Inference in the context of uncertain complex urban environments for climate change conscious planning and design

Jonathan Sykes

University of Sheffield

The
University
Of
Sheffield.

A thesis presented for the degree of
Doctor of Philosophy

December 2, 2022

I dedicate this thesis to my teachers, supervisors, friends and family. One person is never a story, but a story might well be very different without them.

One never notices what has been done; one can only see what remains to be done.

— Marie Curie

Acknowledgements

I would like to express my gratitude to my supervisors Professor Darren Robinson, Parag Wate, Professor Richard Wilkinson and Dr. Chengzhi Peng and my friends and family who have been around me during this work.

Many people have been extremely understanding and supportive of me in completing this piece of work. I am extremely grateful for this support, if you are reading this and know me personally to any extent you are most definitely in this group. Many people in industry who I've met have given me many a friendly conversation helping me broaden my understanding of industrial context and nudges toward things that are just generally interesting.

Declaration, acknowledgements and appreciation

I, Jonathan Sykes, declare that this work is my own work with comments from my supervisors: Darren Robinson, Parag Wate, Richard Wilkinson and Chengzhi Peng, on each chapter individually and on the thesis as a whole except where otherwise acknowledged. The University of Sheffield guidelines on fair use can be found at:

https://www.sheffield.ac.uk/rs/code/plagiarism

Additionally I would like to highlight the sourcing of data for several of the chapters and recognise several individuals in connection with that:

- The Information Commons data, for chapter 4 was provided by the University of Sheffield with permission for research. Peter Webber was my liaison and was invaluable in explaining context to how the systems in the Information Commons work.

- To Alister McLean, Curator at the Weston Park Museum for providing local weather data and advice on its collection for chapter 4.

The data-sets used in chapters 5 and 6 are referenced in those chapters. Appendix A provides some additional background to the Birmingham University Climate Laboratory data-set which provided excellent material to study large scale city design elements.

Abstract

This thesis looks at the urban environment as the centre of human habitation. ~~The complex environment of urban human habitation is important for many reasons.~~ It governs the comfort of much of the human population and is essential to life itself. In the modern world, it is governed at many levels and this thesis approaches two of them: modelling a building's system and elements of urban city design.

Urban climate, in the UK, is being increasingly affected by climate change and urban pollution remains a concern. How cities are maintained and designed is being adjusted to consider these interactions. ~~In t~~This thesis ~~, I~~ looks at the impact of roughness of the cities-scape on wind speed, considered a factor capable of improving air quality . ~~I also~~This thesis will looks at urban albedo and the impact it has on air temperature at ground level compared with the general degree of urban density.

Uncertainty is a part of complex systems such as cities which contain many elements and in order to address this models are used to describe these system. ~~a~~A modeller will not have access to all information or the time to address every element at a high level of detail. The Gaussian processes used in this thesis have inherent uncertainty quantification, and they make estimates that make allowances for inaccuracies. This means conclusions drawn using this method can be considered more robust to uncertainties in the data.

This thesis will examine empirical data using different methodologies to draw conclusions about model fitness of the methods used. The case studies that are used are the problem of emulator construction for the building energy models (BEMs) and two example relationships of urban weather from the Birmingham University Climate Laboratory (BUCL) and the urban fabric.

Building energy use, through domestic, office and industrial consumption, is a major part of how we as a society consumes electricity/gas and this consumption is metered. Building systems are modelled using physical principles which requires a large amount of information about constructed systems, user behaviour and the ambient environment which is very costly justifying alternatives such as statistical modelling. This thesis will showcase how they can be used to address the issue of climate change for a building energy use, in cooling.

Keywords— Uncertain quantification - Urban Climate - Building Energy Modelling - Gaussian Processes - Sensitivity Analysis - Statistical Modelling

Contents

List of Figures

List of Tables

List of Abbreviations

API . . . . Action Protocol Interface

BEM . . . . Building Energy Model

BH-BB albedo Bi-Hemispherical Broad-band albedo

BEIS . . . . (Department of) Business Energy and Industrial Strategy (UK)

CFD . . . . Computational Fluid Dynamics

CIBSE . . . . Civil Institute for Building Service Engineers

DEFRA . . . . . . Department for Environment, Food & Rural Affairs

DSY . . . . Design Summer Year(s)

EC-JRC . . . . . . European Commission Joint Research Centre

GLM . . . . General Linear Model

GP . . . . Gaussian Process

IC . . . . Information Commons

IPCC . . . . International Panel on Climate Change

MLE . . . . Maximum Likelihood Estimation

POE . . . . Post Occupancy Evaluation

RMSE . . . . Root Mean Squared Error

SA . . . . Sensitivity Analysis

SME . . . . Subject Matter Expert

UA . . . . Uncertainty Analysis

UHI . . . . Urban Heat Island

UNCC . . . . United Nations on Climate Change

UWG . . . . Urban Weather Generator (Bueno et al. 2013)

(n.d.) . . . . no date - used for references referring to organisation/web pages. All have accessed within 2019-2020.

Chapter 1

Introduction

1.1 Outline

This thesis investigates~~The premise of this thesis is to investigate~~ methods for evaluating causation in data in UK climate, in particular how models use data and if they are fit for purpose.~~some of the l~~Large amounts of data on the urban fabric, that are being generated in relation to modern cities and ~~how~~ this data can be used to draw conclusions about the sustainability of current building and planning design. ~~There are major difficulties with this process of inference. This work focuses on the relationship of weather or climate data with urban centres.~~ Cities are where an increasing proportion of the Earth's population reside (Leeson 2018). The UK is a prime example of a highly urbanised population~~,~~ and the prime issue for this population~~,~~ has, historically, been ~~considered to be~~ winter heating (Keatinge et al. 1989, Few et al. 2021). As such, models for these environments need to command confidence in their conclusions and have explicable strengths and weaknesses.

1.1.1 On the climate imperative

The UK has signed the Paris climate change agreement (UNCC 2016) ~~to~~ which has implications for the UK economy and infrastructure. The headline figures generated by the UK government (BEIS 2020) are that the UK should achieve net 68% reduction in emission on 1990 levels by 2030 and net zero by 2050. This is set in the context of the IPCC report (IPCC 2021). The mitigation report's summary for policy makers (IPCC 2022) states that limiting global warming to 1.5°C can be achieved with a 50% chance if emissions are limited to 500 giga-tonnes of CO_2. At present global emissions for 1990 to 2019 years were 1000±90 giga-tonnes.

Reaching the CO_2 targets means reducing the energy used in UK heating is ~~highly challenging~~difficult despite supported improvements in the UK housing stock such as the Green Homes Grant (BEIS 2021). ~~An example is the discussion~~ The degree of accuracy of estimation with which home owners can benefit is limited, exemplified by ~~discussed in the paper by the~~ Gillich et al. (2019), in which the identified uncertainty in ~~modelling~~predicting energy use of the building stock is a major concern. The ways in which uncertainty prediction can be carried out in general is also a concern. The ability to quantify the benefits/consequences of policy decisions for building stock and associate uncertainty is important for decision makers to carry out evidence based policy making. It can inform the field of policy such as climate modelling IPCC (2021), or evaluations at much smaller scale such as Gillich et al. (2019) modelling heating of housing stock.

Investigating the atmosphere over and in cities is key to understanding these challenges Taha (1997).; particularly the atmosphere's complex interactions with the urban environment. An example is the impact of greening an urban area during a heatwave, this increase in mean ambient temperature has an impact on the energy consumption of cooling systems~~Examples include the impact of urban greening such as evaluated by~~ Taleghani et al. (2019).~~which examined the~~ Some examples of climate change impact on building energy use are decreased heating costs de Rubeis et al. (2020), however this will not necessarily decrease winter mortality (Staddon et al. 2014).

1.1.2   On model fitness and fitness for users/value for user groups

Model developers should evaluate their model fitness to validate that user requirements have been met. User requirements in the context of environmental modelling in cities are that they relate to the real environment of the city and how decisions would be made using the information they provide. The information must be comprehensible and accurate.

At present many models do not analyse this aspect of their own performance against user requirements examples such as Lukač et al. (2017) which discusses applications but not the if the research is in a suitable form for users. Indeed, as discussed below many model are not concerned with using real world data on micro-climate. This thesis carries out three pieces of work, 1) a building energy model for climate change impact modelling, 2) inference of the impact of roughness length on wind speed, 3) the impact of urban density and albedo on air temperature and evaluates their delivery of user requirements, this particularly key when focusing on the numerical methods fallibility's and capabilities.

1.2   Research Questions

As discussed in subsection 1.1.2, key questions exist about producing user valid models that are able to provide information for policy decisions. How evidence is generated for city fabric at number of different scales has been widely debated in the research community. The performance gap is a well researched topic of model validation, Zou et al. (2018), Yassaghi & Hoque (2019), Menberg et al. (2019), highlighting how model predictions interact with the real world of building construction. The concept of the performance gap highlights how the users of Building Energy Models (BEM) need representations of planned buildings energy use that are accurate, however there is often a significant margin for error in this process (Wate et al. 2019, 2016). Conversely, these models do meet the user requirements of being explainable and being able to quantify smaller scale energy flow that are useful to building designers. The emblematic of being a 'white box model'. Examples of other types of modelling that don't depend on physical modelling or 'white box models', or instead depend on data from energy use to train empirical models or 'black box' models.

