performance different from

other high-rise sites. It is also worth mentioned that in these sites, some rows of buildings lie along direction from north-west to south-east, although the building orientation is north-south. This enhances the shadow overlapping over footprints and other shadows, so that in fact reduces the proportion of small MRT value. These sites may achieve cluster A of good MRT performance, but for this building row character.

S06 is another example from cluster B2, which consists of 9f, 11f and 12f residential buildings. S06 is clustered into fair MRT cluster B2, and into satisfactory cluster B for APSH and WPSH (down in figure F3, right in figure F4). Seen from its distribution, it has well balance in building height and separation of buildings. Even though it is evenly distributed, medium building height with relative large separation compared to its height makes it possible achieving satisfactory APSH and WPSH. While, the considerable medium height of buildings could cast enough shadow to achieve satisfactory MRT performance. This could be summarised that medium level distribution results in satisfactory performance, namely no violation but also no distinction.
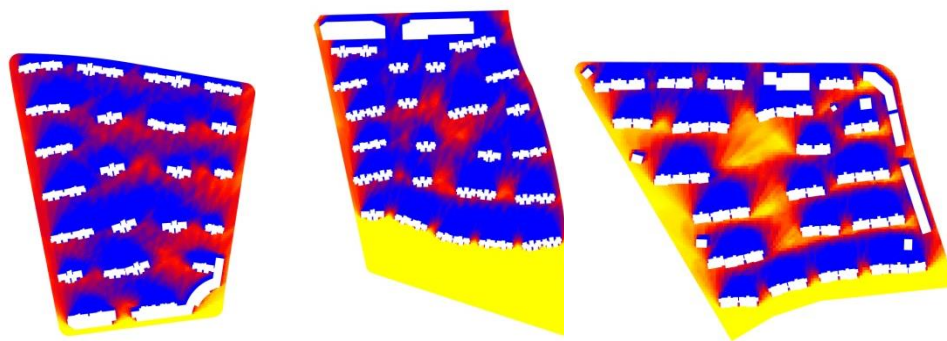
Figure F4 WPSH maps of S2603, S260502 and S06 (from left to right)

To summary, it could be concluded that:

1. Medium level distribution results in satisfactory performance in MRT and sunlight, namely no violation but also no distinction.

2. Rows of buildings lying along direction from north-west to south-east, although the building orientation is north-south, will reduce shadow area relatively due to shadow overlapping; therefore reduce proportion of small MRT value.

3. High-rise community in low density and surrounded by large on-edge open space could achieve a balance of good sunlight performance and fair MRT performance.

iii. Rule and Reason Lying in MRT Maps and Curves of Poor Performance Cluster
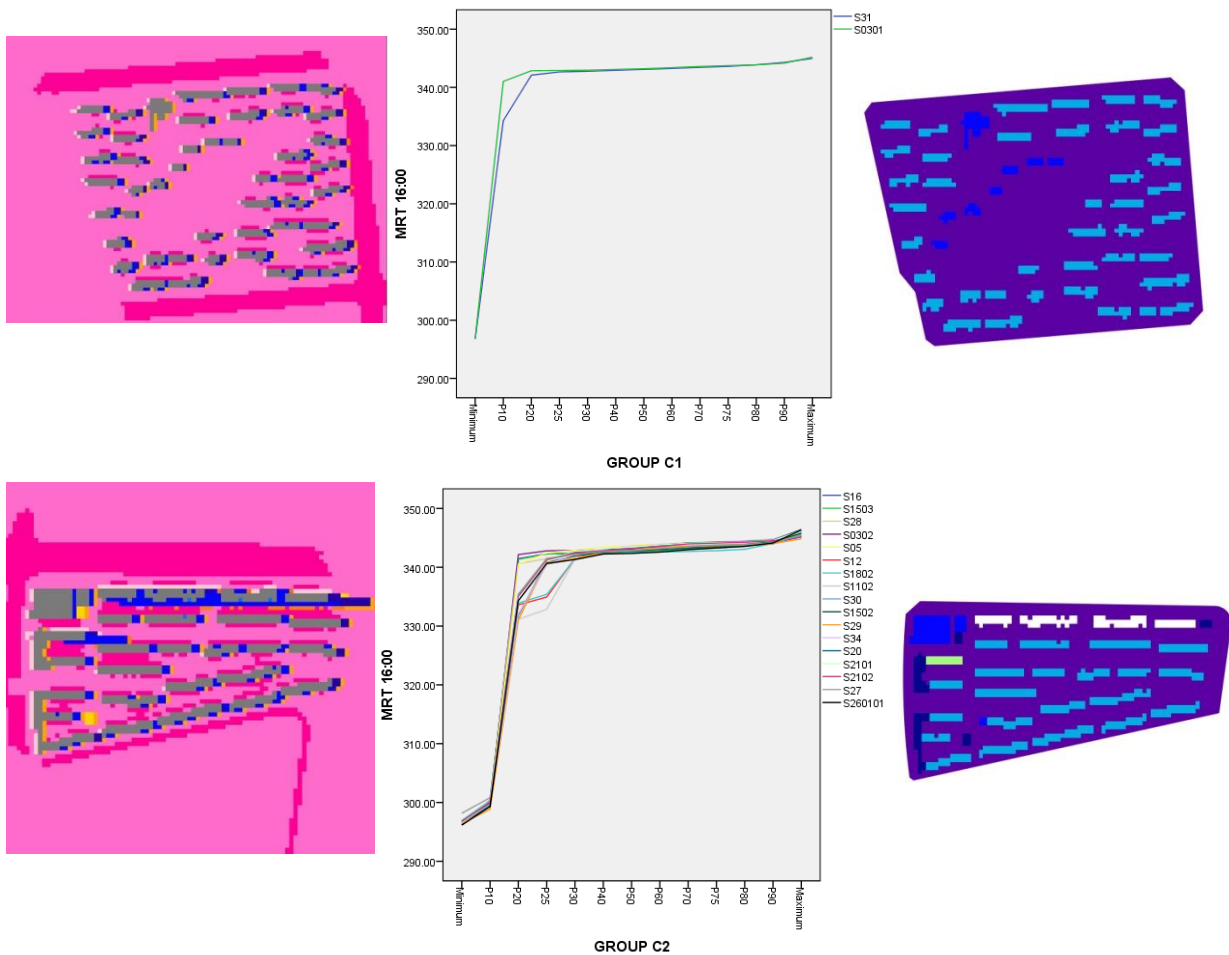
Figure F5 MRT Map, Building Height Map and Curve of Statistical Measure of Group C1 (Up: S0301) and B2 (Down: S05)

Cluster C is of poor MRT performance consisting of two sub-groups of C1 and C2. Curves of C1 and C2 have leaps located at minimum-P10 and P10-P20. This indicates that for C1 the proportion of small MRT value accounts less than 10% and for C2 less than 20%. Cluster C barely has building shadow area that leads to low MRT values.

Sub-group C1 consists of sites with low plot ratio (0.66-0.88), low building density

(0.13-0.15) and zero high-rise ratio, namely shadow casting by residential buildings is very limited. So even P10 of group C1 has reached a relatively high value around 330-340K (56.85-66.85℃). Another moderate transaction occurs on the curve between P10 and P20.

Sub-group C2 consists of sites with low plot ratio (0.83-1.84), low-medium density (0.07-0.20), low facade ratio (42.5-59.3) and low max facade length (31.5-58.9m). The phenomenon shown in cluster C could be summarised as sites with low-rise buildings in low density could barely cast long shadows on various directions therefore, they tend to achieve good sunlight performance but poor MRT performance.

For example, S0302 being one of the most spatially open site (low plot ratio and low density), even having enclosed boundary, is clustered in best APSH and WPSH performance group but in poor MRT group. Some other sites (S05, S12 and S27) still have poor MRT performances but good or satisfactory sunlight performance, even with super high building density and boundary enclosure level. The reason is the shadows from short building height are hardly overlapping the adjacent. Although MRT level could not perform well in this type distribution, this enclose high density pattern by low-rise building is still very encouraged for its distinctive effect of noise barrier and good sunlight performance.
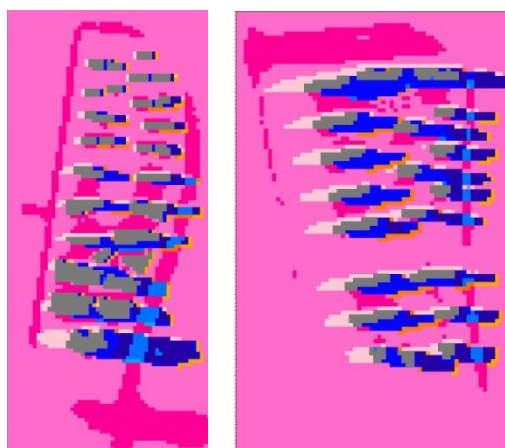

Figure F6 MRT maps of Counter Examples of Cluster C2 (left: S28, right: S260101)

Two counter examples in cluster C are S28 and S260101. They are the only two of high plot ratio even 100% high-rise ratio but of poor MRT performance. The key distribution characteristic they shared is very high shape factor, namely the site in narrow and long shape. Furthermore the long axis of the site is north-south direction. Therefore, the west by south shadow of building at 16:00 which influencing MRT performance would extrude out of the site boundary and the outside part of shadow would not be considered in analysis. So it could be said that the long shadow of high-rises in hot afternoon would have receded contribution to MRT performance within the site, if the shadow extruded out of site boundary due to long-narrow site shape.

To summarise the analysis of cluster C, it could be concluded that:

1. low-rise sites in low density, even some in high-density, tend to have poor MRT performance due to short shadow from building.

2. high-rise buildings may have receded contribution to MRT performance within a certain site, due to the site shape being narrow in east-west direction.

II.   Analysis of Clustering by Wind Speed and Derived Design Guidance

i.   Phenomenon and Reason in Wind Map and Histogram of Best Performance Cluster D

Cluster D has best wind speed performance. They share characteristics of very small site (site 12 and 31), site in very open distribution namely large average SVF (site 301 and 31) or with open space located on south/east side of the site (site 12 and 16) (Figure F7). A single peak with sharp dropping on both sides appears in the histogram of cluster D. The axis of the peak locates around 2m/s, referring to the large area of yellow and orange in wind map as seen in figure F7. The open boundary and large separation between buildings of site 16 on the east side, provide space for higher wind

speed around 3m/s, so that the histogram of 16 has much thicker right tail. It is obviously small sites with less obstacles and having open space of considerable size on south/east side of site (which is dominant wind direction at location) tends to provide better overall wind performance.
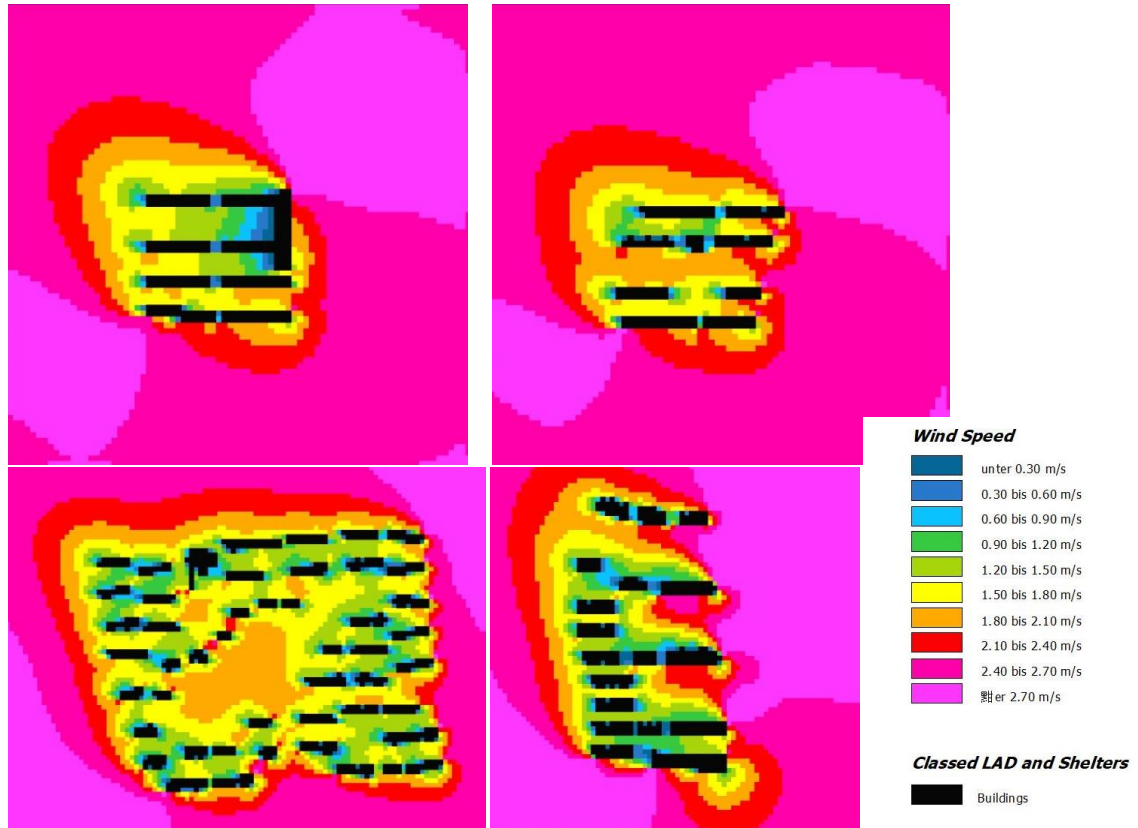
Figure F7 Wind Speed Maps of Site 12(Up-Left), 31(Up-Right), 0301(Down-Left) and 16(Down-Right) and histograms

Counter examples are Site 0302 and 17 (Figure F8). Seen from their maps, wind speed performance is very sensitive to building mass and facade length at influential location although open distribution is vital. Site 17 has generally large SVF, but a very wide podium and a tall tower locate at south-east corner which block wind from entering the site easily. Site 17 is clustered into group C of poor wind speed performance, due to the large area of no-wind shadow cast behind the podium and tower on the dominant wind direction. Site 0302 is one of site with openest and evenest distribution. However, it is clustered into cluster A rather than cluster D, due to dense rows of mid-rise buildings at south-east corner and considerable building

mass at north-west corner of the site. The long facades at down wind direction tend to block wind flows out the site, so that the wind speed in area in front of long facade decreased rapidly.

Based on the anaylsis of wind maps, it could be concluded that smaller site with even and open distribution and open boundary tends to have very good wind speed performance, only if no large building mass blocking the dominant wind direction. Long facade within 90 degrees of wind direction is very influential if locates on upwind direction, but also has negative impact in area in front of the building clusters if locates at down wind direction.
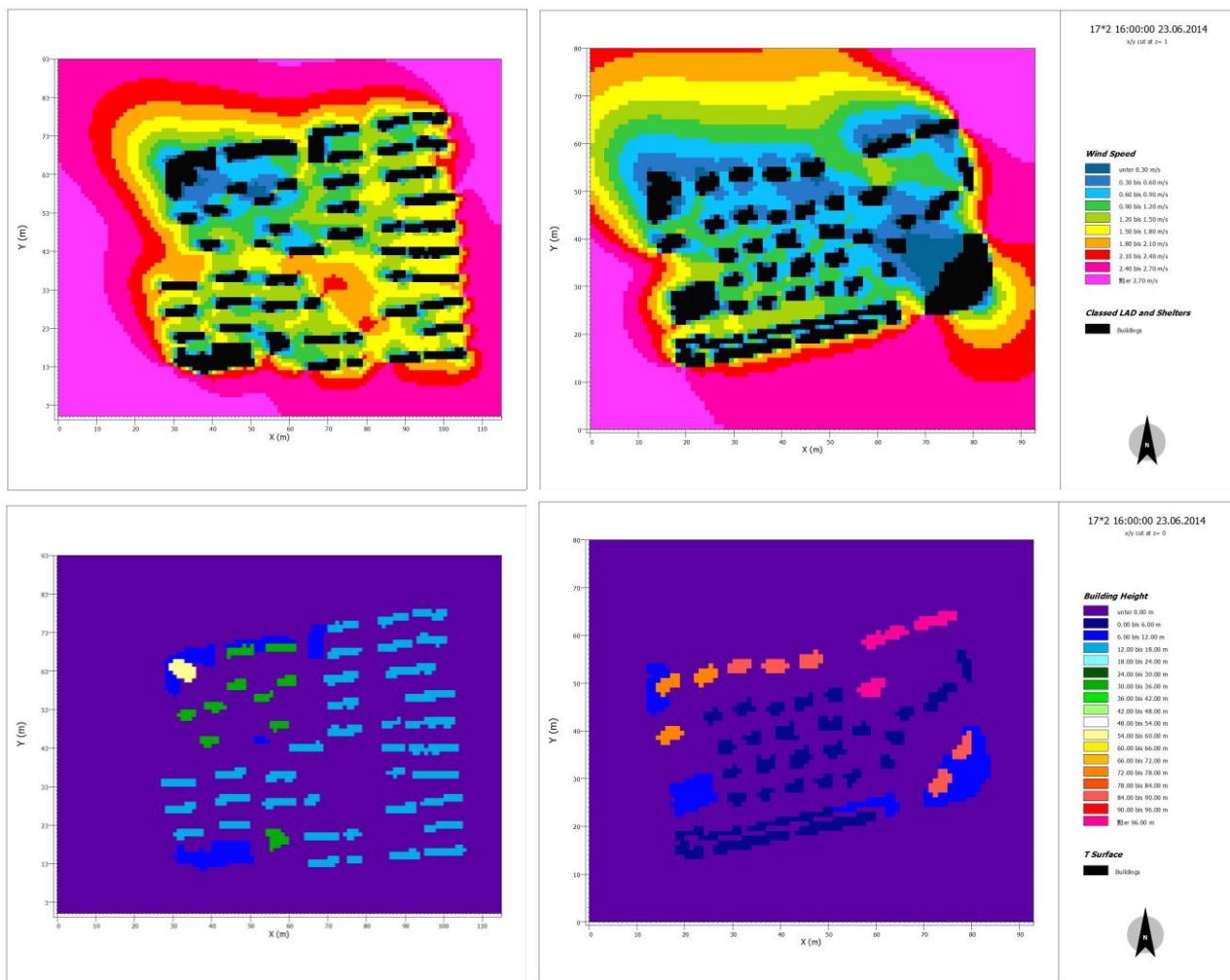

Figure F8 WindMaps and Building Height Maps of Site 0302 (Left) and 17 (Right).

ii. Phenomenon and Reason in Wind Map and Histogram of Poor Performance Cluster

Cluster C is the group with worst wind speed performance. The sites are more or less enclosed with low-rise long facade buildings. There are 4 sub-groups in cluster C, with an exception site 1801. Comparisons between 4 sub-groups show in figure F9.

Cluster C1 show very poor wind performance of large area of no-wind zone. The histogram of C1 appears a single peak axis very close to 0. The tail on right side of peak drops sharply to the bottom around 2m/s and the rest tail remain flat and extremely thin above 2m/s. The similarities of characteristic in building distribution in cluster C1 is the combination of enclosure at site boundary of more than 50% and community of high-rise buildings.

C3 also shows high level of enclosure, even close to 100%, but the community consists of low and mid-rise buildings. Compared to C1, wind maps of C3 shows less no-wind zone, although the general wind performance is poor. This reflects in the histograms of C3 as a single peak close to left side and drop sharply on both sides, but the distance to 0 is large compared to C1. This means the majority values on grid point have very low wind speed but not below 0.3m/s as in C1. Sub-group C1 and C3 indicates that in condition of high level of enclosure at the site boundary, building height has significant influence on overall wind speed condition in site. For site with proper building separation due to sunlight requirement, high-rise community tends to have lower overall wind speed than low and mid-rise community does, if site boundary highly enclosed.

Sub-group C2 consists of high-rise buildings as in C1 but the enclosure at site boundary is limited to 25-50%. The histogram of C2 also has a left skewed single peak, but its distance to 0 is significantly larger than which in C1 and C3. The sharpness of the peak dropping on both sides in C2 also mitigated. This manifests that

less enclosure at site boundary leads to higher overall wind speed in high-rise built site, through raising the wind speed of majority of grid point shown as right-move of the peak axis in histogram.
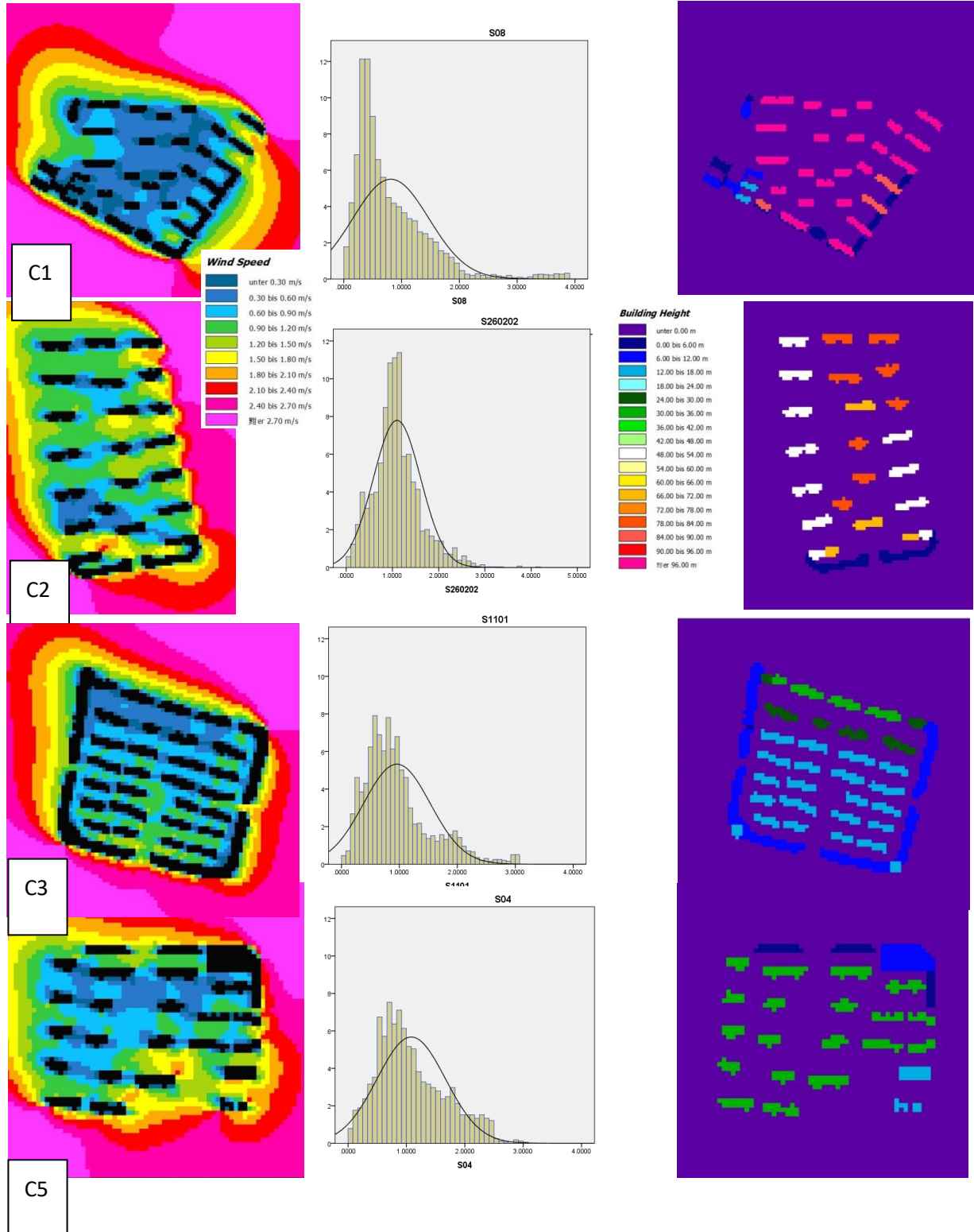
Sub-group C5 is similar to C2 with 25-50% enclosure but consists of low and mid-rise buildings. Seen from its histograms, the position of the peak axis of C5 is slightly left to which of C2, but the right tail becomes much thicker until position close to 3m/s in C5. This indicates that under condition of 25-50% enclosure, low and mid-rise community performances similarly as high-rise community does in wind speed, but each with different emphasis. However, in residential ward, it is more important to avoid no wind area. So under less enclosed condition, high-rise community could be defined as better design with higher wind speed in low wind speed level.

At last comparing the example in C5 versus C3, namely enclosed low and mid-rise community versus enclosed high-rise community, the peak in histogram almost locates at same position. The right tail of example in C5 is thicker than which of in C3, meaning a high frequency of higher wind speed occurrence in site of C5. It indicates that for low and mid-rise community, reducing enclosure at site boundary will improve overall wind speed dedicatedly, shown as thicker right tail of the same positioned single peak in histograms compared to which for C3.

By comparing example sites of C3 (S1101) versus C1 (S08), C5 (S04) versus C2 (S260202), C2 (260202) versus C1 (S08) and C5 (S04) versus C3 (S1101) (Figure F10), it is noticeable that: low and mid-rise enclosed site has higher overall wind speed than high-rise enclosed site. Less enclosed high-rise site has even higher overall wind speed than the both do. Less enclosed low-rise site has lower wind speed than high-rise site in area of low wind speed level, but has higher wind speed in area of high wind speed level. The turning point is at percentile 60. Between curve of S08 and S260202, the difference at low wind speed level is large than that of high wind

speed level. On the contrary, between curve of S1101 and S04, the difference at low end is smaller than that at high end.

Figure F9 Representative Site Sub-Groups of Cluster C with Their Wind Maps, Histograms and Building Height Distribution (S08, S260202, S1101, S04)

By comparing example sites of C3 (S1101) versus C1 (S08), C5 (S04) versus C2 (S260202), C2 (260202) versus C1 (S08) and C5 (S04) versus C3 (S1101) (Figure F10), it is noticeable that: low and mid-rise enclosed site has higher overall wind speed than high-rise enclosed site. Less enclosed high-rise site has even higher overall wind speed than the both do. Less enclosed low-rise site has lower wind speed than high-rise site in area of low wind speed level, but has higher wind speed in area of high wind speed level. The turning point is at percentile 60. Between curve of S08 and S260202, the difference at low wind speed level is large than that of high wind speed level. On the contrary, between curve of S1101 and S04, the difference at low end is smaller than that at high end.
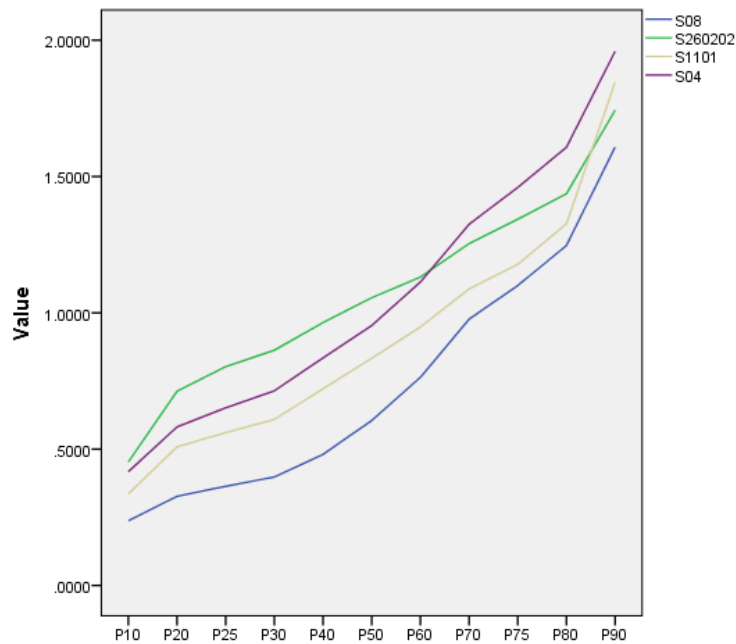

Figure F10 Curve Chart of trend by P10-P90 from example sites for C1, C2, C3 and C5

Regarding to the location of the peak, namely the wind speed majority of area has, C2 is 67% higher than C1, C3 is 53% higher than C1, C5 is 12% higher than C3, and C2 is 19% higher than C5 (table F1).

Table F1 Key Parametric of C1, C2, C3 and C4

Cluster	C1	C2	C3	C5
Example	S08	S260202	S1101	S04
Explanation	High-rise, enclosed	High-rise, less enclosed	Mid-rise, enclosed	Mid-rise, less enclosed
Median	0.6051	1.0555	0.8336	0.9535
Peak Location	0.35	1.05	0.75	0.85

Only one exceptional site 1801 is cluster in C4. Site 1801 is one of the largest sites and consists of mid-rise building. Although no long low-rise facade wrapping the site, the dense distribution causes very poor wind speed condition in site, even with an open space arranged in the site. This case shows that very dense low and mid-rise building arrangement could cause very poor wind speed performance as well as high-rise community of less density. Additionally, larger size of the site will enhance the impact of dense low and mid-rise buildings on wind speed. The appearance of its histogram is very close to which in C1 of worst wind speed condition (Figure F11).
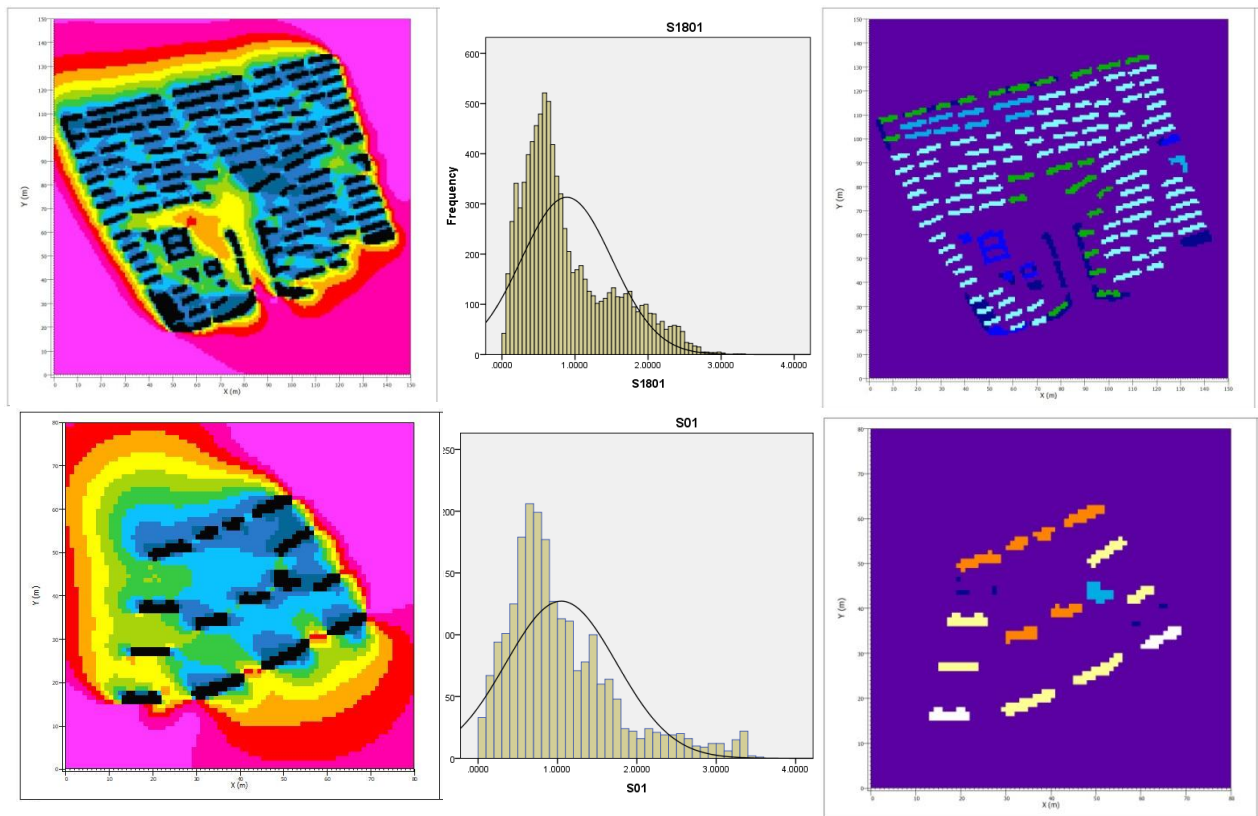
Figure F11 Wind Map, Histogram and Building Height Distribution of Site 1802 and S01

The other reason for the poor wind condition in site 1802 may be that the orientation of long facade array is almost perpendicular to the dominant wind direction. Another counter example is site 01 clustered into C5. Sites with similar building distribution attributes as site 01 of high-rise community, less density, and not enclosed are cluster into cluster A of good wind speed performance. The special characteristic of site 01 is

the long facade array orientated perpendicularly to dominant wind direction (Figure F11).

These two examples raise the necessity of further discussion regarding to site size and orientation of building array. Site size will be discussed in following sections. Orientation of buildings will be discussed in correlation between wind performance indices and building distribution indices. Because the dominant wind direction in SU-ZHE-WAN region is south-east which overlaps main sunlight direction, few sites having building long facade parallel to wind direction as research samples for qualitative observation.

All aforementioned comparisons and analyses lead to the following conclusions:

1. Under highly enclosed condition, low and mid-rise community has higher overall wind speed than high-rise community. However, under less enclosed condition, high-rise community is considered better due to higher wind speed in low wind speed level.

2. In high-rise site, less enclosure help more on improving wind speed in low wind speed level; however, in low and mid-rise site, decreasing enclosure devote more on wind speed in high wind speed level.

3. Decreasing degree of enclosure is more influential on improving wind speed in high-rise community than in low and mid-rise community.

4. Decreasing degree of enclosure is more influential than decreasing building height under same condition on improving in overall wind speed in site.

iii. Phenomenon and Reason in Wind Map and Histogram of Good Performance Cluster (A)

Cluster A shows generally good wind speed condition, and it has 3 sub-groups, of low/mid-rise buildings, high-rise buildings and sites with large open space. Generally speaking, sites in cluster A has loose building distribution or has open spaces.

Sub-group A1 consists of low and mid-rise buildings, with slightly less density compared to sites in C3. Clearly seen from wind maps, sites in A1 have more or less some open spaces inside the site. The less dense site with open space gives looseness for wind to circulate and to recover wind speed. Its histogram appears a single peak with axis at 1.3m/s and relatively sharp drop on both sides. Larger the open space, wider the peak in histogram, especially thicker right tail, as seen in comparison between site 2101 and 2102 in cluster A1 (Figure F12). The examples show that in low and mid-rise site, larger building separation is encouraged and inserting open space inside site could be even better for improving wind speed.
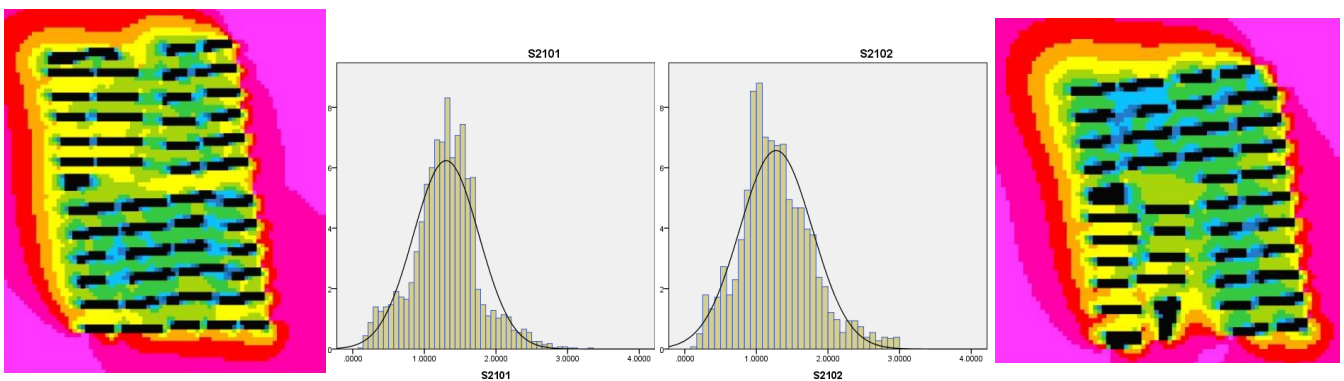

Figure F12 Comparison between Site 2101 and 2102 from A1 of Their Histograms and wind map

High-rise buildings are main component of sub-group A2. Compared to A1, in the histogram, the peak axis locates around 1m/s and the tail on both side drops rapidly. From the wind maps of A2, high-rise buildings tend to have significantly long wind shadow behind the building. This is especially obvious behind the last row or the most north row of buildings, as seen the blue and green area behind last row of buildings in figure F13. This characteristic is unseen in wind maps of low and mid-rise

communities. This also means high-rise sites actually have strong influence to the environment regarding to wind speed, but part of the low wind area is cropped outside the outline of the target site where is not considered and reflected in the research.