White box models are distinguished by the mechanisms for output results being transparent such as in physical system where energy flows are explicitly calculated. The distinction in between 'black' and 'white' box models leads to different fallibilities. Zou et al. (2018) reviews 10 years of evidence on the energy performance gap and illustrates that the user requirements in terms of elements of the design process are well researched. The element that is poorly discussed is statistical analyses of these results.

In additional to BEM elements of city concerns, the micro-climate modelling over cities is often poor in discussion of statistical analysis. There has been discussion of the impact of heat waves on urban areas with warming temperatures. Most such work focuses on the output of case study simulations (Pisello et al. 2015, Gusson & Duarte 2016, Taleghani et al. 2019). There also appears to be limited work on empirical evaluation micro-climate effects over cities. As such, this thesis contributes to the field by providing a methodology for inference and the analysis of model fit.
~~As such the two main questions this thesis will address are:~~

1. ~~How are buildings impacted by the local micro-climate? I propose a method for making use of empirical data to simulate responses.~~

2. ~~How does the local built environment affect the micro-climate? More specifically how can I evaluate empirically the impact of potential and known micro-climatic factors in the built environment.~~

The two questions examined in this thesis to examine model user validation are:

1. When investigating the construction of empirical or black box model for building energy models can one be used to easily describe climate impact on a building?

    - Building energy use is sensitive to external weather conditions (de Rubeis et al. 2020). This sensitivity will be experienced by building energy stocks as climates change. Design summer years (DSY) are a way of providing climate information for building energy simulations. These are being updated to maintain relevance (Eames 2016). These inherently model climate change by incorporating different emissions scenarios and representative scenarios for different emission pathways. Constantly changing weather creates a demand for empirical models that can be rapidly deployed on new buildings without surveys.

    - The data available for this project is provided by the University of Sheffield Estates: energy use data for the Information Commons building and has been paired with public weather data from two sources for current and future climates.

    - This modelling will create a black box modelling project that quantifies uncertainty. The user group would be building operators and how they plan for climate change.

2. When investigating the empirical evidence for interaction between urban fabric and urban micro-climate, can we evaluate the effectiveness of methods?

    - There have been many efforts to study the environment over cities examples include Computational Fluid Dynamic (CFD) models such Gusson & Duarte (2016), Mo & Liu (2018) studying more generic impact urban micro-climate. Alternatively some studies are more focused on terrain summaries such as roughness length such as Wicht & Wicht (2018), Gál & Unger (2009). Work to tally the effectiveness against effect experienced in the real world are challenging. The challenge is not from the physical basis, but from the complexity of urban environments. Vegetation can be highly heterogeneous across a city Vuckovic et al. (2017), which can lead to varied anthropogenic heat flux for example. These features combined with noise in weather from natural climate variability make identifying causation challenging.

- Weather station data is available from the University of Birmingham Climate Laboratory (BUCL) project (Bassett et al. 2016), paired with google maps images of the city and LIDAR height maps (U.K. Environment Agency n.d.)

- The user groups for this analysis would be those in the research community about making decisions about changes to the urban fabric.

1.3   Concept Introductions

It is worth having some higher level discussion of the theory surrounding the topics in this thesis.

- Building energy modelling (BEM) is used to simulate temperature and the associated energy flows in a building. Chapters 3 and 4, create statistical emulators of physics derived models of such building systems. In particular I will be looking at how BEM is influenced by the weather system outside the building allowing me to model the possible impact of climate change.

- One view of statistics and by extension part of machine learning is to summarise information. When this learning is about drawing conclusions, it is sensible to build statistical models that take this into account. ~~Simply put t~~ The models need to ~~express doubt over~~ appropriate certainty in their conclusions.

- The atmosphere over cities is generally modelled using Navier-Stokes equations. There are many adaptations and implementations of this model. I will not discuss this in detail, however it is worth discussing details such as some common approximations that are used modelling in chapter 5.

1.3.1   Building Energy Use

Buildings are complex and often more unique than might be immediately apparent. Systems in different buildings are often not directly comparable. Buildings fundamentally can be thought of as containers for heat. Various ~~things~~factors cause the container to gain heat and this heat also moves out via leaking and active transport. ~~The balance of the gain and loss of heat dictate if~~ The law of conservation of energy means that if energy is transport on net into or out of the building the building heats up or cools down.

Depending on the size and building materials the rate at which the structure will change temperature can be highly variable (Foteinaki et al. 2018). Large well insulated blocks of flats can take months to have substantial temperature change internally under normal operation, ~~small timber framed~~a single family house buildings can take just hours. These are ~~modelled~~ unsurprisingly mostly modelled in terms of control systems and the factor they are meant to control.

The basic example everyone would be familiar with would be a boiler linked to a thermostat in a room. A simple boiler control would keep the heating fluid at a constant temperature. The thermostat then controls flow of the heating fluid into radiators to release heat into the rooms that will cool if not heated. ~~As such the amount of heat the boiler needs to release is the rate at which the heat is lost from the rooms heated plus the system inefficiencies. In this system a model would~~

~~comprise~~ The conservation of energy modelled in these energy flows requires the balancing of the heat loss of the heated room, the system inefficiencies, the state of the thermostat control value and the power output of the boiler.
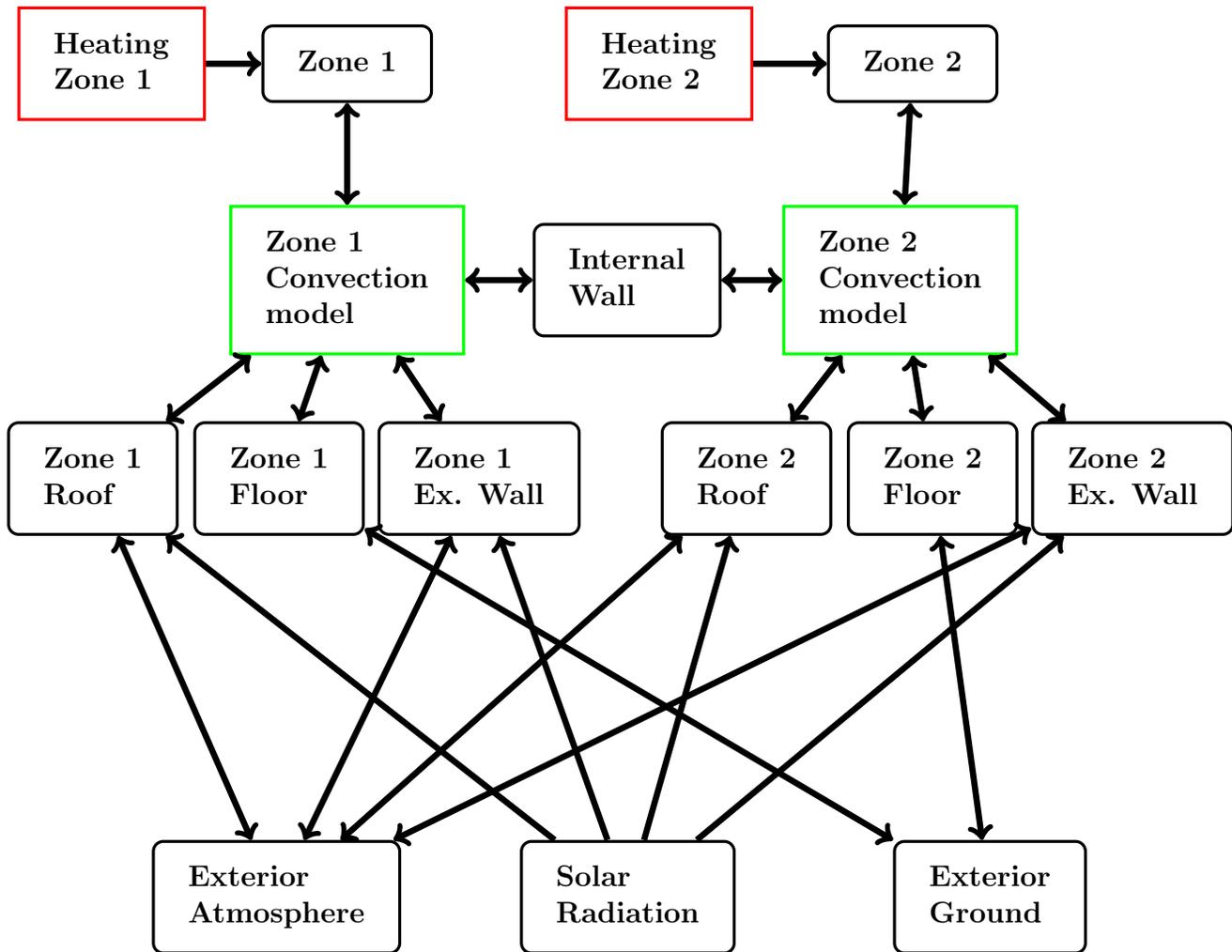
Figure 1.1: Example capacitor network diagram with two zones, with black boxes representing capacitors, arrowheads indicate transfer of heat into nodes. Walls, roofs and floors might often actually comprise several nodes in sequence representing different materials. The convection model could be a fluid dynamics simulation that governs how the zones interact with walls, ceilings and floors. External atmosphere and ground temperatures will be provided by data-sets whether past recorded data or from a linked atmospheric model such as ENVI-met (Simon 2016). The arrow represent capacitor connections as laid out in figure 1.2

The most common model that handles this sort of situation is a lump parameter model. This comprises a capacitor network where each node represents something with the capacity to hold heat and pass it to other nodes as dictated by the shape of the network, an example is laid out in 1.2. Figure 1.1 provides an example of how the physical building might be modelled. Differential equations govern the transfer of energy between components.