Comparing sites in A2 versus A1, high-rise community has larger building separation distance, due to larger obligative distance requirement for high-rise community than for low and mid-rise community, of sunlight availability at ground floor window in China. There is no doubt that high-rise community must be less dense as low and mid-rise community, but high-rise buildings also cast much larger and longer wind shadow behind. That is the reason that cluster A2 does not show great improvement in wind speed compared to A1: the large building separation is compensated by large no-wind area behind high-rise buildings.
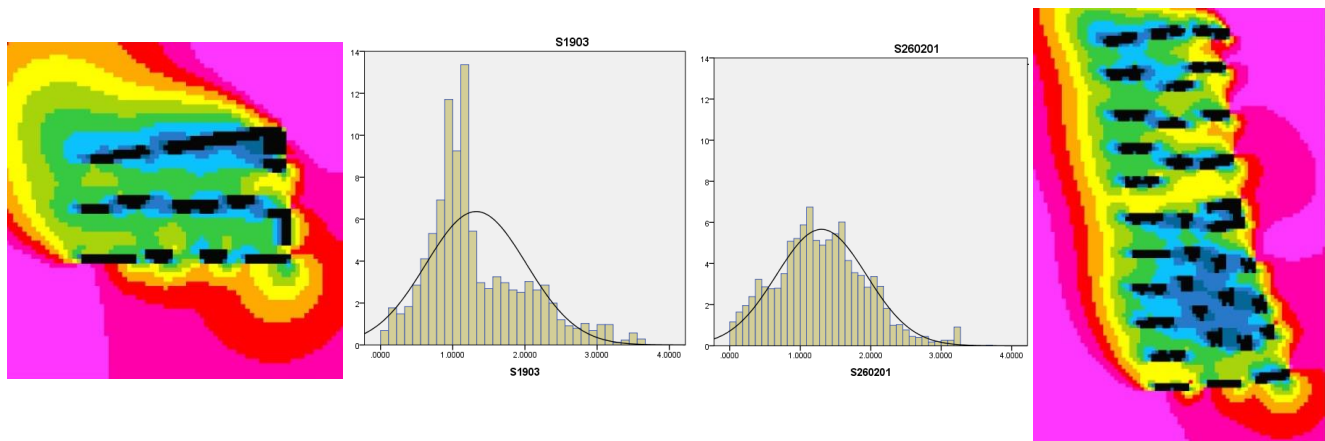

Figure F13 Wind Maps and Histograms of Sites 1903 and 260201 from Cluster A2

Another comparison within A2 of high-rise communities shows small site (S1903) performances better than larger one (S260201) regarding to wind speed (Figure F13). Although the peak locates at similar position in histograms, S1903 has thicker right tail, and stretches further until close to 4m/s. One reason for this phenomenon is that small sites have less accumulative effect of no wind area. Meanwhile, small site has high relative circumference compared to large site. Namely, this leads to lower facade ratio at site boundary if with same facade length and intervals between facades, to allow more wind enter the site.

To spot influence from size of site, a comparison is done between groups of cross-mixed attributes of low-rise, high-rise, small and large site (Figure F14). S2102 and S29 are of low-rise buildings with good wind performance, seen from maps and histograms (Figure F12 and F14). S29 is about half smaller than S2102 in size. Histogram of S29 is holistically right moved compared to which of S2102. This is echoed by curve of P10-P90: which of S29 is almost parallel to which of S2102 but slightly close at P10, and which of P29 is slightly higher. Table F2 indicates that S29 is generally 6% higher than S2102. All these lead to conclusion of in low and mid-rise site, small site size could provide higher overall wind speed.

However in high-rise site, smaller site (1903) have higher wind speed at low wind level section, and high wind level section, but lower wind speed at middle wind level section compared to S260201. Turning points are around P30 and P75. From the table F2, S260201 has higher median as in the curve chart, and the peak is moved 16% to the right than the peak of S1903. So it is proper to say that in high-rise site, smaller and larger site size will lead to difference wind speed distribution and further assessment requires detail condition of each site.
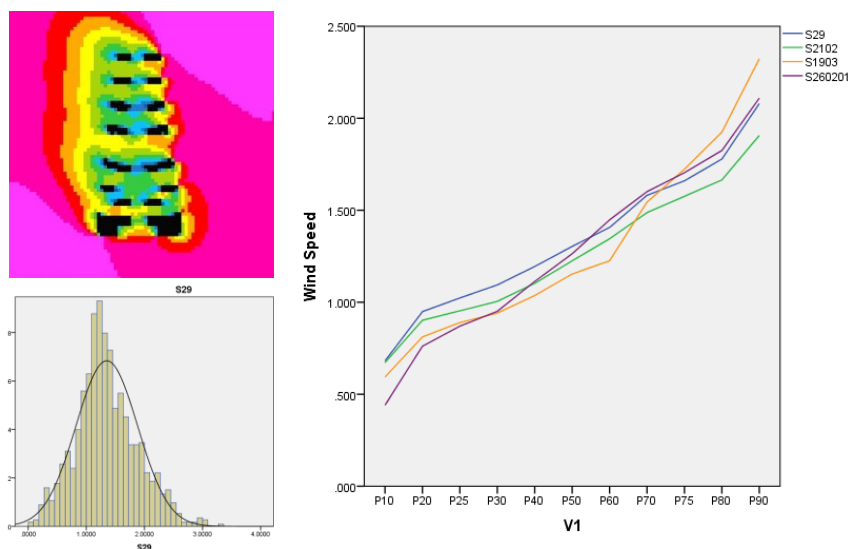

Figure F14 Wind Map And Histogram of S29 from A1, And Curve Chart of P10-P90 from Example Sites from A1 and A2

Table F2 Key Parametric of cross comparison sites from A1 and A2 regarding to site size

Cluster	A1	A1	A2	A2
Example	S29	S2102	S1903	S260201
Explanation	Small, low-rise	Large, low-rise	Small, high-rise	Large, high-rise
Median	1.3036	1.2252	1.1535	1.2635
Peak Location	1.30	1.15	1.05	1.25

Sub-group A3 is the same as cluster B in nature, so it will be discussed together with B.

Therefore, based on the above analysis on cluster A with good wind performance, the conclusion is listed below:

1. In low and mid-rise site, larger building separation and open space inside site could help with wind speed, and having open space is more effective.

2. In low and mid-rise site, small site size could provide higher overall wind speed.

3. In high-rise site, smaller and larger site size will lead to difference wind speed distribution. Assessment need to be based on details of each individual site.

4. In high-rise buildings actually have strong influence to the environment regarding to wind speed, but part of the low wind area is cropped outside the outline of the target site where is not considered and reflected in the research.

iv.   **Phenomenon and Reason in Wind Map and Histogram of Bipolar Cluster (B)**

Seen from the histograms based on wind speed data, most cases shows single peak, however, some characteristics leads to bipolar histograms. Cluster B has two peaks on histograms due to large open space existing in site outside of all building clusters (Figure F15). The left peak shows the poor wind condition inside building clusters,

and the right peak shows the open space is covered in high wind speed. There are two possible locations of open space in sites of cluster B: one is in front of the first row of buildings, which is to the south of all buildings (e.g. S260502); the other is very large open space to the north of all buildings (e.g. S1802).

Sub-group A3 is essentially the same as cluster B with difference in site detail arrangement which causes difference in appearance of histograms. Sites in A3 also have open space, not on the south of the site, but partly wrapped around all buildings, for example S1503 on north and west sides. This can also be interpreted as large separation distance from the road, due to traffic noise avoidance. Based on the comparison of histograms in A3 to B, slices of open spaces wrapping around site hardly form a clear right peak indicating high wind speed isolated from the left peak. Histograms of A3 appear trapezoidal, which is actually two peaks merged together, as seen in 7.12.

From the maps in 7.12, S260502 and S1802 both have relatively poor wind speed performances. The reasons are respectively large and wide building mass on the north edge of site blocking wind from flowing out and very dense mid-rise community with enclosed outline. The data from area of very low wind speed form the left peak in histogram, where the axis is around 0.5m/s.

The right peak in histogram of S260502 is a reflectance of the existence of the open space on the south of the site, where wind enters the site without any obstacles at initial wind speed of 3m/s (as seen in map as bright pink). However as influenced by the front row of buildings causing wind speed reduction, the majority area shows as dark pink and red in the open space. This is the formation of the right peak in its histogram.

Differently, the right peak in histogram of S1802 reflects the high wind speed area behind the northest row of building, majority shown as red in wind map. Because

S1802 consists of low and mid-rise buildings, especially the northest row of buildings being mid-rised, not any long wind shadow is cast on the open space at north. Namely, the wind bypasses the site easily and reaches a high speed in the open space at north, by which to allow a clear right peak shown in histogram.
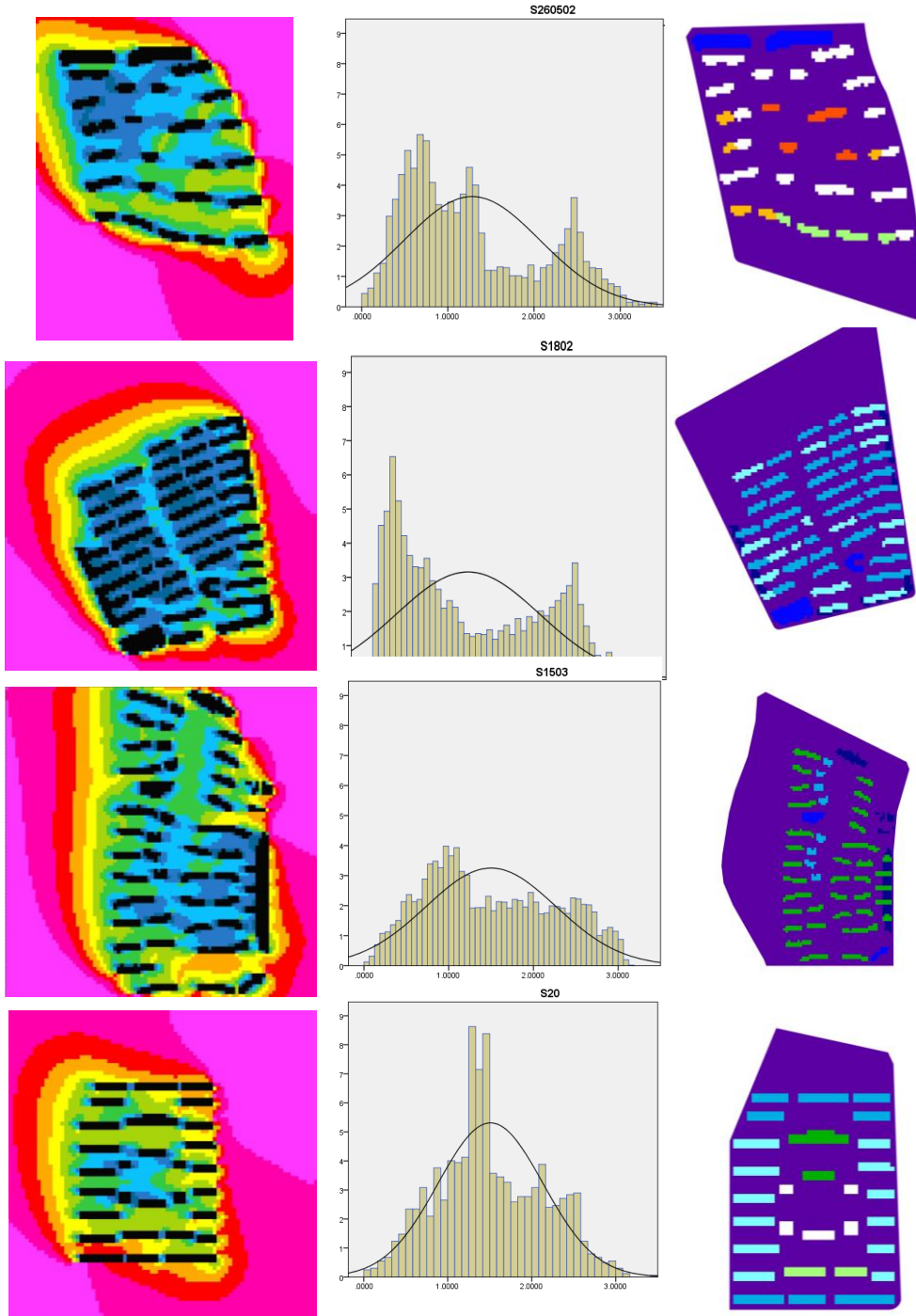

Figure F15 Wind Maps and Histograms of S260502 and S1802 from Cluster B
and S20 and S1503 from A3

Site 20 from sub-group A3 shares similar condition of mid-rise buildings and open space on the north. However, the open space in S20 is not as large as the one in S1802, so that the wind speeds at majority grid point have not reached 2.1m/s within the open space. This causes the histogram of S20 showing a tendency of two peaks, as two close-located peaks get merged (Figure F15). As aforementioned, high-rise buildings will cast very long wind shadow. If the site with north open space is of high-rise buildings, the open space is barely possible covered by wind speed over 2.1m/s (red area in map). So the right peak in histogram will not occur in site of high-rise buildings, only if the open space is even deeper than the wind shadow depth. However, under the control of economic strategy, such large open space compared to the built area is not practical.

It can be concluded that open space on the south and east side of the site tends to lead clearer bipolar-like histogram than open space on the west and north side of the site. Sites in A3 also echo with the conclusion.

However, the bipolar histogram should not be encouraged as a representative of good wind condition. Because the high peak is actually only devoted by the large open space, rather than the space between building clusters. However the aim is to improve wind condition in between buildings, as indoor ventilation highly relies on the wind field around building and outdoor space around building requires enough wind to remove emission from living activities.

To optimise building distributions to make wind field in site more even will be more helpful in improve wind speed performance between buildings. This refers to a single but wide spread peak with right skewed axis in histogram of wind speed. Rather than seeking for the second peak of high wind speed for pretty presentation in design, aiming for less left skewed wide single peak does more for wind-healthy residential ward. But for existing site with poor wind condition, or new design limited by high

development intensity which will definitely cause poor wind, raising ratio of non-segmented open space at least provides area of high wind speed in site.

Based on analysis of sites with bipolar histograms, some conclusions are drawn as below:

1. open space on the south and east side of the sites will generate clearer bipolar histogram of wind speed for a site, than located on the north and west side.

2. bipolar histogram could not be applied as a sign or criterion of good wind speed performance.

3. for new built residential ward, building distribution of a histogram with less-left-skewed wide single peak is more encouraged; for existing site with need of improvement in wind speed and new design limited by high development intensity, building distribution of non-segmented open space with bipolar histogram is encouraged.

Appendix G

Acoustic Performance Simulation Results and Analysis of Street Facing Facades

I. Clustering and Selection of Statistical Measures of Street-facing Facade Attenuation

Street-facing facade attenuation is evaluated by sound pressure level difference in-front-of and behind the facade. Sound pressure level in front of street facing facades is dominated by sound source level and reflections from surroundings, while the SPL behind the street facing building away from sound source shows the degree of quietness in the noise shadow behind building facade. The difference of the two SPL is an indicator of how well the street facing buildings barrier traffic noise, noted as street-facing facade attenuation (SFA)

To acquire SPL difference, SPLs are collected from the pairs of receivers along both sides of the street-facing buildings along each edge of all sites. Based on the data collected from pairs of receives, statistical measures could be calculated for following analysis. The adopted measures are: 1) average value of SPL differences from pairs of receivers on one edge of a site, which is an indicator of average noise barrier ability of street facing buildings on one edge of a site; and 2) maximum value of SPL differences from pairs of receivers on one edge of a site, which is an indicator of maximum noise barrier ability of street facing buildings due to certain arrangement on one edge of a site. These two indices are used for cluster of site and regression (table G1).

Average value of the SPL in front of building facade over one edge of a site, which is indicator of average impact on one edge from sound source level and reflection from

surroundings, is also tested in regression. The result shows that the SPL in front of street facing facade would be increased by strong sound source level, short distance to road and long street-facing facades (due to strong noise reflection).

Table G1 Summary of On-Edge Acoustic Performance Variables Applied in Research

Parameter	Abbreviation	unit	range		Definition	Reason
Average	SFA-avg	dBA	4	31	Average SPL difference from receiver pairs in front of and behind street facing buildings on one edge of a site	An indicator of average noise barrier ability of street facing buildings on one edge of a site
Maximum	SFA-max	dBA	6	36	Maximum SPL difference from receiver pairs in front of and behind street facing buildings on one edge of a site, usually locates in the middle of long facade or enclosed corner	An indicator of maximum noise barrier ability of street facing buildings due to certain arrangement on one edge of a site

II. Analysis of Clustering of Sites Edges by Street-facing Facade Attenuation Performance and Discovery of Qualitative Rules of Distribution

i. Clustering Analysis of 6 Clusters by SFA-max and SFA-avg

Similar to clustering analysis of sound pressure level of grid points, hierarchical clustering is conducted for all edges from sample sites based on street-facing facade attenuation mean and maximum values. The edges are divided into 6 groups. From group 6 to group 1, respectively are edges of very high SFA-max and SFA-avg, high SFA-max and medium-high SFA-avg, high SFA-max but medium SFA-avg, medium SFA-max and fair-medium-high SFA-avg, fair SFA-max and SFA-avg, and poor SFA-max and SFA-avg.

Group 6 is of highest level of boundary attenuation, with max street-facing facade attenuation of 33-35.2dBA and mean value of 25-30dBA. The number of edges in the cluster is 7 accounted for only 4.0% of total amount. The building arrangement in this cluster appears long continuous facade in medium facade ratio parallel to road, located very close to strong sound source. Buildings are high-rise with small building

depth and wide intervals, namely the low interval depth is small. Another condition happens to edges perpendicular to building long facades, also located close to sound source. The edge has gable walls of residential buildings, but enclosed on the edge and corner by commercial buildings of long facades. High interval depth on the edge is actually the length of continuous facade. It also further separate the pair of receives causing higher value of SFA-avg and SFA-max. The second condition shows that the edge of gable walls could perform even better in on edge attenuation than edge of longer south facing building facade parallel to road, only if sufficient enclosure exist. Overall, enclosed yard has best SFA level considering both mean and maximum value.

Group 5 of high SFA-max level, and medium-high SFA-avg level, has 28 edges accounting for 15.9%. The range of maximum SFA is 28.3-34.1 dBA, and of average SFA is 19.5-26dBA. This cluster accommodates the mainly edges perpendicular to long residential building facade, namely east-west facing edges, with some south-north facing edges. The east-west facing edges share narrow facade (gable wall) facing road and large high interval depth, some extent of enclosure with commercial buildings, smaller interval width, and medium density. Large high interval depth separates the receiver pair further apart would result to higher level of SFA. Another condition of edge in this cluster is south-north facing, not close to road, consisted of medium length facade, and either having a building of big mass at corner resulting large high interval depth, or enclosed at corner. In general, some degree of enclosure especially at corner is needed to ensure considerable amount of on edge attenuation.

Group 4 has high level SFA-max of 27.9-34.5 dBA, and medium level SFA-avg of 11.8-18.2 dBA. The counts are 16, accounting for 9.1%. The edges in this cluster appear medium average SFA, however the maximum value of SFA is high. It means that for some reason, the edge has high local on-edge attenuation. This cluster accommodates most of the edge facing east and west, with residential buildings perpendicular to the road. The distribution characteristics are of short facade building

or of wide intervals, but with SFA boost of a section of long continuous facade or building of big mass or enclosure at one corner. There exist another scenario in this cluster that is east-west facing edge in small enclosed site influenced by strong sound source from south and north. In this case, the noise impact results in average SFA limited to medium level.

From group 3 to group 1 the attenuation performance becomes below satisfactory. Group 3 has medium SFA-max level at 20.8-28.3dBA, and full range (fair-medium-good) of SFA-avg level at 9.9-22.1 dBA. 55 edges locate in this cluster accounting for 31.2% of total amount. The components in this cluster include north-south facing and east-west facing edges. They share medium enclosure and continuous facade length. For east-west facing edges, the high interval depth reaches medium level.

Group 2 appears in fair SFA-max (9.8-22.6dBA), and SFA-avg (5.69-14.33dBA) level, which contain the majority of edges of 66 in number, accounting for 37.5%. It consists of edges of open boundary and of short facade length. For east-west facing edge, the high interval depth is below medium level.

Group 1 is of poor performance in both SFA-max (below 10dBA) and SFA-avg (around 6dBA). Only 3 edges in this scenario, accounts for 1.7% of total amount. These edges are east-west facing and of the narrow gable wall of residential buildings. Furthermore, the building separation distance is large due to nature of high-rises, resulting to wide intervals on the edge.

ii. Consolidation of The Clustering and Qualitative Rules

Table G2 Consolidation of Clustering Analysis of Samples Base on SFA Mean and Maximum

Cluster	Parameters		Site Description	Example	conclusion
Cluster 6	SFA-max Range (dBA)	33-35.2dBA	North-south facing: • Long continuous facade • Medium facade ratio • Very close to road • Strong sound source • High-rise • Low interval depth is small East-west facing: • Edge of gable walls • Enclosed on the edge and corner by of long facade commercial buildings	S12R1 S1801R1 S1801R3 S1802R1 S1802R4 S22R1 S22R3	• The edge of gable walls could perform even better in on edge attenuation than edge of longer south facing building facade parallel to road, only if sufficient enclosure exists. • Enclosed yard has best SFA level considering both mean and maximum value.
	SFA-avg Range (dBA)	25-30dBA			
	Proportion	4.0%			
Cluster 5	SFA-max Range (dBA)	28.3-34.1	Mainly east-west facing: • Short facade (gable wall) • Large high interval depth • Some extent of enclosure with commercial buildings • Smaller interval width • Medium density. A few south-north facing: • Not close to road medium length facade • Either having a big mass building at corner resulting large high interval depth, or enclosed at corner.	S04R1 S08R1 S08R2 S09R2 S09R4 S1101R3 S1101R4 S1102R5 S1102R7 S12R2 S14R2 S1501R1 S1501R4 S1801R2 S1802R2 S20R1 S20R3 S2101R2 S2101R4 etc.	• Some degree of enclosure especially at corner ensure considerable amount of on edge attenuation.
	SFA-avg Range (dBA)	19.5-26			
	Proportion	15.9%.			
Cluster 4	SFA-max Range (dBA)	27.9-34.5	• High local on-edge attenuation • Mostly accommodate east-west facing edge • Of short facades or of wide intervals • A section of long continuous facade or corner enclosure or having one high mass building • East-west facing edge in small enclosed site influenced by strong sound source from south and north	S04R2 S05R1 S05R3 S09R1 S10R3 S10R4 S14R1 S1502R4 S23R3 S260102 R7 S260202 R7 S28R2 S34R3 etc.	• Local enclosure or large building mass would boost maximum SFA although average SFA is in medium level.
	SFA-avg Range (dBA)	11.8-18.2			
	Proportion	9.1%			
Cluster 3	SFA-max Range (dBA)	20.8-28.3	• Medium enclosure • Medium continuous facade length • For east-west facing edges, the high interval	S04R1 S08R1 S08R2 S09R2 S09R4	
	SFA-avg Range	9.9-22.1			

	(dBA)		depth reaches medium level	S1101R3 etc.	
	Proportion	31.2%			
Cluster 2	SFA-max Range (dBA)	5.69-14.33	▪ Open boundary ▪ Short facade length for east-west facing edge, the high interval depth is below medium level	S0201R1 S0201R2 S0201R4 S0202R5 S0202R6 S0202R8 etc.	▪ Lack of boundary enclosure would dramatically lost attenuation ability on the edge
	SFA-avg Range (dBA)	9.8-22.6			
	Proportion	37.5%			
Cluster 1	SFA-max Range (dBA)	<10	▪ East-west facing narrow gable wall of residential buildings ▪ Large building separation distance (wide intervals) ▪ High-rise residential buildings	S01R1 S01R2 S01R3 S01R4 S0201R3 S0202R7 S0301R3 etc.	
	SFA-avg Range (dBA)	<6			
	Proportion	1.7%			

Clustering results by SFA maximum and mean are consolidated in table xxx. A few qualitative rules could be summarised from the analysis:

1. Street-facing facades locate close to sound source could result in increase in both SFA-max and SFA-avg.

2. Enclosure is the most effective approach to achieve boost in facade attenuation. Enclosed edge constituted of gable walls of residential buildings, could produce higher SFA than less enclosed edge of long continuous facades.

3. Building of high body mass could improve SFA, because the large building depth could separate the receiver pair further apart.

4. Enclosed yard has best street-facing facade attenuation ability.

III. Regression of Street-facing Facade Attenuation Statistical Measure and Building Distribution Indices

i. Regression Parameter Selection

The regression of street-facing facade attenuation involves parameters representing noise barrier effect and building distribution pattern. The data for either of the two aspects are collected along edges of a site, instead of overall level from the whole site.

As discussed in Chapter 4 of building distribution indices, indicators of building distribution pattern includes overall average indices and boundary condition indices. The distribution parameters adopted for SFA regression are the boundary condition indices related to factors of facade enclosure (facade length and ratio, interval length and ratio, interval depth amount), and facade building height (average height over one edge). As for study of edges of site, the statistical measures of boundary condition indices are calculated over one edge of a site rather than over the whole site. The indices are listed in table G3

Table G3 Representative Building Distribution Indices for On-Edge Performance Regression

Group	Representatives	Abbreviation
Edge scale	Total outline length of one edge	TOL
Building height	Average facade height on one edge	FHmean
Facade amount	Total Façade Length on one edge	TFL
	Façade Ratio on one edge	FR
Interval depth amount	Average Low Interval Depth on one edge	IDL
	Average high Interval Depth on one edge	IDH
Interval amount	Average interval area on one edge	IAmean
Separation from road	Distance To Road from street facing façade	D
Sound source	Sound source level on one edge	SSL

The parameters of assessing SPL difference across street facing buildings are listed in table G1, as discussed in Chapter 3 of parametric study. The average is also calculated over each edge of all sites. Details of selected variables refer to table G1.

ii.   Regression Result of Average Street-Facing Facade Attenuation

Average street-facing facade attenuation (SFA-avg) is calculated as mean value of SPL differences from receiver pairs located across and along street-facing buildings. It is clear that to have higher level of noise barrier effect by buildings, the SFA-avg is expected as high as possible.

The regression of SFA-avg is just acceptable. The R square is 0.503 and adjusted R square is 0.485, with generalisation shrinkage of 3.6%. Coefficients are listed in table G4.

Table G4 Details of Regression of SFA-avg with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-30.998	8.447		-3.669	.000
TotalFacadeLength	.018	.004	.318	4.730	.000
AvgHeight	-.050	.013	-.231	-3.883	.000
AvgIntervalDepthLow	.122	.025	.320	4.820	.000
soundsource	.523	.113	.291	4.628	.000
DistanceToRoad	-.076	.016	-.306	-4.823	.000
FacadeLengthRatio	6.125	1.872	.255	3.272	.001

The regression shows to achieve high SFA-avg, it is required to have: long total facade length, large facade ratio, large lower interval depth, large sound source level, lower average building height on the edge and short distance to road. To summarise in other words, edge with high level of enclosure, low building height and compressed to close distance to stronger sound source could reach higher noise attenuation. Worth to mention that lower building height is related to longer and more enclosed building facade due to residential ward design habits: the low height commercial buildings tend to have long facade and wrapping around the site.

Close to strong sound source also helps to increase the SPL difference value however it is not the real method to adjust design to achieve quieter environment. But for case with strong sound source, and capable of sacrificing front facade in strong noise, compressing the separation distance to road is still very effective.

iii.   Regression Result of Maximum Street-Facing Facade Attenuation

Maximum street-facing attenuation shows the maximum ability of noise barrier by buildings on an edge of site. It usually occurs in the middle of the longest facade or at

the enclosed corner. High value of maximum SFA is desired for quieter inner acoustic performance.

However, the explanation power of the regression model on the variation of SFA-max is barely satisfactory. The R square is 0.479 and adjusted R square is 0.457, with generalisation threshold of 4.6%. Regression results are listed in table G5.

Table G5 Details of Regression of SFA-max with Representative Indices

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-32.980	10.819		-3.048	.003
AvgHeight	-.045	.016	-.166	-2.714	.007
AvgIntervalArea	-.002	.001	-.266	-3.084	.002
AvgIntervalDepthHigh	.151	.031	.431	4.906	.000
SoundSource	.581	.146	.259	3.989	.000
DistanceToRoad	-.079	.020	-.258	-3.920	.000
TotalOutlineLength	.023	.003	.423	7.381	.000
FacadeLengthRatio	8.366	2.266	.280	3.692	.000

As seen in the table, to achieve high SFA-max, it is required to have large higher interval depth, long total outline length, large façade ratio, strong sound source level, small interval area, small distance to road and short average building height. The most dominant indices are high interval depth and total outline length.

iv. **Consolidation and Discussion of Regressions of Street-facing Facade Attenuation**

The two regressions of SFA are compared with their standardised coefficients in table G6.

Table G6 Comparison of Standardised Coefficients of Regressions of SFA Mean and Maximum

Index Group		Edge scale	Building height	Facade amount		Interval depth		Interval amount	Set back	Sound source
Models	Expect	TOL	FHmean	TFL	FR	IDL	IDH	IAmean	D	SSL
SFA-avg	Large		-.231	.318	.255	.320			-.306	.291
SFA-max	Large	.423	-.166		.280		.431	-.266	-.258	.259

Comparing the difference in indices selection and in amount of their values, average street-facing facade attenuation shows stronger impact by short separation distance

and sound source level, represented by higher standardised coefficient of -0.306 on D and 0.291 on SSL compared to -0.258 and 0.259 for regression of SFA-max respectively. On the contrary, SFA-max is more sensitive to high facade ratio, large total outline length and small interval area on one edge with coefficients of 0.280, 0.423 and -0.266. This discrepancy could be applied in practice that if required boosting performance on quiet noise shadow behind street-facing facade, it is suggested to design long continuous facade in priority instead of locating street-facing buildings close to the road; vice versa if need average or even noise attenuation across street-facing buildings.

The interval depth parameter involves in two regression are different, average SFA adopted lower interval depth, while maximum SFA adopted higher interval depth. So it means noise leaking from passage at interval is very influential to maximum SFA which usually locates in the middle of long continuous facade. Therefore, building cluster forms large high interval depth could provide considerable protection from noise leaking which contribute to maximum SFA.

Comparison of acoustic regression result with thermal and sunlight, with its own noise maps

Table G7 Equivalent Change of Building Distribution to One Unit Improvement of On-Edge
Performance Variables

Independent Variables			SFA-avg +3dBA	SFA-max +3dBA
Cluster	Name	Range		
Edge scale	TOL	75-706		130.435
Building height	FHmean	3-152	-60.000	-66.667
Facade amount	TFL	23-587	166.667	
	FR	0.13-1.0	0.490	0.359
Interval depth	IDL	0-66	24.590	
	IDH	0-73		19.868
Interval amount	IAmean	0-6410		-1500.000
Set back	D	7-163	-39.474	-37.975
Sound source	SSL	73.4-84.6	5.736	5.164

The equivalent building distribution compromises are calculated for the two regressions of SFA. Compared to the range of each index, all compromises for one unit change of SFA are operable.

Decrease average building height around 60m, to achieve increase in SFA, should be interpret as involving in low-rise commercial and residential buildings with long facade which could improve barrier effect, rather than decrease residential building height located on the edge of site.

Increase facade ratio is very effective in improving the barrier effect, however the step size of 0.4-0.5 is significant when changing distribution pattern.

The step size of increase on low interval depth is 25m for 3dBA increase of SFA-avg, and on high interval depth is 20m for 3dBA increase of SFA-max. Increase interval depth is operable by applying L-shape buildings or to form L-shaped mass with multiple buildings.

Regression equations are consolidated in table G8.

Table G8 Consolidation of Regression Equations of SFA Performances

| Model	Regression Equation	R^2	Sig
SFA -avg	SFA-avg=0.018×TFL-0.050×FHmean+0.122×IDL+0.523×SSL-0.076×D +6.125×FR-30.998	0.503	.000
SFA -max	SFA-max=-0.045×FHmean-0.002×IAmean+0.151×IDH+0.581×SSL -0.079×D+0.023×TOL+8.366×FR-32.980	0.479	.000

Appendix H

Cluster Analysis of Feasible Parato Solutions of MD-MOO

The cluster analysis result according to average linkage between groups is presented with a dendrogram as below.
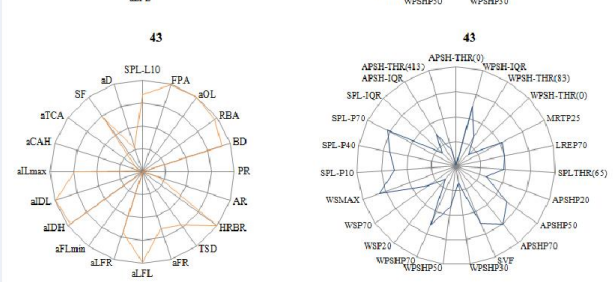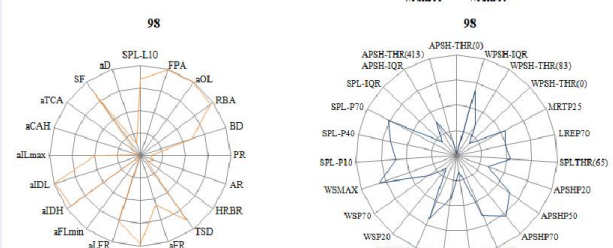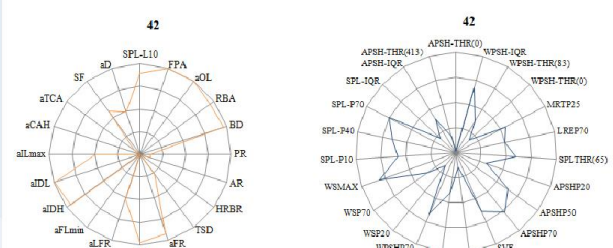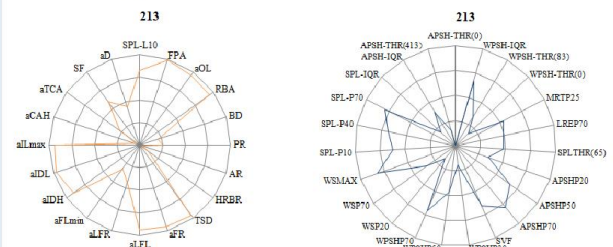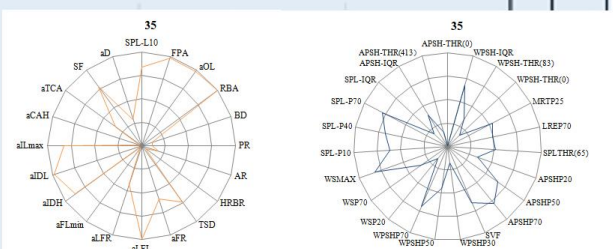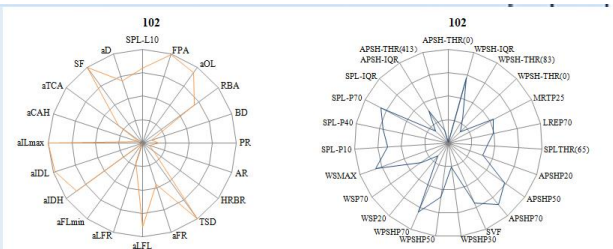