The guidelines in the UK that cover how the modelling is done for real buildings in in-
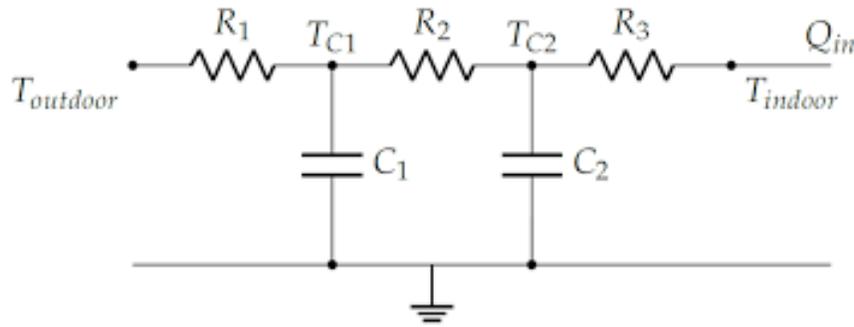
Figure 1.2: Formal capacitor network diagram representing a connection in figure 1.1 taken from Boodi et al. (2020), this case between indoor and outddor

dustry are provided by the Chartered ~~Civil~~ Institute for Building Service Engineers (CIBSE) (Civil Institute for Building Service Engineers n.d.). Overheating is being evaluated as a risk to UK building and methodology is being continuously developed to handle this issue, examples being Jentsch et al. (2013), Eames (2016) and Ramallo-González et al. (2020). However as was highlighted by the CIBSE report in 2017, Bonfigli et al. (2017), there is conflict ~~between~~with UK climate change commitments to energy efficiency. This trade off comes from increased heating efficiency by increasing air tightness of modern buildings. This can compromise air quality and the ability of the building to shed heat during the night period in heat waves. These sort of trade-offs are referred to in the Fifth Carbon Budget (Change 2017), where it refers to accessing increasingly difficult areas for savings.

As a result Bonfigli et al. (2017) recommends energy savings and overheating in buildings need to be considered in parallel to manage the trade offs. This means an increasing demand for simulation of the performance of buildings, old and new. Chapters 3 and 4 are targeted at providing a way using empirical data to meet this need.

1.3.2   Uncertain machine learning and Gaussian processes

Uncertainty is a catch all phrase used both in wider society and in more more technical contexts to say 'we don't know'. In the context of science and more specifically machine learning ~~the most~~a common way of building a cohesive way of ~~describing~~quantifying 'not knowing', is to use the theory of probability and by extension likelihoods, (Rasmussen et al. 2006). Exactly what probability means is an interesting and complex debate. ~~If the reader has some further interest I would suggest~~ For discussion of the theory of probability see O'Hagan & Oakley (2004) ~~as a paper. There are also many good books that discuss it. From a personal perspective I~~The author believes ~~that~~ it is sufficient that it is a model of how we as decision makers should weigh the frequency of a possible outcome. This is consistent with interpretations such as repeating identical experiments and expecting proportions of different outcomesthough closer to a Bayesian subjective interpretation of probability such as in O'Hagan (2004) . It does address three criteria:

1. Assuming decision making is to be made on an empirical basis, the process should take into account evidence from the formulator's perspective.

2. The belief is graded in that it permits comparisons between judgements.
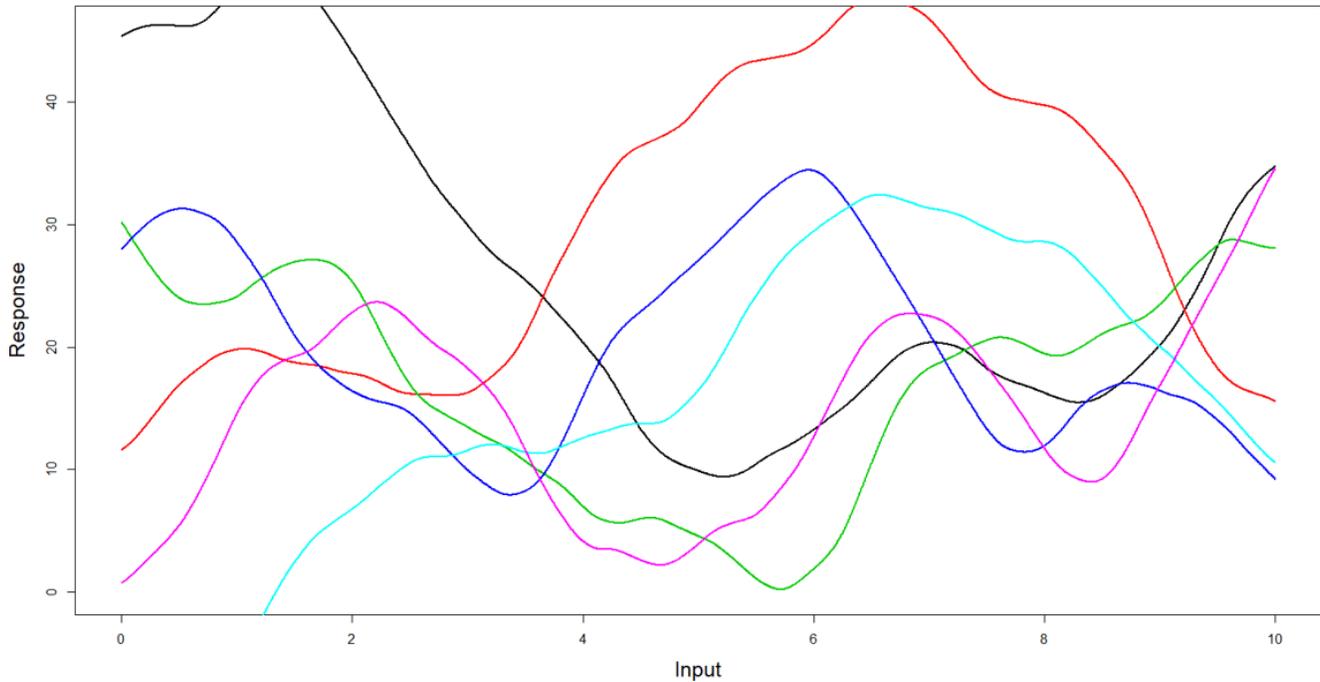
Figure 1.3: Samples from a Gaussian process prior distribution with squared exponential covariance function plotted as lines as GP samples are continuous.. These are 'draws' from the prior distribution

3. It should take into account the wider objective world in that the decision to base the decision on evidence should be robust, in that another individual should be able to understand the conclusions in different circumstances.

The difficulty in the fields relevant to this thesis is that the events being predicted are highly connected, in that the value of one unknown is known to correspond to a pattern of the outcome in another. In chapters 3 and 4, I will be predicting building energy use on the basis of the external weather conditions and the time of the year. With similar weather conditions it would be reasonable to assume that the predictions might be related by a pattern.

1.3.2.1 Stochastic processes

'Patterns', in associated unknown behaviour, can be modelled using 'probabilistic' or stochastic processes, a group of random variables that are indexed, for example being organised into a timeline. In this thesis I will use a stochastic process called a Gaussian process. A detailed comparison of different alternative methodologies can be found in chapter 4 of Wate et al. (2020). These processes make use of a series of Gaussian distributions that are indexed in relation to each other by regressor or 'prediction' variables. Gaussian processes have their origins in geostatistics for mineral exploitation in South Africa (Krige 1951).

Figure 1.3 shows a series of samples from a Gaussian process. In this case the regressor is the horizontal or 'x' axis and the prediction variable is the vertical or 'y' axis. The samples are curves and every point on those curves is a random variable again indexed by the position on the x-axis.
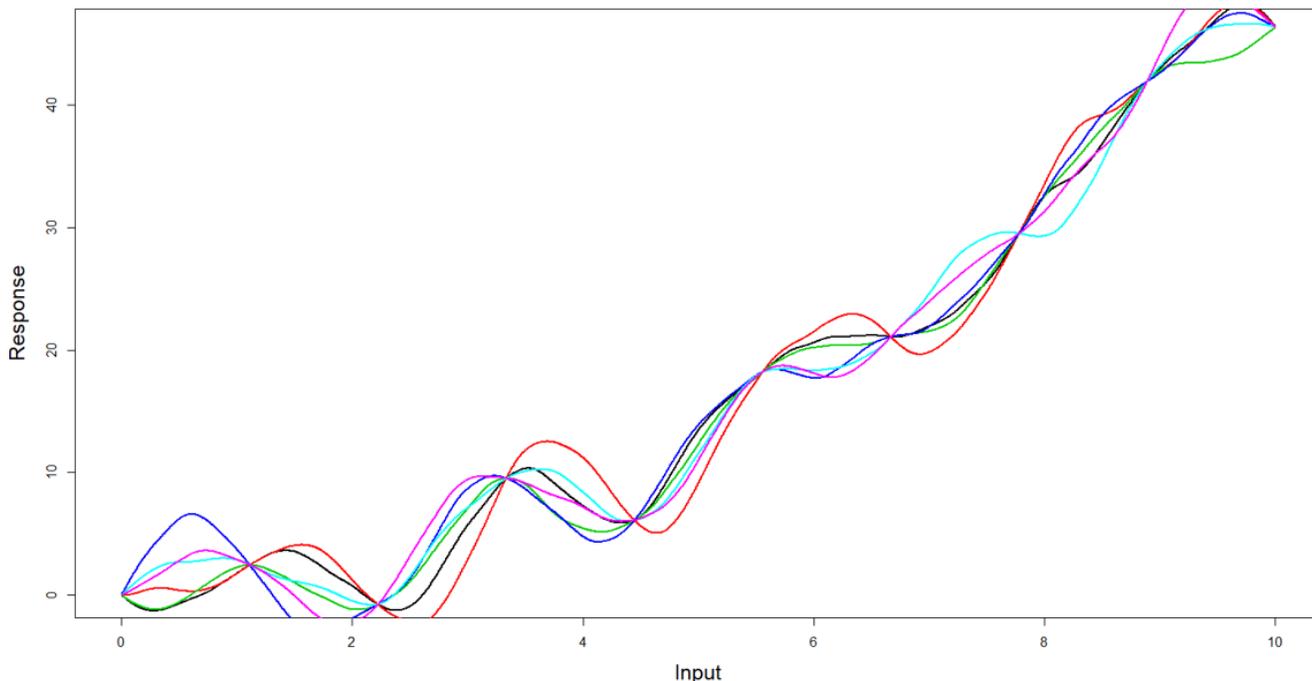
Figure 1.4: Sample from a Gaussian process posterior distribution with squared exponential co-variance function, the observations are at the 'pinch' points where the GP has learnt the data. In between the data points the samples demonstrate the possibilities the Gaussian process allows for.

Gaussian processes can as such be seen as a potentially infinite number of random variables and a description of how they relate to each other. The smooth nature of the samples in figure 1.3 is particular to some of the modelling choices for the Gaussian process.

1.3.2.2 ~~Model fitting~~

~~These s~~Stochastic processes ~~are~~can be built to summarise and explain data where the data contains uncertainty and variability. ~~As such these processes need to~~The process of matching the 'fit' to the data is important and needs to match the reasons for the model creation. In more general circumstances in machine learning or statistics this is described as model fitting. This can involve several elements carrying out calculations to analytically calculate the optimal model, numerical optimising model parameters to produce approximate solutions and manual ~~checking~~review of additional criteria for goodness of fit.

Empirical models are designed to model data rather than than make predictions from building a simulation of the physical mechanics. The model fitness is a process that examines how well the model numerically fits the data and then adjusts the model to improve it. Model fitness is a slightly more qualitative process where the model is examined

~~In the case of~~For Gaussian processes there are commonly two integrated parts to this. The first is to optimise the hyper-parameters of the Gaussian processes which with most model choices requires numerical optimisation. The~~ the second is to produce the posterior of the Gaussian process given the known hyper-parameters.

~~The numerical optimisation requires a criterion against which to optimise the hyper-parameters. With probabilistic models a commonly used criterion is the likelihood, called maximum likelihood estimation (MLE). The likelihood is the number given by the model to the possibility of the data occurring if it were the values of random variables from the model. This is not necessarily always available and there are occasionally issues with MLE that can produce issues~~problems ~~such as parameter estimation bias (Gibson et al. 2009)(in this case of a model of wave heights and crests).~~

~~Evaluating the posterior of the Gaussian process can commonly only be done analytically if the hyper-parameters are held fixed. Figure 1.4 shows a Gaussian process posterior that has been 'trained'. The samples all pass through the known values however moving away from these values on the x-values they vary more. The optimisation of the hyper-parameters will adjust and improve the amount and type of this variation away from the data points.~~

An introduction to the literature of Gaussian processes can be found in section 2.6, whereas justification for choices of covariance function and more mathematical detail can be found at the end of sections 3.2.5, 4.3.2 and 5.2.2.1. Chapters 3 and 4 provide application of GP learning to learn behaviour for prediction. Chapter 5 provides an example of how GPs can be used for inference where responses are non linear.

1.3.3 Atmosphere dynamics

The atmosphere over cities is subject to a number of factors that differentiate it from the micro-climate in other areas. Firstly cities produce a lot of heat directly from human activity such as combustion, for example transport, cooking, heating, or from electrical power inefficiencies such as lighting, cleaning, manufacture and leisure. The urban boundary layer is the layer in which the atmosphere of the Earth is affect by urban fabric. This is discussed in more detail in section 2.8. A particular example of this interaction is the urban heat island.

1.3.3.1 Urban heat island

~~This excess heat~~The effect of urban fabric warming the local climate is collectively referred to as an urban heat island (UHI). This effect has been empirically measured, for example in Birmingham using the Birmingham University Climate Laboratory (Bassett et al. 2016). The magnitude of the effect can vary: in Bassett et al. (2016) the magnitude is demonstrated to vary substantially on the basis of wind direction. Rajasekar & Weng (2009) and Rajasekar & Weng (2009) are papers that examine the magnitude of the UHI and both suggest that the urban heat island effect is much stronger during the ~~day~~night, this is explored in the results in chapter 6. **?** provides a more holistic review of the estimates of the intensity and magnitude literature suggesting that most studies do not quantify in a standard fashion the volume of urban heat islands. It recommends developing automated methods for assessing urban heat islands impact.

The UHI effect in Northern Europe is of limited concern and may even be of some limited positive effect in the most Northern cities providing marginally lower heating costs in winter. However in summer over-heating in the UK is becoming an more of a concern discussed by CIBSE (Bonfigli et al. 2017). As mentioned earlier in section 1.3.1, the issue of the UHI effect compounds these concerns. Chapters 3 and 4 directly aim to provide a modelling technique that can make providing building projection for future climates more manageable. The impact of temperature

change can also cause issues health as highlighted in Pastore et al. (2017), evaluating the impact of urban greening.

1.3.3.2   Reynolds Number, Roughness length and Rugosity

Rough surfaces are something that are ~~generally familiar~~describe as providing friction on fluids Stokes (1851). ~~When considering how a fluid flows over a rough surface it becomes considerably more complex than dealing with a flat surface.~~~~Generally speaking, in f~~Fluid dynamics differential equations ~~work on the basis that a fluid has~~ describe an infinitely thin selection of layers that ever so slightly slip over each other. The rate at which they are permitted to do this ~~is~~ and compress dictates the properties of the fluid.

Atmospheric modelling at low altitude over a city generally speaking is describing how the atmosphere slips over the ground combined with heating effects. One effect that is considered by some architects and city planners is the roughness of the city surface. In ~~modelling fluid flow~~CFD there are two main types of flow: turbulent and laminar. Laminar flow is where all these infinitely thin layers align into recognisable layers with different speeds and in the context of flow over a surface corresponds to a smooth surface. Turbulent flow is where the above is not true, flow is highly disrupted and changes direction over small length scales, unsurprisingly this coincides with rough surfaces. The Reynold's number provides a guide as to whether the flow is likely to be laminar or turbulent. It was suggested initially by Stokes in Stokes (1851), and later developed and named. It is derived from the ratio of speed of flow times length scale to the viscosity of the fluid. Low Reynolds numbers give rise to laminar flows with little turbulence. High Reynolds's numbers give rise to more turbulent flows that tend to chaotic behaviour.