Dendrogram using Average Linkage (Between Groups)
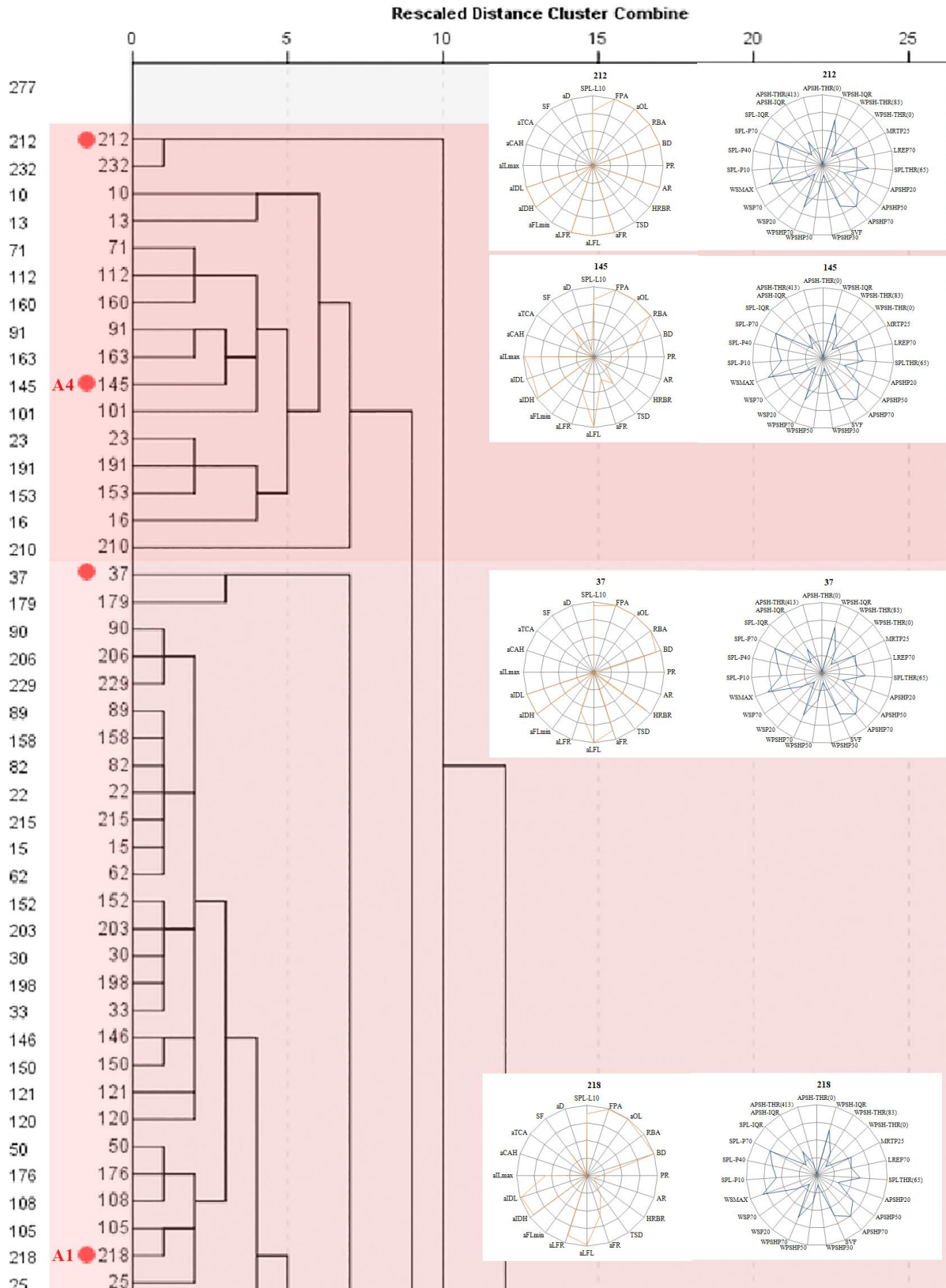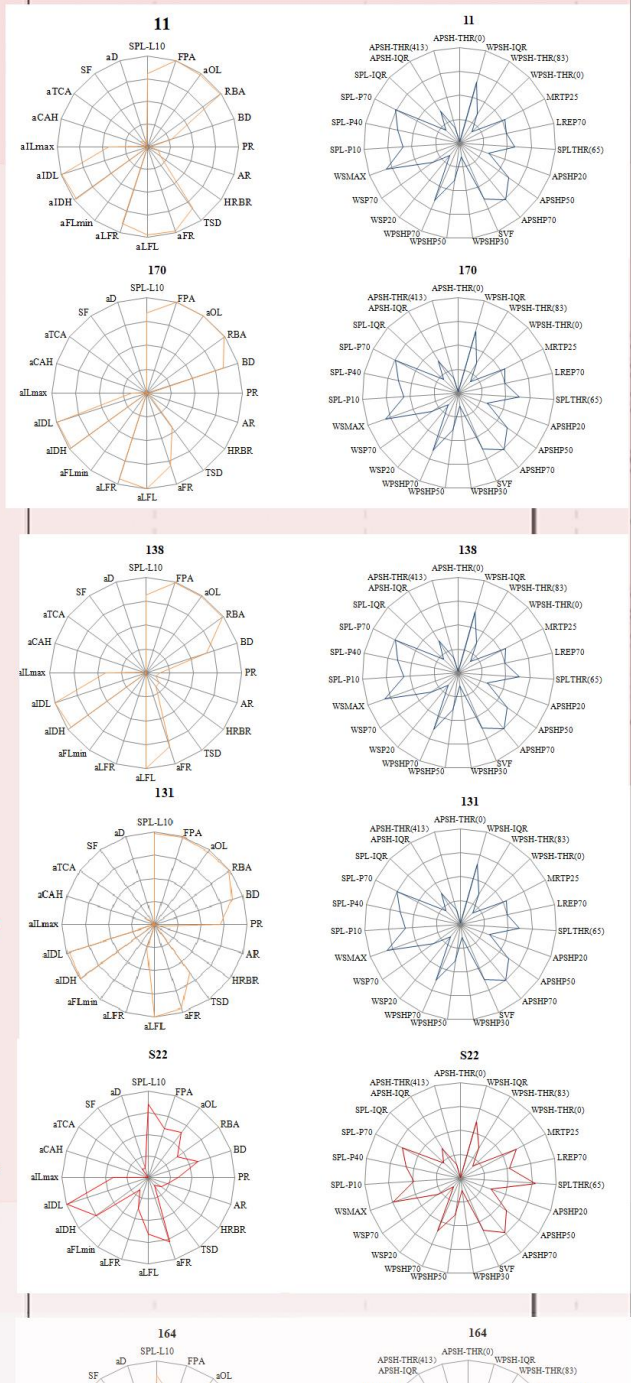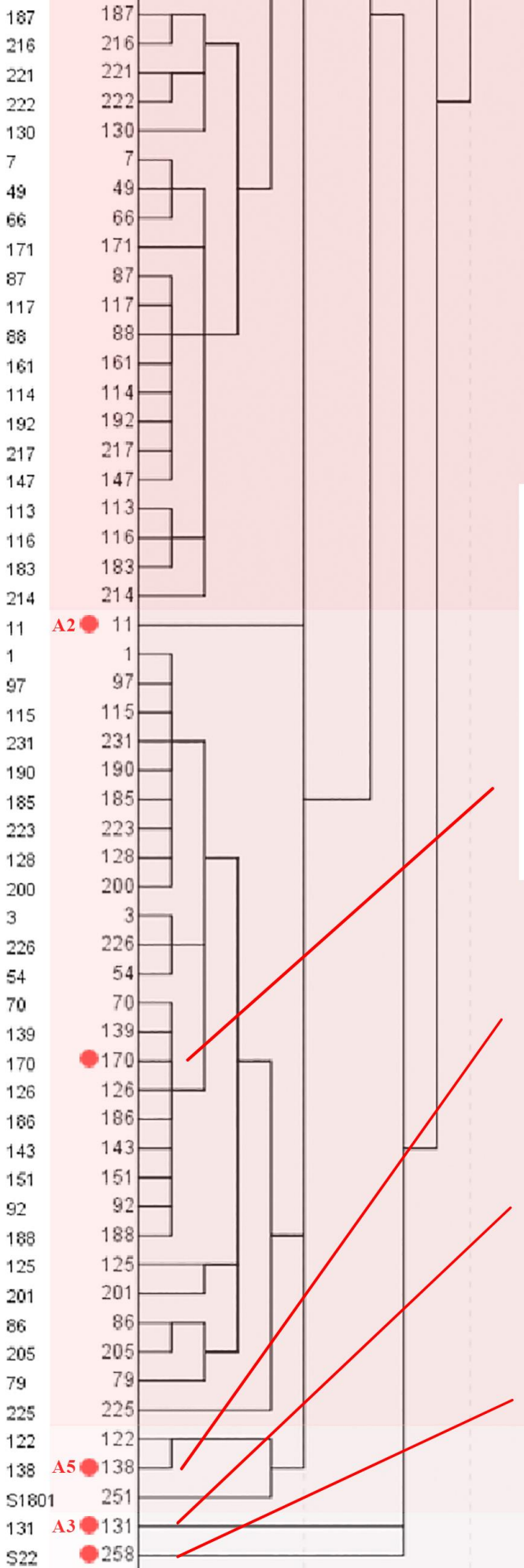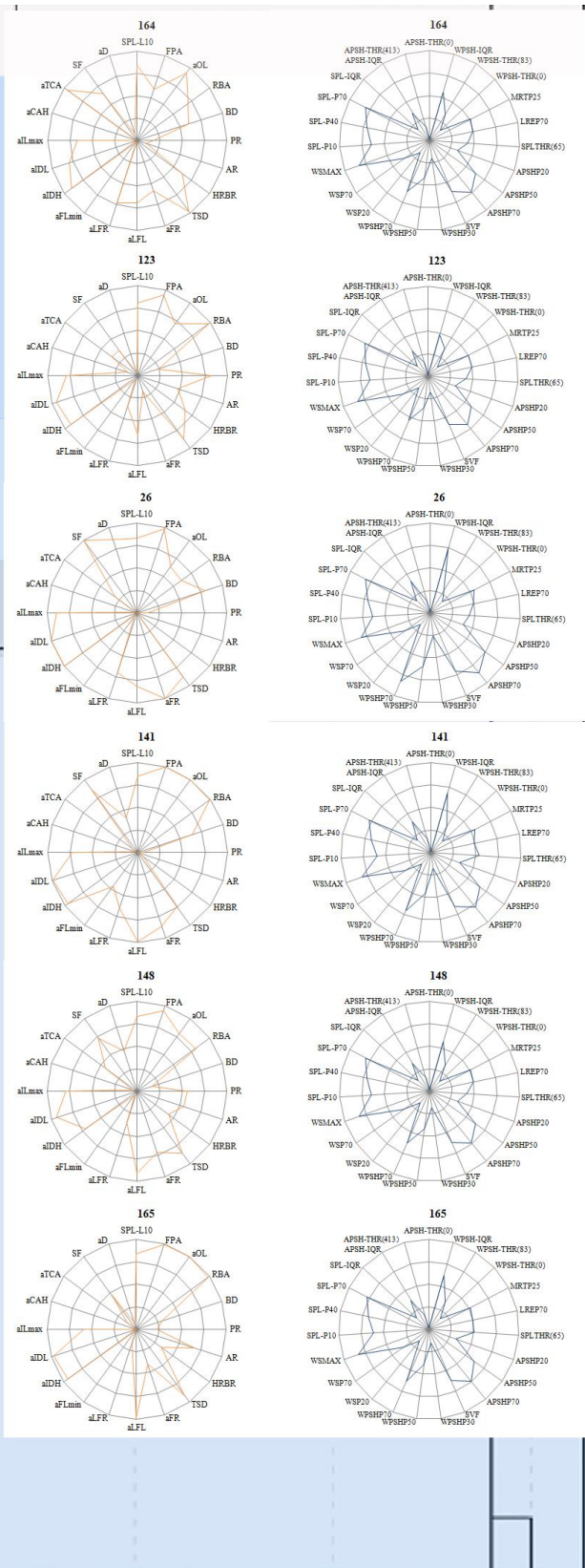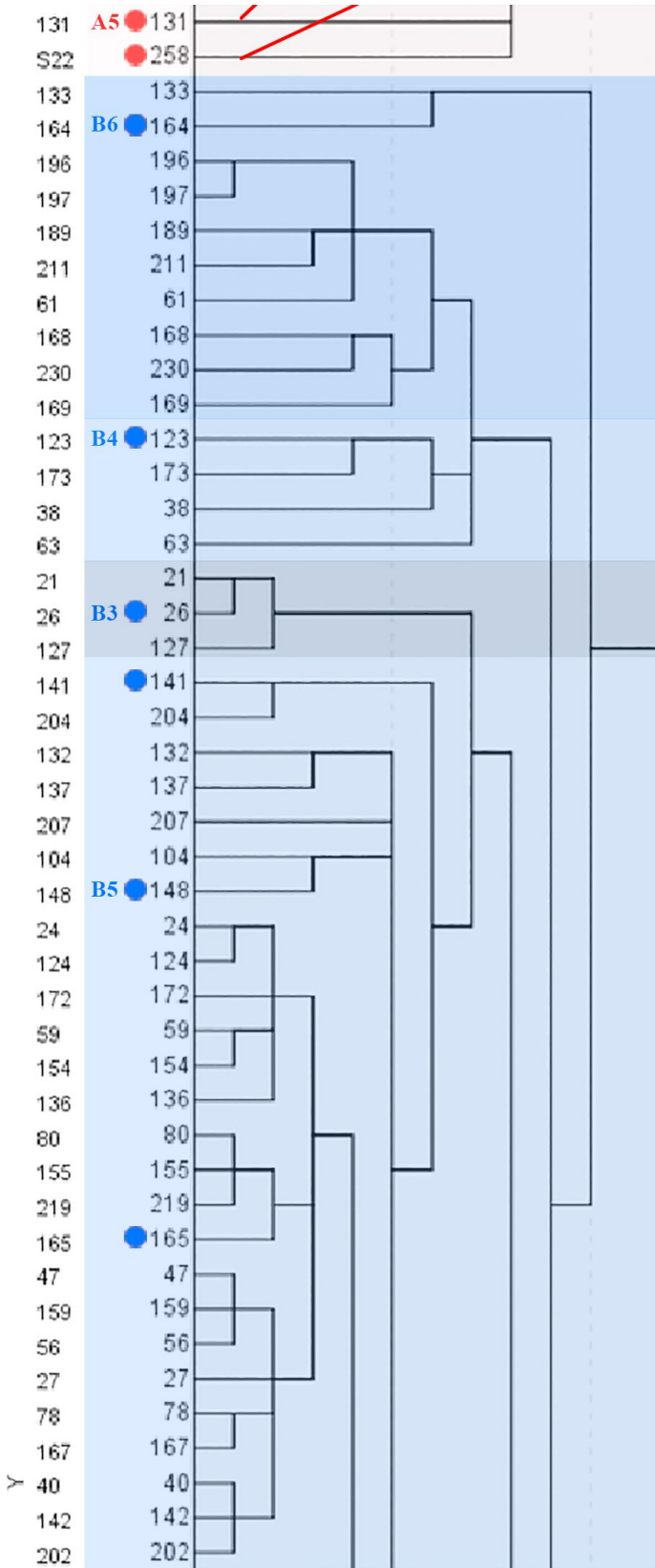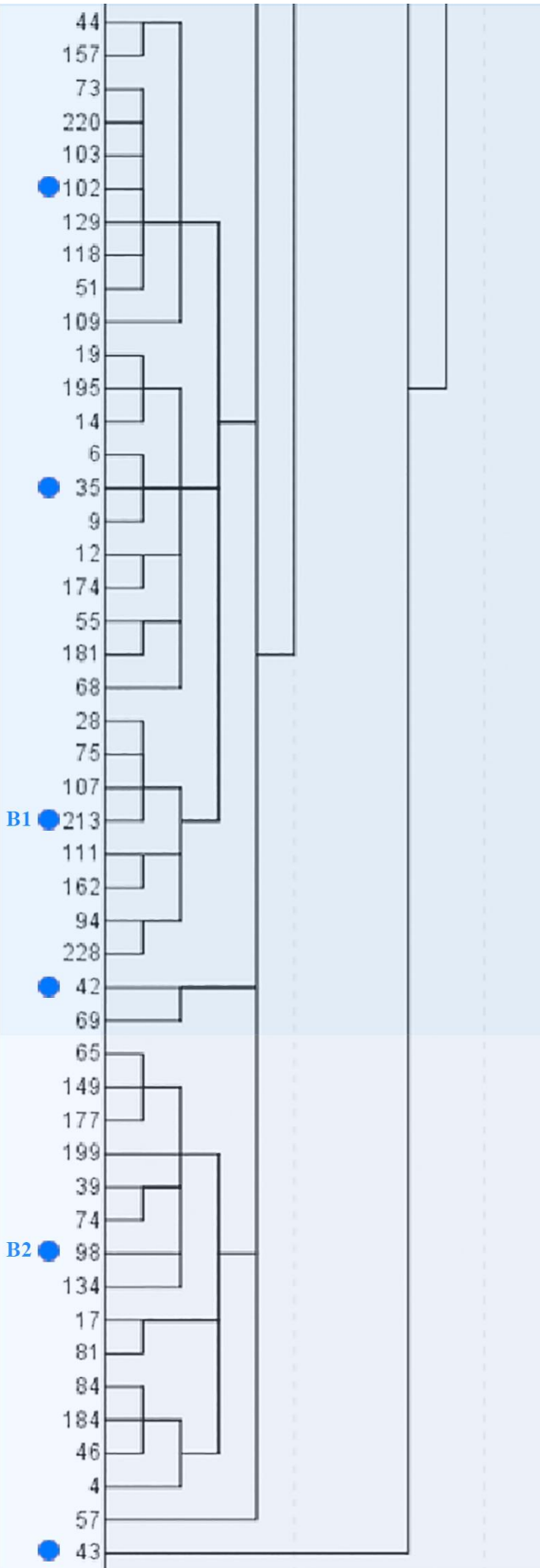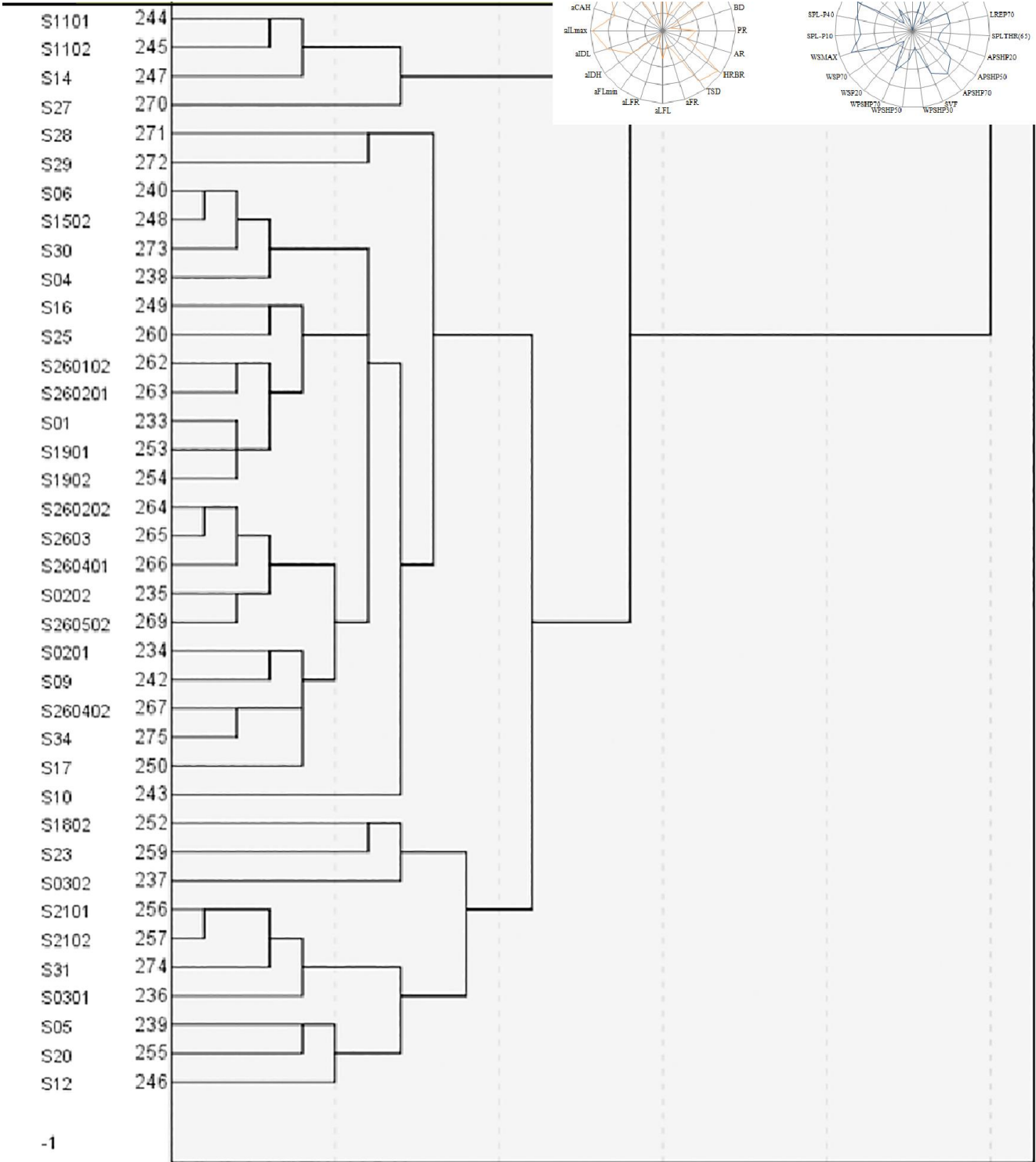
Abbreviations and Glossary