The length scale in the context of cities is important as rougher surfaces lead to turbulence at lower wind speeds and are considered to be important for pedestrian comfort (Liu et al. 2019), adverse effects include wind tunnels between buildings and poor pollution dispersal away from cities as discussed in Wicht & Wicht (2018). Roughness length is an approximation in modelling laminar flow past a rough surface. It is the distance, where for the laminar flow away from the surface, for which the actual surface could be replaced with a zero slip smooth surface. ~~This is used to avoid the complexities of turbulent flow, which needs full element-wise approximations, instead analytic solutions can be used which are computationally cheap.~~This is due turbulent flow around heterogeneous structure causing the average velocity of flow to be essentially zero.   Rugosity is one of the metrics used to find actual city features that correspond to the roughness length and Reynold's number, examples of this work are Adolphe (2001) and again Wicht & Wicht (2018). In chapter 5, I explore this in terms of whether the rugosity has an effect on the air temperature over the city.

1.3.3.3   Albedo

Another element of the UHI effect is thought to be caused by the optical properties of a city surface. Unsurprisingly cities are generally a different ~~colour~~optical properties, which manifest in the visible spectrum as colour, from the surrounding rural areas. Tarmac is substantially darker than surrounding areas, this means more energy is absorbed from sunlight which is then remitted at lower wavelengths as infra-red leading to higher temperatures. Of course the magnitude of this effect is open to evaluation. This is the goal of chapter 6 to see what the magnitude of the effect

is compared to how urban the area is in general. The goal would be to see if areas with different albedo values in cities might be notably affected, which could lead potential adaptive measures that have been applied in other parts of the world (Bande et al. 2019), such as changing the colour of street and building roofs.

1.4 Summary Thesis Progression

In summary, the main research chapters in this thesis are:

Chapter 3 Introduces a method for reproducing the results of building energy model simulations, more cheaply than using the full simulations. It also provides an example of using likelihood to make decisions, in this case how to construct the replacement model or emulator.

Chapter 4 Practical implementation of using an emulator or statistical model to replicate the behaviour of the energy use of the chillers in a real building in order to create climate change impact projections for these chillers. This chapter builds on work from chapter 3 making use of the learnt modelling choices.

Chapter 5 This chapter examines evidence from the Birmingham University Climate Laboratory (BUCL) as to the impact of the shape of the landscape on wind speed over the city of Birmingham, with a view to whether it should be considered for city planning.

Chapter 6 Continuing analysis of the BUCL data this chapter brings in several other data sources to try estimate whether albedo has a distinguishable impact on the air temperature over Birmingham from general urbanisation.

I have also carried out a discussion of the literature in chapter 2.

1.4.1 Users groups

In summary, this thesis is composed of a number of projects that combine together to allow explore model user validation in different contexts and for different user groups. These user groups are:

- model architecture validation for modellers of building energy data. Chapter 3

- building operator researchers that use constructed models for planning. Chapter 4

- modelling for researchers drawing inference about the impact of the urban canopy on specific aspect of the urban micro-climate. Chapters 5 & 6

The first two users groups have related but different tasks and interests. This is how the chapters 3 and 4 relate to one another. The users of chapter 3 will be delivering to server the second user group. This gives the thesis the ability to provide insight on the workflow of a modelling process.

Modelling building energy data has a number of requirements. The first is accuracy of the predictions and the second is realistic characteristics of the predictions. These relate to the models they are building to for research on building operation. The accuracy is derived for the need for the results research to draw correct conclusions at decision points. The realistic components are to ensure that the model can be demonstrated to stakeholder groups. Application to climate change modelling aside from direct reference to the challenges facing buildings stock allow this thesis to address validation in relation to communication of uncertainty.

1.4.2   Role of chapters

The research model validation in chapters 5 & 6 is different in that it covers hypothesis testing and causal inference. They make use of different public resources in terms of data. Chapter 5 makes use of a resources designed for research as such outputs are well defined and standardised. Chapter 6 however makes use of the google maps resource that is not standardised for research, rather for public consumption. This gives a coverage of purpose and non purpose built data to evaluate as suitable for research on the impact of the urban fabric.

In addition to covering casual inference, chapter 5 & 6 expand the modelling targets to cover two weather variables that affect thermal comfort. Section 2.12 provides discussion of how the ENVI-met model was designed to take into account both of these elements. The plants modelled in this particular study highlight sensitivity of elements the urban fabric to the weather. Wind speed and air temperature were picked as they are well characterised and provide opportunities for examining the impact of the roughness length on wind speed and albedo on air temperature. A third concept of the the impact of vegetation on air humidity would be interesting, however was not explored in this thesis.

The element in common between chapter 5 & 6 of holistic micro-climate research in Birmingham form a contrast to the BEM model example in chapters 3 & 4. This gives coverage of two key areas of the urban micro-climate: buildings affects by climate and climate affected by urban fabric.

This thesis also covers both black and white box statistical methodology:

- Black box approaches

  - Chapter 3 focuses on model architecture and making choices between different models. The metric used to prefer one model are challenging to interpret for non experts and their explain-ability is an important user requirement. Black box models being generally more complex to explain it is useful to examine how this compromises the user's needs.

  - Chapter 4 provides inference for a real world asset. The user group has real world problems to consider such as future investment decisions and managing the building on a day to day basis.

  - Chapter 5 experiments with using using additive model with a Gaussian Process model to determine if there is significant predictive power in the variable.

- White box approaches

  - Chapter 3 include a classic BEM model as the basis for it is model selection which helps assist with explicability as building energy model are physics-ally interpretable by nature. Allows for a direct side by side comparison with it's black box model.

  - Chapter 5 uses linear regression which is highly interpretable and gives estimated confidence of results.

  - Chapter 6 uses linear regression on partitioned data and provides evidence for causal links between input and output variables in the models. Casual inference is an important part of determining the distinction between correlation and causation. Chapter 6 is able to present inference on this topic and facilitate the discussion on how researchers deal with the problem.

Chapter 2

Background ~~and state of knowledge~~

~~Drawing inference from data and simulators in urban environments~~

2.1 Introduction

General analysis of the urban environment is complex. City fabrics are constantly evolving from changing technologies, the wider environment, human behaviour and policy. Climate change of the weather system is an influence on various parts of the urban fabric, such as building energy use Yassaghi & Hoque (2019). Another example of this change is increased prevalence of population heat stress from higher summer temperatures requiring greater air conditioning capability (Chow et al. (2002)).

The goal of this review is to provide an introduction to models and model fitting. The purpose of models in this thesis is to provide statistical inference models with confidence information. This thesis has a common thread of examining how different models are suitable for their users in uncertain circumstances, in the case of studies found in chapters 3-6 how microclimate and cities interact. As such this chapter introduces uncertainty and sensitivity analysis and the complexities of applying them to building energy modelling (BEM) and urban weather environment literature. In order to facility sensitivity analysis this thesis will need models that ares suitable for modelling complex unknown responses.

This chapter provides a discussion followed by two case studies. ~~This provides an opportunity to~~It explores the nature of EnergyPlus as an example of BEM, ENVI-met as an example of urban environment modelling and examine some of the statistical context around for this type of modelling.

Kavgic et al. (2010) identifies various approaches to modelling of building energy use, choosing to subdivide methods into bottom-up and top-down approaches. Top-down approaches, compared with bottom-up approaches, link the performance of a class of buildings such as residential four bed properties in the UK and their response to macro-economic factors such as GDP, national holidays or average temperature in a time period with a goal to predicting national/regional energy use without getting bogged down by uncertainty relating to specific local circumstances (Summerfield et al. (2010)).

Bottom-up modelling is approached by working from elements specific to the area being modelled. For a BEM this might be: building size and degree of occupancy (Zhang et al. (2013)). Whilst in the case of urban mirco-climate modelling specific elements could be: building topology and current local environmental conditions. This contrasts with top-down approaches that link

classes of buildings across geographic or economic regions and the energy use's link with information such as GDP and average temperatures across the region. Forms of statistical regression are used in both cases.

Simulation or physics based models are commonly associated with bottom-up models where components of the model can represent relatively simple systems such as walls or latent heat flux from a specific green area. A comparison of one example of ~~each approach~~ using both a black box neural network model to model building energy use and a 'physics-based' BEM can be found in Neto & Fiorelli (2008), this is discussed in one of the case studies.

Bottom-up evaluation of energy use and weather for specific sites is complicated by the numerous potential interactions: latent heat flux from water or plants; on/offshore breezes for coastal areas; seasonal weather patterns; changes to building use. Whilst simulations and statistical models can account for some of these, many interactions, inevitably remain outside the sphere of available information, and therefore have a detrimental impact on predictions. This is referred to in Kavgic et al. (2010): "Currently, models often fail to deal adequately with the interactions that occur with different aspects of energy demand, particularly socio-technical factors."

Much research has been carried out looking at the sources of uncertainty in building energy such as climate change reviewed in general in Yassaghi & Hoque (2019) or for example a specific model for occupancy behaviour Wilke et al. (2013). Whilst the charge of 'models often fail to deal adequately with interactions' is obviously subjective and nearly 10 years old, the phrase represents a discomfort with the so-called performance gap that exists in post-occupancy buildings. The gap between modelled performance prior to construction and actual performance. Marston (2018) is an example of how this gap can justify remedial workin the context of providing a retrofit to achieve the predicted energy performance in the design. This thesis focuses on bottom-up approaches specifically wind speed and air temperature.