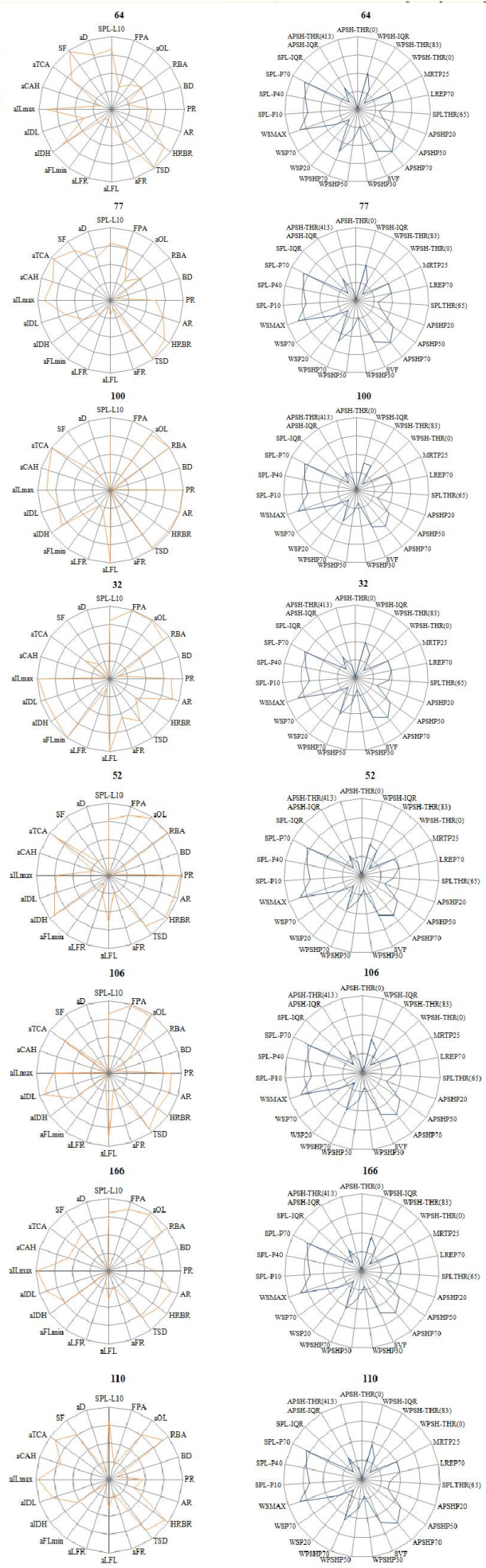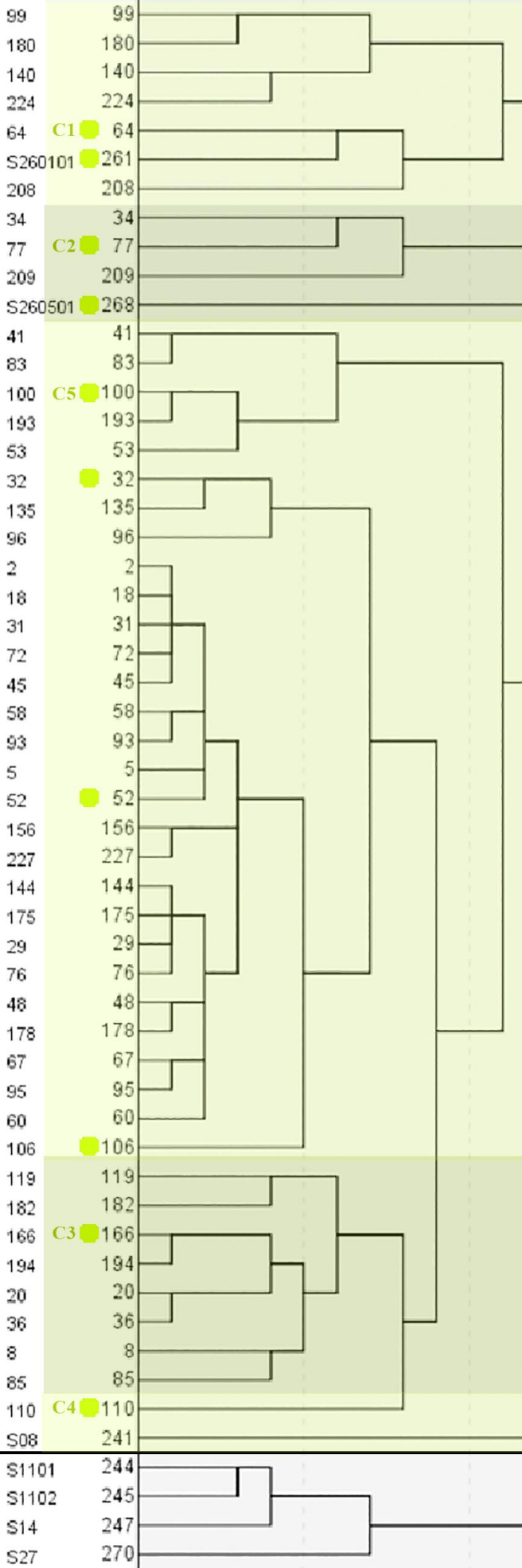|	
ACO	ant colony optimisation
AAO/SAND	appositional analysis optimisation/simultaneous analysis and design
ABC	artificial bee colony optimisation
aCAH/AvgCornerHigh	average corner area high value
aCAL/AvgCornerLow	average corner area low value
aCAmean/AvgCornerArea	average corner area
aD/AvgDistance/Dmean	average distance to road of a site
ADF	average daylight factor
aFHmean/AvgFrontHeight	average front façade height
aFL/AvgTotalFaçadeLength	average total façade length
aFLmax/AvgMaxFaçadeLength	average max façade length
aFLmean/AvgFaçadeLength	average façade length
aFLmin/AvgMinFaçadeLength	average min façade length
aFR/AverageFaçadeRatio	average façade ratio
aFSmean/AvgFrontStorey	average front façade storey
AHP	analytical hierarchy process
aIAmean/AvgIntervalArea	average interval area
aIDH/AvgIntervalDepthHigh	average interval length high value
aIDL/AvgIntervalDepthLow	average interval depth low value
aILmax/AvgMaxIntervalLength	average max interval length
AIO	all-in-one method
aLFL/AvgLowFaçadeLength	average low-rise façade length
aLFR/AvgLowFaçadeRatio	average low-rise façade ratio
ANN	artificial neural network
ANSI	American National Standards Institute
aOL/AvgOutlineLength	average outline length
API	application programming interface
APSH	annual possible sunlight hour
APSHF, WPSHF, SHWF	APSH, WPSH or SHW value at 1.6m facade level height grid
APSHG, WPSHG, SHWG	APSH, WPSH or SHW value at ground level height grid
APSH-IQR	interquartile range of APSH
APSH-P20	20% percentile of APSH grid value of a site
APSH-P50	50% percentile of APSH grid value of a site
APSH-P70	70% percentile of APSH grid value of a site
APSH-THR(0)	ratio of grid points with sunlight duration equal to 0 hours

APSH-THR(413）	ratio of grid points with sunlight duration below 413 hours
AR/AspectRatio	aspect ratio
aRS/AvgResiStorey	average residential storey
ASHRAE	The American Society Of Heating, Refrigerating And Air Conditioning Engineers
aTCA/TotalCornerArea	total corner area
ATHR(0)	ratio of grid points with sunlight duration equal to 0 hours
ATHR(413)	ratio of grid points with sunlight duration below 413 hours
BA	bat algorithm
BD/BuildingDensity	building density
BIM	building information modelling
BLISS	bi-level integrated system synthesis
BPNN	backward propagation neural network
BPS	building performance simulation
BRE BR	Building Research Establishment Building Regulation
BS	British Standardisation
Building Morphology Parameter	building geometry and morphology characteristics of single building and neighbourhood in residential building wards
CAD	computer aided design
CECS	China Association For Engineering Construction Standardisation
CFD	computational fluid dynamics
CIE	International Commission On Illumination
CMA-ES	covariance matrix adaptation evolution strategy
CO	collaborative optimisation
CRTN	calculation of road traffic noise
CS	cuckoo search
CSSO	concurrent subspace optimisation
DB	Chinese Regional Standards
DE	differential evolution
dendrogram	a tree diagram that is used to represent and categorize hierarchical relationships among objects, created as an output from hierarchical clustering
DL/DiagonalLength	diagonal length of site
Dmax	max distance of the front facade to the faced road in one site
DoCE	design of computational experiment
Domain Performance Metric	weighted summation of Ovs, interchangeable with SD-MOO

DV	design variable from optimisation viewpoint, input variable for MD-MOO. Normalised building morphology parameter
EA	evolutionary algorithm
EC	evolutionary computation
EIA	The U.S. Energy Information Administration
FPA/FootprintArea	foot print Area
GA	genetic algorithm
GB	Chinese National Standardisation (Guobiao)
GRNN	genetic regression neural networr
GRNN-PSO	hybrid generalised regression neural network particle swarm optimisation
GUI	graphical user interface
GWO	grey wolf optimiser
GWO-GRNN	genetic regression neural network hybrid with grey wolf optimiser
GWO-Kriging	improved Kriging with grey wolf optimiser
HCCI	Homogeneous-Charge Compression Ignition
HDE	hybrid differential evolution
HRBA/HighriseResiArea	high-rise building area
HRBR/HighriseRatio	high-rise ratio
IDF	individual disciplinary feasible method
ISA	interior search algorithm
JG	Chinese Professional Standard Systems
KH	krill herd
L(A)eq	the A weighted equivalent sound level
L10	a percentile at 10% of all data which suggests that in 10% of the time this value is exceeded during the whole record period and it is an indicator of the average peak value of the sound source.
LHS	Latin hypercube sampling
LMRBA/Low/medium-riseResiArea	low/medium-rise building area
LRE	longwave radiation from environment
LRE-P70	70% percentile of LRE grid value of a site
mae	mean absolute error
mape	mean absolute percentage error
MARS	multivariate adaptive regression splines
MBA	mine blast algorithm
MCDM	multi-criteria decision making
MDF	multi-disciplinary feasible method
MD-MOF	multi-domain multi-objective optimisation function, solving maximisation or minimisation for MD-MOO
MD-MOO	multi-domain multi-objective optimisation

MDO	multi-domain optimisation
Meta-model	estimative prediction model of relationship between DVs and OVs, mentioned as Dm (m=1, 2, 3), the estimation of SD-MOF in this research
MFO	moth-flame optimiser
MLR	multiple linear regression
MOCBO	multi-objective colliding bodies optimiser
MODE	multi-objective differential evolution
MOEA	multi-objective evolutionary algorithm
MOF	multi-objective optimisation function
MOGA	multi-objective genetic algorithm
MOO	multi-objective optimisation
MOPSO	multi-objective particle swarm optimiser
MOSOS	multi- objective symbiotic organism search
MRT	mean radiant temperature
MRT-P25	25% quatile of MRT grid value of a site
mse	mean square error
NDO	normalise domain objective
NSGA	non-dominated sorting genetic algorithm
NSGAII	non-dominated sorting genetic algorithm with elitist strategy
NSGWO	non-dominated sorting grey wolf optimiser
OAT	one-at-a-time approach of simulation
OGBA/OvergroundBuiltArea	over ground building area
OV	objective variable from optimisation viewpoint, input variable for MD-MOO. Normalised performance indices
P10, P20, P30, P40, P50, P60, P70, P80, P90	ten percentiles of all collected simulation data over one grid
P25, P75	two quatiles of all collected simulation data over one grid
PBD	the performance-based design
PCA	performative computational architecture
PDD	predictive percentage dissatisfied
pdf	probability density function
Performance Index	a statistical measurement selected for cross domain outdoor environmental performance index values on a certain simulation grid points on the moment or period of interest, equals to component of domain performance metric
PMV	predictive mean vote
PR/PlotRatio	plot ratio

PR/PlotRatio	plot ratio
PSO	particle swarm optimisation
RBA/ResiBuildingArea	residential building area
RBFNN	radical base function neural network
RC/Resicircumference	residential circumference
RFPA/ResiBuildFootprintArea	residential foot print area
RH	relative humidity
RSA/ResiSuperfacialArea	residential superficial area
SD-MOF	single-domain multi-objective optimisation function, solving maximisation or minimisation for SD-MOO
SD-MOO	single domain multi-objective optimisation, weighted summation of SD-SOO, weighted summation of Ovs
SD-SOF	single domain single objective function
SD-SOO	single domain single objective optimisation
SF/ShapeFactor	site shape factor
SHW	sunlight hour in winter
SI	swarm intelligence
SOF	single objective optimisation functions
SOO	single objective optimisation
SPEA-2	strength pareto evolutionary algorithm 2
SPL	sound pressure level
SPL(A)	A weighted sound pressure level
SPL-IQR	interquatile range, range between P25 to P75 of SPL grid value of a site
SPL-L10	sound pressure level L10
SPL-P10	10% percentile of SPL grid value of a site
SPL-P40	40% percentile of SPL grid value of a site
SPL-P70	70% percentile of SPL grid value of a site
SPL-THR(65)	threshold ratio of grid points met acceptable threshold of less than 65dBA in one acoustic simulation
sRC	residential circumference
sRSA	residential superficial area
STI	the speech transmission index
SVF	sky view factor
SVM	support vector machine
T30	reverberation time under T30 measurement
T_{db}	outdoor dry bulb temperature
TIN	The triangular irregular network tool
TLA/TotalLandArea	total land area
TSD/TriangleSD	standard deviation of triangle area
UDI	useful daylight illuminance

UHI	the urban heat island intensity
VIF	variance inflation factor
VSC	vertical sky component
WPSH	winter possible sunlight hour
WPSH-IQR	the difference between two quartiles of WPSH grid value of a site
WPSH-P30	30% percentile of WPSH grid value of a site
WPSH-P50	50% percentile of WPSH grid value of a site
WPSH-P70	70% percentile of WPSH grid value of a site
WPSH-THR(0)	ratio of grid points with sunlight duration equal to 0 hours in winter
WPSH-THR(83)	ratio of grid points with sunlight duration below 83 hours in winter
WS	wind speed
WS-Max	maximum of wind speed
WS-P20	20% percentile of WS grid value of a site
WS-P50	50% percentile of WS grid value of a site
WS-P70	70% percentile of WS grid value of a site
Z-APSH	normalised APSH value
ZPRED	standardised predicted value
ZRESD	standardized residuals
Z-SHW	normalised SHW value
Z-WPSH	normalised WPSH value

Bibliography and References

Alexandrov, N. M. and Lewis, R. M. (2002) 'Analytical and computational aspects of collaborative optimization for multidisciplinary design', *AIAA journal*, 40(2), pp. 301–309.

Appleby, P. (2012) *Integrated sustainable design of buildings*. Routledge.

Attia, S. *et al.* (2012) 'Simulation-based decision support tool for early stages of zero-energy building design', *Energy and Buildings*, 49, pp. 2–15. doi: https://doi.org/10.1016/j.enbuild.2012.01.028.

Attia, S. *et al.* (2013) 'Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design', *Energy and Buildings*, 60, pp. 110–124. doi: https://doi.org/10.1016/j.enbuild.2013.01.016.

Balesdent, M. *et al.* (2012) 'A survey of multidisciplinary design optimization methods in launch vehicle design', *Structural and Multidisciplinary optimization*. Springer, 45(5), pp. 619–642.

Balling, R. J. and Wilkinson, C. A. (1997) 'Execution of multidisciplinary design optimization approaches on common test problems', *AIAA journal*, 35(1), pp. 178–186.

Banerjee, A., Quiroz, J. C. and Louis, S. J. (2008) 'A model of creative design using collaborative interactive genetic algorithms', in *Design Computing and Cognition'08*. Springer, pp. 397–416.

Bendu, H., Deepak, B. B. V. L. and Murugan, S. (2017) 'Multi-objective optimization of ethanol fuelled HCCI engine performance using hybrid GRNN–PSO', *Applied Energy*, 187, pp. 601–611. doi: https://doi.org/10.1016/j.apenergy.2016.11.072.

Beyer, H.-G. and Sendhoff, B. (2007) 'Robust optimization – A comprehensive survey', *Computer Methods in Applied Mechanics and Engineering*, 196(33), pp. 3190–3218. doi: https://doi.org/10.1016/j.cma.2007.03.003.

Bizjak, Marko ; Žalik, Borut ; Lukač, N. (2015) 'Evolutionary-driven search for solar building models using LiDAR data', *Energy and buildings*. Elsevier B.V, 92, pp. 195–203. doi: 10.1016/j.enbuild.2015.01.051.

Botteldooren, D., Dekoninck, L. and Gillis, D. (2011) 'The influence of traffic noise on appreciation of the living quality of a neighborhood', *Int J Environ Res Public Health*. Switzerland: Switzerland: Molecular Diversity Preservation International (MDPI), 8(3), pp. 777–798. doi: 10.3390/ijerph8030777.

Braun, R. D. (1996) 'Collaborative optimization: an architecture for large-scale distributed design'. Stanford University.

Brownlee, J. (2011) *Clever algorithms: nature-inspired programming recipes*. Jason Brownlee.

Bruse, M. (2005) 'ITCM–A simple dynamic 2-node model of the human thermoregulatory system and its application in a multi-agent system', 41, pp. 398–401.

Chantrelle, F. P. *et al.* (2011) 'Development of a multicriteria tool for optimizing the renovation of buildings', *Applied Energy*, 88(4), pp. 1386–1394. doi: https://doi.org/10.1016/j.apenergy.2010.10.002.

Chapman, L., Thornes, J. E. and Bradley, A. V (2001) 'Rapid determination of canyon geometry parameters for use in surface radiation budgets', *Theoretical and applied climatology*. Wien: Wien: Springer-Verlag, 69(1), pp. 81–89. doi: 10.1007/s007040170036.

Chen, H. (2021) 'Multidisciplinary Design Optimization (MDO)', *Encyclopedia of Ocean Engineering, W. Cui, S. Fu, Z. Hu, Eds. Singapore: Springer*, pp. 1–7.

Chen, K. W., Janssen, P. and Schlueter, A. (2018) 'Multi-objective optimisation of building form, envelope and cooling system for improved building energy performance', *Automation in construction*. AMSTERDAM: AMSTERDAM: Elsevier B.V, 94(July), pp. 449–457. doi: 10.1016/j.autcon.2018.07.002.

Chen, L. and Ng, E. (2012) 'Outdoor thermal comfort and outdoor activities: A review of research in the past decade', *Cities*. Elsevier Ltd, 29(2), pp. 118–125. doi: 10.1016/j.cities.2011.08.006.

Chen, S., Zhang, F. and Khalid, M. (2002) 'Evaluation of three decomposition MDO algorithms', in *Proceedings of 23rd international congress of aerospace sciences, Toronto, Canada*.

Chen, X., Yang, H. and Sun, K. (2016) 'A holistic passive design approach to optimize indoor environmental quality of a typical residential building in Hong Kong', *Energy (Oxford)*. OXFORD: OXFORD: Elsevier Ltd, 113, pp. 267–281. doi: 10.1016/j.energy.2016.07.058.

Chen, X., Yang, H. and Sun, K. (2017) 'Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings q', *Applied Energy*. Elsevier Ltd, 194, pp. 422–439. doi: 10.1016/j.apenergy.2016.08.180.

Cheng, V. and Ng, E. (2006) 'Thermal Comfort in Urban Open Spaces for Hong Kong', *Architectural Science Review*. Taylor & Francis, 49(3), pp. 236–242. doi: 10.3763/asre.2006.4932.

Ciftcioglu, O., Sariyildiz, I. S. and Bittermann, M. S. (2007) 'Building performance analysis supported by GA', *CEC*. IEEE, pp. 859–866. doi: 10.1109/CEC.2007.4424560.

Citherlet, S. (2001) 'Towards the holistic assessment of building performance based on an integrated simulation approach'. Citeseer.

Coello, C. A. C. (2002) 'Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art', *Computer methods in applied mechanics and engineering*. Elsevier, 191(11–12), pp. 1245–1287.

Coello, C. A. C., Pulido, G. T. and Lechuga, M. S. (2004) 'Handling multiple objectives with particle swarm optimization', *IEEE Transactions on Evolutionary Computation*, 8(3), pp. 256–279. doi: 10.1109/TEVC.2004.826067.

Conceição António, Carlos A ; Monteiro, João Brasileiro ; Afonso, C. F. (2014) 'Optimal topology of urban buildings for maximization of annual solar irradiation

availability using a genetic algorithm', 73, pp. 424–437. doi: 10.1016/j.applthermaleng.2014.08.007.

Cramer, E. J. *et al.* (1994) 'Problem formulation for multidisciplinary optimization', *SIAM Journal on Optimization*. SIAM, 4(4), pp. 754–776.

Cui, Y. *et al.* (2017) 'Review : Multi-objective optimization methods and application in energy saving', *Energy*. Elsevier Ltd, 125, pp. 681–704. doi: 10.1016/j.energy.2017.02.174.

Deb, K. *et al.* (2000) 'A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II BT-Parallel Problem Solving from Nature PPSN VI', in Schoenauer, M. et al. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 849–858.

Deb, K. *et al.* (2002) 'A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II', *IEEE Transactions on Evolutionary Computation*, 6(2), pp. 182–197. doi: 10.1109/4235.996017.

Deb, K. and Goel, T. (2001) 'Controlled Elitist Non-dominated Sorting Genetic Algorithms for Better Convergence BT-Evolutionary Multi-Criterion Optimization', in Zitzler, E. et al. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 67–81.

Deb, K. and Goldberg, D. E. (1993) 'Analyzing Deception in Trap Functions', in WHITLEY, L. DARRELL, F. of G. A. (ed.) *Foundations of Genetic Algorithms*. Elsevier, pp. 93–108. doi: https://doi.org/10.1016/B978-0-08-094832-4.50012-X.

Dorigo, M., Birattari, M. and Stutzle, T. (2006) 'Ant colony optimization', *IEEE Computational Intelligence Magazine*, 1(4), pp. 28–39. doi: 10.1109/MCI.2006.329691.

Dutta, K. and Sarthak, S. (2011) 'Architectural space planning using evolutionary computing approaches: a review', *The Artificial intelligence review*. Dordrecht: Dordrecht: Springer Netherlands, 36(4), pp. 311–321. doi: 10.1007/s10462-011-9217-y.

Eberhart, R. and Kennedy, J. (1995) 'A new optimizer using particle swarm theory', in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39–43. doi: 10.1109/MHS.1995.494215.

Ekici, B. *et al.* (2019) 'Performative computational architecture using swarm and evolutionary optimisation : A review', *Building and Environment*. Elsevier, 147(May 2018), pp. 356–371. doi: 10.1016/j.buildenv.2018.10.023.

Eltaweel, A. and Su, Y. (2017) 'Parametric design and daylighting : A literature review', 73(October 2016), pp. 1086–1103. doi: 10.1016/j.rser.2017.02.011.

Ercan, B. and Elias-ozkan, S. T. (2015) 'Performance-based parametric design explorations: A method for generating appropriate building components', *Design Studies*. Elsevier Ltd, 38, pp. 33–53. doi: 10.1016/j.destud.2015.01.001.

EU Directive (2002) 'Directive 2002/49/EC of the European parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise', *Official Journal of the European Communities, L*, 189(18.07), p. 2002.

Evins, R. (2013) 'A review of computational optimisation methods applied to sustainable building design', *Renewable and Sustainable Energy Reviews*. Elsevier, 22, pp. 230–245. doi: https://doi.org/10.1016/j.rser.2013.02.004.

Fanger, P. O. (1972) *Thermal Comfort: Analysis and Applications in Environmental Engineering*. New York: McGraw-Hill Book Company.

Futrell, B. J., Ozelkan, E. C. and Brentrup, D. (2015) 'Bi-objective optimization of building enclosure design for thermal and lighting performance', *Building and Environment*. Elsevier Ltd, 92, pp. 591–602. doi: 10.1016/j.buildenv.2015.03.039.

Gagge, A. P., Fobelets, A. P. and Berglund, L. (1986) 'A standard predictive index of human response to the thermal environment', *ASHRAE Transactions*. ASHRAE, 92(2), pp. 709–731.

Gagge, A. P., Stolwijk, J. A. J. and Nishi, Y. (1971) 'An Effective Temperature Scale Based on a Simple Model of Human Physiological Regulatiry Response', *ASHRAE Transactions*. ASHRAE, 77(1), pp. 247–262. Available at: http://hdl.handle.net/2115/37901.

Gagne, J. and Andersen, M. (2012) 'A generative facade design method based on daylighting performance goals', *Journal of building performance simulation*. Taylor & Francis, 5(3), pp. 141–154. doi: 10.1080/19401493.2010.549572.

Gandomi, A. H. (2014) 'Interior search algorithm (ISA): A novel approach for global optimization', *ISA Transactions*, 53(4), pp. 1168–1183. doi: https://doi.org/10.1016/j.isatra.2014.03.018.

Gandomi, A. H. and Alavi, A. H. (2012) 'Krill herd: A new bio-inspired optimization algorithm', *Communications in Nonlinear Science and Numerical Simulation*, 17(12), pp. 4831–4845. doi: https://doi.org/10.1016/j.cnsns.2012.05.010.

Gandomi, A. H., Yang, X.-S. and Alavi, A. H. (2013) 'Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems', *Engineering with Computers*, 29(1), pp. 17–35. doi: 10.1007/s00366-011-0241-y.

Gehl, J. (1971) *Life Between buildings:Using Public Space*. London:Isand Press.

Gidlöf-Gunnarsson, A. and Öhrström, E. (2010) 'Attractive" quiet" courtyards: a potential modifier of urban residents' responses to road traffic noise?', *International journal of environmental research and public health*. Molecular Diversity Preservation International, 7(9), pp. 3359–3375.

Givoni, B. (1976) *Man, climate and architecture*. London: Applied Science Publishers.

Goldberg, D. E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st edn. Addison-Wesley Longman Publishing Co., Inc.

Gossard, D., Lartigue, B. and Thellier, F. (2013) 'Multi-objective optimization of a building envelope for thermal performance using genetic algorithms and artificial neural network', *Energy and buildings*. LAUSANNE: LAUSANNE: Elsevier B.V, 67, pp. 253–260. doi: 10.1016/j.enbuild.2013.08.026.

Gou, S. *et al.* (2018) 'Passive design optimization of newly-built residential buildings in Shanghai for improving indoor thermal comfort while reducing building energy demand', *Energy & Buildings*. Elsevier B.V., 169, pp. 484–506. doi: 10.1016/j.enbuild.2017.09.095.

Great Britain. Department of the Environment, Great Britain. Ministry of Housing and Local Government, Great Britain. Ministry of Public Building and Works, Great Britain. Ministry of Transport, Great Britain. Department of National Heritage, Great

Britai, T. and the R. (1975) *Calculation of road traffic noise*. Edited by G. B. W. Office. London: London : H.M.S.O., 1975.

Gulyás, Á., Unger, J. and Matzarakis, A. (2006) 'Assessment of the microclimatic and human comfort conditions in a complex urban environment: Modelling and measurements', *Building and Environment*, 41(12), pp. 1713–1722. doi: https://doi.org/10.1016/j.buildenv.2005.07.001.

Gustafson, E. J. (1998) 'Quantifying landscape spatial pattern: what is the state of the art?', *Ecosystems*. Springer, 1(2), pp. 143–156.

Haftka, R. T. (1985) 'Simultaneous analysis and design', *AIAA journal*, 23(7), pp. 1099–1103.

Hamdi, M., Lachiver, G. and Michaud, F. (1999) 'A new predictive thermal sensation index of human response', *Energy and Buildings*, 29(2), pp. 167–178. doi: https://doi.org/10.1016/S0378-7788(98)00054-1.

Holland, J. H. (1992) 'Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence'. The MIT Press. doi: 10.7551/mitpress/1090.001.0001.

Höppe, P. (2002) 'Different aspects of assessing indoor and outdoor thermal comfort', *Energy and Buildings*, 34(6), pp. 661–665. doi: https://doi.org/10.1016/S0378-7788(02)00017-8.

Huang, Y.-S., Chang, W.-S. and Shih, S.-G. (2015) 'Building Massing Optimization in the Conceptual Design Phase', *Computer-Aided Design and Applications*. Taylor & Francis, 12(3), pp. 344–354. doi: 10.1080/16864360.2014.981465.

Huang, Y. and Niu, J. (2016) 'Optimal building envelope design based on simulated performance: History, current status and new potentials', *Energy and Buildings*. Elsevier B.V., 117, pp. 387–398. doi: https://doi.org/10.1016/j.enbuild.2015.09.025.

Hurtley, C. (2009) *Night noise guidelines for Europe*. WHO Regional Office Europe.

Hygh, J. S. *et al.* (2012) 'Multivariate regression as an energy assessment tool in early building design', *Building and Environment*, 57, pp. 165–175. doi: https://doi.org/10.1016/j.buildenv.2012.04.021.

Jangir, P. and Jangir, N. (2018) 'Engineering Applications of Artificial Intelligence A new Non-Dominated Sorting Grey Wolf Optimizer (NS-GWO) algorithm : Development and application to solve engineering designs and economic constrained emission dispatch problem with integration of wind power', *Engineering Applications of Artificial Intelligence*. Elsevier Ltd, 72(September 2017), pp. 449–467. doi: 10.1016/j.engappai.2018.04.018.

Kämpf, J. H. *et al.* (2010) 'Optimisation of buildings' solar irradiation availability', *Solar Energy*, 84(4), pp. 596–603. doi: 10.1016/j.solener.2009.07.013.

Kämpf, J. H. and Robinson, D. (2010) 'Optimisation of building form for solar energy utilisation using constrained evolutionary algorithms', *Energy and Buildings*, 42(6), pp. 807–814. doi: https://doi.org/10.1016/j.enbuild.2009.11.019.

Kang, J. (2006) *Urban sound environment*. CRC Press.

Kelley, C. T. (1999) 'Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition', *SIAM journal on optimization*. SIAM, 10(1), pp. 43–55.

Kennedy, J. and Eberhart, R. (1995) 'Particle swarm optimization', in *Proceedings of ICNN'95 - International Conference on Neural Networks*, pp. 1942–1948 vol.4. doi: 10.1109/ICNN.1995.488968.

Kenny, N. A. *et al.* (2009) 'Part A: Assessing the performance of the COMFA outdoor thermal comfort model on subjects performing physical activity', *International Journal of Biometeorology*, 53(5), p. 415. doi: 10.1007/s00484-009-0226-3.

Kheiri, F. (2018) 'A review on optimization methods applied in energy-efficient building geometry and envelope design', *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, 92(May 2017), pp. 897–920. doi: 10.1016/j.rser.2018.04.080.

King, E. A. and Rice, H. J. (2009) 'The development of a practical framework for strategic noise mapping', *Applied Acoustics*. Elsevier, 70(8), pp. 1116–1127.

Klæboe, R. (2011) 'Environmental exposure—annoyance relationships in black and gray urban areas', in *10th International Congress on Noise as a Public Health Problem (ICBEN) Proceedings, London, UK*.

Klæboe, R., Engelien, E. and Steinnes, M. (2006) 'Context sensitive noise impact mapping', *Applied acoustics*. Oxford: Oxford: Elsevier Ltd, 67(7), pp. 620–642. doi: 10.1016/j.apacoust.2005.12.002.

Knowles, J. D., Watson, R. A. and Corne, D. W. (2001) 'Reducing Local Optima in Single-Objective Problems by Multi-objectivization BT-Evolutionary Multi-Criterion Optimization', in Zitzler, E. et al. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 269–283.

Kolarevic, B. (2004) 'Back to the Future: Performative Architecture', *International Journal of Architectural Computing*. SAGE Publications, 2(1), pp. 43–50. doi: 10.1260/1478077041220205.

Kroo, I. *et al.* (1994) 'Multidisciplinary optimization methods for aircraft preliminary design', in *5th symposium on multidisciplinary analysis and optimization*, p. 4325.

Kyu, Y. and Kim, H. (2015) 'Agent-based geometry optimization with Genetic Algorithm (GA) for tall apartment's solar right', *Solar Energy*. Elsevier Ltd, 113, pp. 236–250. doi: 10.1016/j.solener.2014.11.007.

Lam, K.-C. *et al.* (2009) 'Annoyance response to mixed transportation noise in Hong Kong', *Applied acoustics*. Kidlington: Kidlington: Elsevier Ltd, 70(1), pp. 1–10. doi: 10.1016/j.apacoust.2008.02.005.

Lam, K.-C. *et al.* (2013) 'Relationship between road traffic noisescape and urban form in Hong Kong', *Environ Monit Assess*, 185(12), pp. 9683–9695. doi: 10.1007/s10661-013-3282-4.

Lam, K.-C. and Ma, W.-C. (2012) 'Road traffic noise exposure in residential complexes built at different times between 1950 and 2000 in Hong Kong', *Applied Acoustics*, 73(11), pp. 1112–1120. doi: http://dx.doi.org/10.1016/j.apacoust.2012.05.001.