Urban weather can broadly be termed as meso-climate to micro-climate.The distinction is in the agglomeration of area: meso-climate refers to areas between a few kilo-meters square to tens of kilo-meters square; micro-climate in contrast is considered to be around 1-2 kilo-meters square. Meso-climate effect for example might be the atmospheric drag provided by a city Kanda et al. (2013). Micro-climate is a term used to refer to effects such as airflow around a building or the shading of the urban canopy. Efforts to provide sample weather incorporating these effects can be found in the Urban weather generator Bueno et al. (2013). Models are often integrated to provide larger scale models such as the Met-Office climate projections (Met Office: Hadley Centre (2019)). These models are linked into collections of models such as HAD-CM3 (cite) as the UK met-office model.

For wind speed and air temperature over the urban fabric, research are based on the known threats of overheating (Wong et al. (2010), Solecki et al. (2005)) and air pollution(Jeanjean et al. (2016)) and attempts to mitigate them. In urban weather the research analysis of producing observations of the urban climate and simulation of the urban climate are surprisingly disjointed. Simulations of the climate are generally computational fluid dynamics (CDF) models which have their predictions consisting of complete fields of temperature, wind speed and various pollution concentrations, whilst observations are lists of values at specific locations, for example ENVI-met discussed in the case study at the end this chapter 2.12 contrasts with the observations in the BUCL discussed in the appendix A. Here ENVI-met allows the CFD method to be carried out.

2.2 Modelling physical systems and model fitness methodology

Physical system modelling aims to capture empirically measurable elements in the real world parallel. Inevitably this leads to prioritisation about which elements need to be captured due to resource constraints. As such discussions of the model's validity centre on whether measurable elements of the real world system are captured.

The concept of value can be exhibited as the use case of the model or the needs of the user group. In Robinson (2011) 5.2.2 the omission of thermal flows is discussed. The first model the focuses on is on urban texture and micro-climate therefore condones the decision to not model occupant behaviour. However, this motivates the development of a wider resource flow model in the chapter to support master planning decision making. The new context demands that new elements are included in a revised model. At the end of each chapter a section discusses the limitations of the approach adopted in the chapter from this perspective.

The methodology used to examine model fitting is largely speculative based on the author's experience. The goal of this thesis is to provide a systematic structure discussion that is comparable between chapters on different topics. The resultant analysis could make arranging interviews to fully evidence the analysis. Model fitting will be analysed on the following points:

- User requirements: the requirements of the modelling/inference carried out in the chapter. This included specifying the nature of the end user an their expertise, hence their capability to examine and understand the model results.

- Chapter Outputs: The nature of the results of the chapter.

- Needs met and evaluated: A examination of how the Chapter outputs meet the user requirements.

- Recommendations: How the work in the chapter might be improved to meet the user requirements better.

In the context of modelling for statistical inference, the main tools are sensitivity and uncertainty analysis. As will be discussed in section 2.5 there the interpretability of models is an important element in the model's composition. However, before considering this, an introduction into sensitivity analysis and uncertainty analysis is discussed below.

2.3 Introducing sensitivity analysis

Broadly, statistical sensitivity analysis should describe how a response varies ~~in response to~~ with it's inputs and quantify the degrees with which the response varies to inputs individually and in combination. Usually this is associated with the sensitivity of a model to inputs, often a computer model. Moonen & Allegrini (2015) carried out sensitivity analysis on a CFD simulation to the impact of simulation boundary wind direction, speed and other effects on the rest of the simulation. Saltelli et al. (n.d.), a book introducing sensitivity analysis methods, highlights the very broad variety in sensitivity analysis methods.~~, can be considered a good reference point for the theoretical work on sensitivity analysis.~~

Challenges with applying this broad approach are that both in the actual world environment and in models there are often very large numbers of inputs. Some are continuous such as building height (even if they do not vary in reality) and others such as topology of buildings are more complex. Sensitivity analysis is generally subdivided into two areas of application: local and global sensitivity. Local sensitivity considers the impact of change to the input parameters at a particular set of values. Global sensitivity analysis tries to analyse the sensitivity of a system with respect to the full range of input values.

The textbook Castillo et al. (2007), ~~is an example of~~ details multiple methods for carrying out ~~a form of~~ local sensitivity analysis. Closer to the specific subject matter, Liu et al. (2019) evaluated a large eddy simulation that is used to study airflow around buildings for pedestrian comfort. This study examines the impact of various simulation parameters on the airflow at street level. Liu et al. (2019) assessed recommended values for mesh resolution, upstream distance, length of sample period and vortices number. The sensitivity analysis demonstrated that in each case there is substantial variation for different values on each parameter in the wind speeds which would have an impact on pedestrian comfort.

Tian & De Wilde (2011), evaluates the impact of uncertainty from probabilistic climate change forecasts, providing an example of simple global sensitivity analysis. This paper examines the impact of climate projection on a building simulation. The value of each weather variable at each time point is essentially an input affecting building performance. The possible scope of variation in the input space is huge, and in places probably of limited relevance (perfectly uniform weather for example is unlikely). Tian & De Wilde (2011) uses sampling from a collection of realistic inputs, in this case UKCP09 met-office weather generator Met Office: Hadley Centre (2019), todetermine the sensitivity of these inputs with respect to realistic values. The sensitivity is testing the broad sensitivity in four different regimes of inputs, in this case, baseline climate scenario and three scenarios for 2050's. The paper examines the spread and location of values for each scenario and compares them using box plots. The discussion in this work concluded that the future scenarios represent a significant increase in cooling degree days and decrease in heating degree days for the modelled for a UK multipurpose university building.

Several classes of sensitivity analysis methods highlighted by Saltelli et al. (n.d.):

- Elementary effects methods work by estimating the derivatives in local sensitivity analysis focusing on a subset of input variables. The derivatives explain how fast the output varies in relation to the input. In terms of the real world this is often close to the practical scenario in that data is gathered on key related variables to ~~the variable~~ the investigation is interested in predicting. For example in chapter 4; I use information on the local micro-climate to assess its impact on the building's cooling energy use. Boukouvalas et al. (2014), providing methodology for screening inputs of computer experiments, such as building energy models, CDF and other finite difference models, uses this approach to propose a method to reduce the complexity of the optimisation problems.

- Variance based methodology assesses the impact of an input of the output by considering the amount of variation in the output caused by varying the input. The box plots that are used to summarise the output of different scenarios in Tian & De Wilde (2011) is an example of this sort of analysis. It is also important for scenarios where the input is categorical rather than continuous as derivatives cannot exist. Delgarm et al. (2018), in a comparison of sensitivity analysis methods for BEMs, uses this type of approach for global sensitivity analysis. This work identifies window size amongst other factors as being the most significant factor in

global sensitive analysis of cooling, heating and lighting electrical energy use, which support the general industry expectations.

- Experimental design approaches are used with ~~computer~~simulation and real world experiments to make each data point more useful for doing sensitivity analysis. This is much more cost efficient in terms of computational resources in the case of computer models or physical world cost if carrying out practical experiments. One of the earliest papers to introduce experimental design was Sacks et al. (1989). For applications, Busby (2008) is an example of using experimental design to improve carrying out optimisation using Gaussian processes.

2.4 Distinction between uncertainty analysis and sensitivity analysis

The distinction between uncertainty analysis and sensitivity analysis is not entirely distinct andis often defined by the model used. ~~Broadly s~~Sensitivity analysis is summarising a change in a response to an input or inputs. Uncertainty analysis is about evaluating the lack of knowledge about a response. It does not need to be with respect to particular inputs. However when examining the structure of this uncertainty, whether it relates to any known inputs, distinctions of whether a method is ~~UA or SA~~uncertainty or sensitivity analysis become dependent on how the methods are framed or phrased rather than whether it is solely associated with sensitivity or uncertainty analysis techniques. In general, more specific methods relating to the differential of a response are sensitivity analysis whereas meta-analysis of such problems using other approach will be uncertainty analysis, exemplified in MacDonald & Strachan (2001).

Chapter four will look at decomposing the source of uncertainty where the knowledge of a piece of information is uncertain. Often measurements at input values will leave a lack of knowledge as to the output value. This lack of knowledge is usually summarised by a probability distribution which gives structure to the lack of knowledge about the uncertain values. Probabilistic classification models have a simple example of uncertainty summary in the form of assigning a probability to each class~~, m~~. More complex examples such as Wilke et al. (2013) ~~which~~ provide~~s~~ a model for modelling the occupancy patterns of the building occupants.

2.5 Comparison of different non-linear response models

The main different model that might be compared within machine learning for non linear regression vary substantially in terms of their properties. An authoritative description of neural network and Gaussian processes comparison can be found in Rasmussen et al. (2006)'s preface and later chapters. This thesis provides a short summary of linear regression, Gaussian processes and neural networks:

- Gaussian process and other statistical models are aimed to be explicable in various ways:
  - Linear regression makes simple assumptions about how input relate to outputs. The basis functions, which are the function that modifies model input, are multiplied by a coefficient. The value of this coefficient therefore can be interpreted to find the importance of the basis function. These basis functions can be the identity function so the input variable is unmodified. This highly interpretable and this thesis uses this

approach in chapters 5 & 6to provide inference as to the importance of input features with physical significance.

– Gaussian process are discussed in contextual detail below are an extension of linear modelling making use of a mathematical manipulation known as the kernel trick to represent different groups of infinite basis functions. The most common interpretation for these groups are the resultant covariant functions which represent how model outputs varies as input changes. Information such smoothness can be encoded by picking/optimising covariance functions and hyper-parameters. The down side Gaussian processes is the training methodology is highly costly for large data sets. Being order n^2 for memory and n^3 for computation times where n is the number of data points. Simple Gaussian processes are suitable data-sets of the order of several thousand data points. This thesis uses Gaussian processes in chapters 3, 4 & 5.

– In summary, linear regression is used for creating highly interpretable models and Gaussian process for flexible models encoded smoothness of outputs.

- Neural networks represent a simpler representation of the realisation that sufficient number of non linear basis functions can capture any behaviour.

  – This primitive nature means understanding the problem can be approached with a poorer understanding the nature of the data. Often this entirely expected with large complex data-sets.

  – Lower complexity of the model setup mean that with respect to statistical properties they encode less information of the model from the creator's understanding. This information then needs to be derived from the data-set. Neural networks are often parameterised, making optimisation difficult. They are simply searching too large a parameter space.

  – Neural network carry substantial advantages with large data-sets due the way back-propagation works. The standard gradient descent method will pass through the data line by line opposed to simultaneously evaluating the relationships between all data points. In the comparatively simple standard linear regression requires evaluating a n^2 matrix for n data points which is memory prohibitive if not careful evaluated.

  – In summary neural networks are excellent for large data-sets where other methods infeasible and encoding the model creators expectation is difficult, some other alternative include random forests and genetic algorithms, for a comparison of genetic algorithms and neural networks see Wood et al. (2015).

2.6   Introducing Gaussian processes

Gaussian processes (GPs) are a type of probabilistic model with parameters that do not have direct interpretation. They are ideally suited to tasks with small to mid sized data-sets due to needing to load the full data-set, the strong assumptions encoded through covariance choices allow them to be informative in scenarios with limited data. However they have an advantage over neural network models, with which they often compared (Neto & Fiorelli 2008), in that they are more easily informed by smaller data-sets making them much more useful in the context of limited data. This is particularly relevant around urban environments, this thesis encounters in chapter 5, as

whilst in some ways there is a lot of information the coverage is sometimes very poor compared to variation, though as will be discussed in the results of chapter 5; these scenario are challenging for models.

The most common formulations have scale and length parameters. The scale parameters adjust how fast the GP will change ~~movign~~moving through the input space. The variance parameter adjusts the maximum possible variation in the GP response. A GP also uses the input data as additional parameters. The training of a GP adjusts how the GP uses this information, rather than no longer needing it after the training is complete. However this does mean the model can be transferred 'pre-trained' to another data-set if retraining isn't feasible ~~for some reason~~. The exact formulation of a GP's parameters will depend on the implementation, as such they are introduced in chapter 3 & 4 as they are implemented for the models in those chapters, sections 3.2.5 & 4.3.2.

Gaussian processes have been used in an array of different contexts. Gaussian processes are advantageous when dealing with uncertainty for several reasons. They naturally quantify uncertainty when fitted by maximum likelihood and can extend to high dimensions relatively well Rasmussen et al. (2006). They have also been successfully applied to many applications from Galaxy formation models Vernon et al. (2010), cardiac models Ghosh et al. (2018), CFD emulation domain for BEM Wood et al. (2015), Heo & Zavala (2012).

Gaussian processes work by associating the output prediction at location with each other and the training data. These are not perfectly correlated but have increasing correlation as their input values get closer together. The definition of this closeness is defined by a covariance function for which there are many known definitions that make Gaussian processes a very flexible class of machine/statistical learning tool. The uncertainty grows as the prediction is made further away from the input locations of the training data/observations, allowing the model to provide a good summary of the state of knowledge at different points in the input space. Most covariance functions also give the Gaussian process analytic derivatives which is useful for local sensitivity analysis.

2.6.0.1   Model fitting with Maximum Likelihood

~~In the case of~~For Gaussian Processes model fitting there are commonly two integrated parts to this. The first is to optimise the hyper-parameters of the Gaussian processes which with most model choices requires numerical optimisation. The ~~the~~ second is to produce the posterior of the Gaussian process given the known hyper-parameters.

The numerical optimisation requires a criterion against which to optimise the hyper-parameters. With probabilistic models a commonly used criterion is the likelihood, called maximum likelihood estimation (MLE). The likelihood is the number given by the model to the possibility of the data occurring if it were the values of random variables from the model. This is not necessarily always available and there are occasionally issues with MLE that can produce ~~issues~~problems such as parameter estimation bias (Gibson et al. 2009)(in this case of a model of wave heights and crests).

Evaluating the posterior of the Gaussian process can commonly only be done analytically if the hyper-parameters are held fixed. Figure 1.4 shows a Gaussian process posterior that has been 'trained'. The samples all pass through the known values however moving away from these values on the x-values they vary more. The optimisation of the hyper-parameters will adjust and improve the amount and type of this variation away from the data points.

2.7   Industry application of uncertainty and sensitivity analysis in building energy modelling

The industry application of modelling of building energy simulation is to inform decision making prior to construction or modelling of a building. When considering modification or construction there is a modelled expectation of the building performance. The gap between this expected performance and actual building performance once occupied can be substantial. Cases of the actual performance being two to five times the predicted values have been reported Marston (2018). Industry interviews discussed in Zou et al. (2019) provide some explanations such as uncertainty in models, inaccuracies in design and limited feedback.

From the perspective of uncertainty analysis, lack of feedback and poor problem specification are more easily considered. There is a larger issue in the industry that most projects do not have after-completion assessment, in the case of construction more specifically post-occupancy evaluation, Li et al. (2018) (p12) highlights how ~~the that~~ there are a number of conflicts that act as barriers to uptake of post occupancy evaluation (POE):

> . . . .the ambiguity of who pays for POE, defending professional territory, split incentives within the procurement and operation processes, lack of agreed-upon and reliable indicators, potential liability issues, exclusion from current delivery expectations, and exclusion from professional curriculum. . . .

As such the academic community extolls the benefits of post occupancy evaluation, there is also an awareness that wide scale adoption of post occupancy evaluation is not imminent. Durosaiye et al. (2019) highlights that more information is being accumulated by building management. Referring this information back to the building design process is largely unimplemented.

Uncertainty in terms of the performance gap occurs in two main contexts, new build and retrofits:

- Design of new buildings generally are designed using computer models that provide forecasts of building performance. The gap between the performance of the end product and the design is the performance gap. Identifying the sources of this certainty is an area of research interest: Singh et al. (2016) investigates the uncertainty around the impact of blind shading on energy use and lighting; Wang et al. (2012) explores the uncertainty from building operation and weather.

- Retrofitting buildings involves proposing changes to existing structures to improve the performance of the building. The performance gap comes from the improvements not achieving the performance expected. Heo et al. (2012) uses Gaussian processes to better evaluate the baseline for evaluating the actual improvement in building performance in building energy use. Marston (2018) reviews the retrofit of an air conditioning system, highlighting the mistakes in the original design which lead to the retrofit. In this case the goal was to improve the design to get the energy use more in line with original design forecasts.

2.7.0.1   Sensitivity analysis applied to BEM

In terms of sensitivity analysis there is significant use of linear modelling with significance testing as a method of global sensitivity analysis. Zhang et al. (2013) Bamdad et al. (2018) both make use of linear models without mention of other methods. It should be noted that this is in line with

industry practice for real building design which obviously follows guidelines. For example ASHRAE (2014) for which the recommended way to demonstrate a retrofit building energy change to external influences is with a linear regression model.

This is in contrast to many papers, such as Neto & Fiorelli (2008), Bamdad et al. (2018) , discussing the non-linearities in building energy systems: Bamdad et al. (2018) raises non-linearities as motivation for the proposing Ant Colony Optimisation algorithm for Multiple Variables for building energy systems; Neto & Fiorelli (2008) discusses in more detail of the non-linearities of actual building energy response to external stimuli. This is highlighted by comparing a predictive linear model in the form of a neural network actual building performance, finding the predictive model has very limited capabilities. This knowledge that building systems have only very weakly linear responses makes inference based on linear models limited in terms of its accuracy.