Lam, K. and Chung, Y. T. (2012) 'Exposure of urban populations to road traffic noise in Hong Kong', *Transportation Research Part D: Transport and Environment*. Elsevier, 17(6), pp. 466–472.

Law, C. *et al.* (2011) 'Advancement of three-dimensional noise mapping in Hong Kong', *Applied acoustics*. Elsevier Ltd, 72(8), pp. 534–543. doi: 10.1016/j.apacoust.2011.02.003.

Lee, H. S. (2004) 'Sequential approximate individual discipline feasible method using enhanced two-point diagonal quadratic approximation method. Hanyang University'. Master Thesis (in Korean).

Lee, S.-W., Chang, S. Il and Park, Y.-M. (2008) 'Utilizing noise mapping for environmental impact assessment in a downtown redevelopment area of Seoul, Korea', *Applied Acoustics*. Elsevier, 69(8), pp. 704–714.

Leung, B. K. H. and Mak, C. M. (2008) 'Is the CRTN Method Reliable and Accurate for Traffic Noise Prediction in Hong Kong?', *Transactions (Hong Kong Institution of Engineers)*. Taylor & Francis Group, 15(2), pp. 17–23. doi: 10.1080/1023697X.2008.10668113.

Li, D. H. W. *et al.* (2006) 'A study of the daylighting performance and energy use in heavily obstructed residential buildings via computer simulation techniques', *Energy and buildings*. Elsevier B.V, 38(11), pp. 1343–1348. doi: 10.1016/j.enbuild.2006.04.001.

Li, K. *et al.* (2017) 'Multi-Objective Optimization for Energy Performance Improvement of Residential Buildings: A Comparative Study', *Energies (Basel)*. MDPI AG, 10(2), p. 245. doi: 10.3390/en10020245.

Lin, Y. *et al.* (2018) 'Design Optimization Considering Variable Thermal Mass, Insulation, Absorptance of Solar Radiation, and Glazing Ratio Using a Prediction Model and Genetic Algorithm', *Sustainability (Basel, Switzerland)*. MDPI AG, 10(2), p. 336. doi: 10.3390/su10020336.

Littlefair, P. J. (2011) *Site Layout planning for daylight and sunlight*. 2nd edn. IHS BRE Press.

Liu, X., Liu, H. and Duan, H. (2007) 'Particle swarm optimization based on dynamic niche technology with applications to conceptual design', *Advances in engineering software (1992)*. Elsevier Ltd, 38(10), pp. 668–676. doi: 10.1016/j.advengsoft.2006.10.009.

Machairas, V., Tsangrassoulis, A. and Axarli, K. (2014) 'Algorithms for optimization of building design: A review', *Renewable and Sustainable Energy Reviews*, 31(1364), pp. 101–112. doi: https://doi.org/10.1016/j.rser.2013.11.036.

Magnier, L. and Haghighat, F. (2010) 'Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network', *Building and Environment*. Elsevier Ltd, 45(3), pp. 739–746. doi: 10.1016/j.buildenv.2009.08.016.

Mak, C. M., Leung, W. K. and Jiang, G. S. (2010) 'Measurement and prediction of road traffic noise at different building floor levels in Hong Kong', *Building services engineering research & technology*. London, England: London, England: SAGE Publications, 31(2), pp. 131–139. doi: 10.1177/0143624410361223.

Marriage, C. and Martins, J. R. R. A. (2008) 'Reconfigurable semi-analytic sensitivity methods and mdo architectures within the pimdo framework', in *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, p. 5956.

Marsault, X. (2017) *Eco-generative Design for Early Stages of Architecture*. John Wiley & Sons.

Marsault, X. and Torres, F. (2019) 'An interactive and generative eco-design tool for architects in the sketch phase', in *Journal of Physics: Conference Series*. IOP Publishing, p. 12136.

Matzarakis, A., Rutz, F. and Mayer, H. (2010) 'Modelling radiation fluxes in simple and complex environments: Basics of the RayMan model', *International Journal of Biometeorology*, 54(2), pp. 131–139. doi: 10.1007/s00484-009-0261-0.

Mayer, H. and Höppe, P. (1987) 'Thermal comfort of man in different urban environments', *Theoretical and Applied Climatology*, 38(1), pp. 43–49. doi: 10.1007/BF00866252.

McGarigal, K. and Marks, B. J. (1995) 'Spatial pattern analysis program for quantifying landscape structure', *Gen. Tech. Rep. PNW-GTR-351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station*, pp. 1–122.

Menges, A. (2012) 'Biomimetic design processes in architecture: morphogenetic and evolutionary computational design', *Bioinspir. Biomim*. England: England: IOP Publishing, 7(1), p. 15003. doi: 10.1088/1748-3182/7/1/015003.

Miedema, H. M. and Oudshoorn, C. G. (2001) 'Annoyance from Transportation Noise: Relationships with Exposure Metrics DNL and DENL and Their Confidence Intervals', *Environ Health Perspect*. United States: United States: National Institute of Environmental Health Sciences. National Institutes of Health. Department of Health, Education and Welfare, 109(4), pp. 409–416. doi: 10.1289/ehp.01109409.

Mirjalili, S. (2015) 'Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm', *Knowledge-Based Systems*, 89, pp. 228–249. doi: https://doi.org/10.1016/j.knosys.2015.07.006.

Mirjalili, S., Mirjalili, S. M. and Lewis, A. (2014) 'Grey Wolf Optimizer', *Advances in Engineering Software*. Elsevier Ltd, 69, pp. 46–61. doi: https://doi.org/10.1016/j.advengsoft.2013.12.007.

Mohammad Zadeh, P. and Sadat Shirazi, M. (2017) 'Multidisciplinary design optimization architecture to concurrent design of satellite systems', *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*. SAGE Publications Sage UK: London, England, 231(10), pp. 1898–1916.

Nault, E. *et al.* (2017) 'Predictive models for assessing the passive solar and daylight potential of neighborhood designs: A comparative proof-of-concept study', *Building and Environment*, 116, pp. 1–16. doi: 10.1016/j.buildenv.2017.01.018.

Navarro-Mateu, D. and Cocho-Bermejo, A. (2020) 'Evo-devo strategies for generative architecture: colour-based patterns in polygon meshes', *Biomimetics*. Multidisciplinary Digital Publishing Institute, 5(2), p. 23.

Negendahl, K. and Nielsen, T. R. (2015) 'Building energy optimization in the early design stages : A simplified method', *Energy & Buildings*. Elsevier B.V., 105, pp. 88–99. doi: 10.1016/j.enbuild.2015.06.087.

Ngo, M. and Labayrade, R. (2014) 'Multi-genomic algorithms', in *2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM)*. IEEE, pp. 48–55.

Nguyen, A., Reiter, S. and Rigo, P. (2014) 'A review on simulation-based optimization methods applied to building performance analysis', *Applied Energy*. Elsevier Ltd, 113, pp. 1043–1058. doi: https://doi.org/10.1016/j.apenergy.2013.08.061.

Nguyen, L. *et al.* (2016) 'Evolutionary processes as models for exploratory design', in *Biomimetic Research for Architecture and Building Construction*. Springer, pp. 295–318.

Ochoa, C. E. and Capeluto, I. G. (2009) 'Advice tool for early design stages of intelligent facades based on energy and visual comfort approach', *Energy and Buildings*, 41(5), pp. 480–488. doi: https://doi.org/10.1016/j.enbuild.2008.11.015.

Öhrström, E. *et al.* (2006) 'Effects of road traffic noise and the benefit of access to quietness', *Journal of sound and vibration*. Elsevier, 295(1–2), pp. 40–59. doi: http://dx.doi.org/10.1016/j.jsv.2005.11.034.

Oliveira Panão, M. J. N., Gonçalves, H. J. P. and Ferrão, P. M. C. (2008) 'Optimization of the urban building efficiency potential for mid-latitude climates using a genetic algorithm approach', *Renewable energy*. Oxford: Oxford: Elsevier Ltd, 33(5), pp. 887–896. doi: 10.1016/j.renene.2007.04.014.

Østergård, T., Jensen, R. L. and Maagaard, S. E. (2016) 'Building simulations supporting decision making in early design – A review', *Renewable and Sustainable Energy Reviews*. Elsevier, 61, pp. 187–201. doi: 10.1016/j.rser.2016.03.045.

Ouarghi, R. and Krarti, M. (2006) 'Building shape optimization using neural network and genetic algorithm approach'. Atlanta GA: Atlanta GA: ASHRAE, pp. 484–491.

Padula, S., Alexandrov, N. and Green, L. (1996) 'MDO test suite at NASA Langley Research Center', in *6th Symposium on Multidisciplinary Analysis and Optimization*, p. 4028.

Panda, A. and Pani, S. (2016) 'A Symbiotic Organisms Search Algorithm with Adaptive Penalty Function to Solve Multi-Objective Constrained Optimization Problems', *Appl. Soft Comput.* NLD: Elsevier Science Publishers B. V., 46(C), pp. 344–360. doi: 10.1016/j.asoc.2016.04.030.

Park, C.-G. and Lee, J.-S. (2001) 'Improvement of sensitivity based concurrent subspace optimization using automatic differentiation', *Transactions of the Korean Society of Mechanical Engineers A*. The Korean Society of Mechanical Engineers, 25(2), pp. 182–191.

Petersen, S. (2011) *Simulation-based support for integrated design of new low-energy office buildings*. Technical University of Denmark, Department of Civil Engineering.

Price, A. R., Keane, A. J. and Holden, C. M. E. (2011) 'On the coordination of multidisciplinary design optimization using expert systems', *AIAA journal*, 49(8), pp. 1778–1794.

Rakha, T. and Nassar, K. (2011) 'Genetic algorithms for ceiling form optimization in response to daylight levels', *Renewable energy*. Oxford: Oxford: Elsevier Ltd, 36(9), pp. 2348–2356. doi: 10.1016/j.renene.2011.02.006.

Ranjbar, H. R., Gharagozlou, A. R. and Nejad, A. R. V. (2012) '3D analysis and investigation of traffic noise impact from Hemmat highway located in Tehran on buildings and surrounding areas'. Scientific Research Publishing.

Renaud, J. E. and Gabriele, G. A. (1994) 'Approximation in nonhierarchic system optimization', *AIAA journal*, 32(1), pp. 198–205.

Sacks, J. *et al.* (1989) 'Design and Analysis of Computer Experiments', *Statistical science*. Institute of Mathematical Statistics, 4(4), pp. 409–423. doi: 10.1214/ss/1177012413.

Sadollah, A. *et al.* (2013) 'Mine blast algorithm: A new population based algorithm for solving constrained engineering optimization problems', *Applied Soft Computing*, 13(5), pp. 2592–2612. doi: https://doi.org/10.1016/j.asoc.2012.11.026.

Salomons, E. M. and Berghauser Pont, M. (2012) 'Urban traffic noise and the relation to urban density, form, and traffic elasticity', *Landscape and Urban Planning*, 108(1), pp. 2–16. doi: https://doi.org/10.1016/j.landurbplan.2012.06.017.

Sariyildiz, I. S. (2012) 'Performative computational design', in *Keynote speech in: Proceedings of ICONARCH-I: International congress of architecture-I, Konya, Turkey, 15-17 November 2012*. Selcuk University.

Schwarz, N. (2010) 'Urban form revisited—Selecting indicators for characterising European cities', *Landscape and urban planning*. Elsevier B.V, 96(1), pp. 29–47. doi: 10.1016/j.landurbplan.2010.01.007.

Sghiouri, H. *et al.* (2018) 'Shading devices optimization to enhance thermal comfort and energy performance of a residential building in Morocco', *Journal of Building Engineering*. Elsevier Ltd, 18, pp. 292–302. doi: 10.1016/j.jobe.2018.03.018.

Shi, X. *et al.* (2016) 'A review on building energy efficient design optimization rom the perspective of architects', *Renewable and Sustainable Energy Reviews*. Elsevier, 65, pp. 872–884. doi: 10.1016/j.rser.2016.07.050.

Shi, Z., Fonseca, J. A. and Schlueter, A. (2017) 'A review of simulation-based urban form generation and optimization for energy-driven urban design', *Building and environment*. Elsevier Ltd, 121, pp. 119–129. doi: 10.1016/j.buildenv.2017.05.006.

Skånberg, A. and Öhrström, E. (2002) 'Adverse Health Effects in Relation to Urban Residential Soundscapes', *Journal of sound and vibration*. Elsevier, 250(1), pp. 151–155. doi: http://dx.doi.org/10.1006/jsvi.2001.3894.

Sobieszczanski-Sobieski, J. (1982) *A linear decomposition method for large optimization problems. Blueprint for development*.

Sobieszczanski-Sobieski, J. (1988) 'Optimization by decomposition: a step from hierarchic to non-hierarchic systems', in *NASA/Air Force Symposium on Recent Advances in Multidisciplinary Analysis and Optimization*.

Sobieszczanski-Sobieski, J. *et al.* (2002) *Bi-level system synthesis (BLISS) for concurrent and distributed processing. American Institute of Aeronautics and Astronautics, New York*. AIAA-2002-5409.

Sobieszczanski-Sobieski, J., Agte, J. and Sandusky   Robert, J. (1998) 'Bi-level integrated system synthesis (BLISS)', in *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, p. 4916.

Solar gain in buildings (2017) *Design Building Wiki*. Available at: https://www.designingbuildings.co.uk/wiki/Solar_gain_in_buildings (Accessed: 1 August 2018).

Spagnolo, J. and de Dear, R. (2003) 'A field study of thermal comfort in outdoor and semi-outdoor environments in subtropical Sydney Australia', *Building and Environment*, 38(5), pp. 721–738. doi: https://doi.org/10.1016/S0360-1323(02)00209-3.

Specht, D. F. (1991) 'A general regression neural network', *IEEE transactions on neural networks*. Citeseer, 2(6), pp. 568–576. doi: 10.1109/72.97934.

Storn, R. and Price, K. (1997) 'Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces', *Journal of Global Optimization*, 11(4), pp. 341–359. doi: 10.1023/A:1008202821328.

Su, Z. and Yan, W. (2015) 'A fast genetic algorithm for solving architectural design optimization problems', *AIEDAM*. New York, USA: New York, USA: Cambridge University Press, 29(4), pp. 457–469. doi: 10.1017/S089006041500044X.

Sun, C., Han, Y. and Feng, H. (2015) 'Multi-objective building form optimization method based on GANN-BIM model', *Next Generation Building*, 2(1).

Tappeta, R. V and Renaud, J. E. (1997) 'Multiobjective collaborative optimization', in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, p. V002T29A017.

Tian, Z. *et al.* (2018) 'Towards adoption of building energy simulation and optimization for passive building design : A survey and a review', *Energy & Buildings*. Elsevier B.V., 158, pp. 1306–1316. doi: 10.1016/j.enbuild.2017.11.022.

Tsou, J.-Y., Chow, B. and Lam, S. (2003) 'Performance-based simulation for the planning and design of hyper-dense urban habitation', *Automation in construction*. Amsterdam: Amsterdam: Elsevier B.V, 12(5), pp. 521–526. doi: 10.1016/S0926-5805(03)00039-6.

Turner, M. G. and Gardner, R. H. (1991) *Quantitative methods in landscape ecology : the analysis and interpretation of landscape heterogeneity*. New York: New York.

Turrin, M., von Buelow, P. and Stouffs, R. (2011) 'Design explorations of performance driven geometry in architectural design using parametric modeling and genetic algorithms', *Advanced engineering informatics*. Elsevier Ltd, 25(4), pp. 656–675. doi: 10.1016/j.aei.2011.07.009.

Vermeulen, T. *et al.* (2015) 'Urban layout optimization framework to maximize direct solar irradiation', *Computers, Environment and Urban Systems*, 51, pp. 1–12. doi: 10.1016/j.compenvurbsys.2015.01.001.

Wang, L. *et al.* (2016) 'Solar radiation prediction using different techniques: model evaluation and comparison', *Renewable & sustainable energy reviews*. Elsevier Ltd, 61, pp. 384–397. doi: 10.1016/j.rser.2016.04.024.

Weber, N., Haase, D. and Franck, U. (2014) 'Traffic-induced noise levels in residential urban structures using landscape metrics as indicators', *Ecological indicators*. Elsevier, 45, pp. 611–621. doi: http://dx.doi.org/10.1016/j.ecolind.2014.05.004.

WHO (2011) *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe.

Yang, X.-S. (2010) 'A New Metaheuristic Bat-Inspired Algorithm BT-Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)', in González, J. R. et al. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 65–74. doi: 10.1007/978-3-642-12538-6_6.

Yi, H. (2014) 'Automated generation of optimised building envelope: simulation based multi-objective process using evolutionary algorithm', *International Journal of Sustainable Building Technology and Urban Development*. Taylor & Francis, 5(3), pp. 159–170. doi: 10.1080/2093761X.2014.906333.

Yi, S. I., Shin, J. K. and Park, G. J. (2008) 'Comparison of MDO methods with mathematical examples', *Structural and Multidisciplinary Optimization*, 35(5), pp. 391–402. doi: 10.1007/s00158-007-0150-2.

Yu, W. *et al.* (2015) 'Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design', *Energy & Buildings*. Elsevier B.V., 88, pp. 135–143. doi: 10.1016/j.enbuild.2014.11.063.

Zemella, G. *et al.* (2011) 'Optimised design of energy efficient building faГ§ades via Evolutionary Neural Networks', *Energy and buildings*. LAUSANNE: LAUSANNE: Elsevier B.V, 43(12), pp. 3297–3302. doi: 10.1016/j.enbuild.2011.10.006.

Zhang, A. *et al.* (2017) 'Optimization of thermal and daylight performance of school buildings based on a multi-objective genetic algorithm in the cold climate of China', *Energy & Buildings*. Elsevier B.V., 139, pp. 371–384. doi: 10.1016/j.enbuild.2017.01.048.

Zhang, L. L., Zhang, L. L. and Wang, Y. (2016) 'Shape optimization of free-form buildings based on solar radiation gain and space efficiency using a multi-objective genetic algorithm in the severe cold zones of China', *Solar energy*. Elsevier Ltd, 132, pp. 38–50. doi: 10.1016/j.solener.2016.02.053.