Whilst linear models are well understood there have not been examples of attempts to use many higher order interaction terms to study responses on the whole , for BEM. This is could be due to difficulty to exploring the vast number possible basis functions and interaction terms at higher orders. Wood et al. (2015) is an example of utilising nonlinear models to optimise systems and Heo & Zavala (2012) is an example of using non-linear models to assess energy saving using nonlinear models: indeed the ASHRAE (2014) guidance is to use linear models for estimation of retrofit savings.

Occupancy is a complex contributor to general uncertainty. Most models are built with significant assumptions about how occupants use buildings, the strength these assumption makes the issues around occupancy uncertainty large. For example, Neto & Fiorelli (2008) makes assumptions about how the windows are opened, which was found to create issues with the model being unable to model energy use early to mid morning. However as the modellers' generally note Neto & Fiorelli (2008), these assumptions are a lot less complex than the pattern of behaviour by actual occupants, however there is considerable debate and cost around obtaining more accurate data. Wilke et al. (2013)provides a statistical model structure for modelling occupant behaviour. Design models are only as accurate as the information provided. As a result of most buildings having some structural or functional modifications made during the construction ~~phase~~, the predictions and models from the design phase are ~~partially invalidated~~much less accurate.

In summary dealing with uncertainty in a wider context of building energy modelling:

- Current modelling does not reach high precision accuracy for predicting building energy use. Reasons are varied and partly logistical, and not easily accounted for such as the building being used differently to its original design intention.

- Evaluation of uncertainty is limited due to lack of post-occupancy evaluation and the linearity of statistical models which means the uncertainty is poorly specified.

- Global sensitivity analysis is usually evaluated using linear modelling significance tests or complex situations makes use of variance based methods.

2.8 Boundary layer in the urban environment

In this thesis the focuses on the atmosphere surface interaction at the local scale. The boundary layer, in general fluid dynamics, is the layer of the fluid where fluid flow is not part of the main fluid flow (Young 1989), this a well establish concept (n.d.a). Urban environments interact mostly with this layer and it can be broken down further (Roth 2000), exact length scale will vary on the underlying terrain:

- Urban Boundary Layer: grouping of other layers, defined as proportion of atmosphere affected by presence of an urban area.

- Urban Canopy Layer:Produced by microscale or micro-climate effects or site characteristics, thermal characteristics are dominated by the local environment. Complex non-uniform flows. Combined with the roughness sub-layer this is the environment in which the BUCL data (Warren, Young, Chapman, Muller & Grimmond 2016) resides for chapters 5 & 6.

- Roughness sub-layer:Transition layer where mechanically affected by roughness length scales, turbulence is non uniform.

- Constant-flux layer: Mechanically and thermally the profiles obey semi-logarithmic laws. This is general well above the height of most buildings so less well measured, and will of be less interest to this thesis.

- Mixed layer: Highest level, characterised by interaction with air mixing after interacting with rough and warm surfaces with the wider atmosphere. Similar to the constant flux layer.

When considering which factors affect the atmospheric parameters as such velocity, air temperature, which this thesis works with chapters 5 & 6, Bou-Zeid et al. (2020) talks about the issue in identifying issues with heterogeneity in under lying terrain being driving factors in heterogeneous boundary layer behaviour. Chapter 5 looks at trying to find the effect of heterogeneous roughness length in the BUCL data. This addresses the mechanical effects and is contrasted in chapter 6, which is in response to evaluating the empirical evidence for the strength in albedo effect on local air temperature, posited as a key part of surface energy budget studies highlighted in work such as Qu et al. (2015).

2.9   Uncertainty in urban weather

Within cities, urban weather can be quite heterogeneous with small pockets of sunny, still-aired protected back gardens to large urban canyons in city centres which can act as wind tunnels with considerable shading. Considerable bodies of research exist, which look at the impact of urban features such as the urban roughness length for example Bottema (1997), Kent et al. (2018a,b).

Rose & Apt (2016) in particular looks at uncertainty in wind speed modelling and produced a model designed to produce corrections for ~~computation fluid~~CFD simulations by fitting a model with information on seasonal roughness lengths and height profiling. The results of their uncertainty analysis are that the wind speed is poorly predicted in the unstable region of the atmosphere inside the roughness layer. This roughness length ~~in fluid dynamics~~ is the ~~distance above a rough surface~~ratio of surface area to width facing into a flow~~, rough~~ in the atmosphere from trees and buildings~~, at which point the mean lateral wind speed is zero~~, roughness length and other related concepts are formally defined in section 5.2.1.

Rose & Apt (2016) identifies several other possible sources of uncertainty for the wind speed from the literature such as state variables (ie instantaneous measurements) from Decker et al. (2012), simplification of terrain Carvalho et al. (2014), and high wind speeds from higher altitudes Sharp et al. (2015), in all these cases there was no unambiguous evidence that they contributed to the uncertainty. That is not to say that they are not significant rather that evidence could not be found. Wind speed and other variables are often difficult to draw conclusions about due to the limit of sensor placement locations compared to the complexity of local environments.

Comparisons between these wind fields and the point observations within this field are particularly difficult due to the noise that is often associated with such observational sensors, studies that involve such networks include Munir et al. (2019), analysing the use of air low quality sensors for monitoring pollution and Bassett et al. (2016), analysing the extent of an urban heat island in Birmingham. Aligning empirical observations with models such as BEM or CDF models is difficult. The community have the approach of 'forcing' models to meet empirical measurements. This usually involves using the observations as boundary conditions to the model.

There exist two main approaches to align CFD models with observations. The primary way in which this is done is by forcing the CFD simulations to have consistent values with the observations at the locations, for example Gusson & Duarte (2016). This is somewhat different to a statistical attempt to calibrate computer models, such as is more commonly done with BEM with examples being Heo et al. (2012), Menberg et al. (2019). The main advantages are that the forced model is consistent with the observations, it is relatively computationally cheap to carry out and that model physics are adhered to. The disadvantage is that the method is not designed to improve general predictive accuracy measures such as Root Mean Squared Error (RMSE). The other class of approaches to fit the CFD model using criteria such as likelihood or RMSE. Jiang et al. (2017) carries out a case study of an indoor thermal heat map with sparse sensor observations (4 sensors). In order to calibrate their CFD model with their sensor network, they make use of an internal probabilistic model to create the likelihood criteria.

ENVI-met V4 tool, an urban mirco-climate modelling tool, presented in the example below contains a forcing tool that enables users to force their simulation to match with observations, two examples being Gusson & Duarte (2016), studying the effect of built density on urban mirco-climate in São Paulo and Shinzato et al. (2019), comparing built up and vegetated areas around São Paulo.

Building energy use is often modelled to respond to external weather conditions, which can in turn be outputs of CFD simulations P. Rastogi (2015). For urban weather CFD simulations differentiations urban terrain such as urban greenery Vuckovic et al. (2017) or roughness length Kent et al. (2018b), Lukač et al. (2017) are considered inputs with uncertainty associated with them being issues that make drawing conclusions more complex.

In summary urban weather has particular features for uncertainty analysis:

- Observation data is often quite sparse compared to the modelled information such as wind profiles and pollution density.

- Calibration of models is often done by forcing models to be consistent with observations which doesn't necessarily improve predictive performance.

2.10 Summary

Holistically, sensitivity and uncertainty are problems the research communities around BEM modelling and urban weather are very aware of. Both operate in a highly complex environment that has many influencing factors. They also face major limits on information gathering to handle some of these uncertainties: both cost and lack of post occupancy evaluation uptake limit BEM modelling knowledge of uncertainty; for urban weather the cost of measuring weather at many locations at different height profiles means it is difficult to get good coverage of the area of interest.

There are techniques that exist in the statistics community that can be used to address this uncertainty, this thesis provides a review at the end every chapter on their fitness to the task

at hand. Gaussian processes are used by both research in BEM and urban climate to manage uncertainty in modelling and observations. Gaussian processes are more often used in BEM, but as Moonen & Allegrini (2015) shows there is ~~penetration~~increasing use ~~into~~ urban weather research. Variance based methods are used to summarise the uncertainty in BEM around weather based uncertainty analysis. The picture is somewhat unclear with work around CFD simulations focusing on forcing rather than predictive accuracy of models.

2.11 Case Example: Comparison between detailed model simulation and neural network for forecasting building energy consumption

Forecasting building energy use is often done to meet a variety of tasks such as optimising building management in real time or making longer term management decisions. In Neto & Fiorelli (2008) the Administration Building of the University of São Paulo is an office block with two six-story blocks, angled 43 degrees Northwest and user controlled air conditioning. It is occupied by around 1000 people mostly between 0800-1800 over its approximately $3000m^2$ of floor space.

2.11.1 EnergyPlus building energy model

At the start of the study several inspections of the building were carried out. The building was then modelled using the EnergyPlus modelling software (Energy2010). EnergyPlus is a well established industry standard that is originally funded by the U.S. Department of Energy's (DOE) Building Technologies Office (BTO). Models are comprised of a thermal capacitor resistor network that is based on the input 3D model with parameters derived from specified details such as wall elements, their thickness and composite materials. The exterior of this model is subjected to the exterior climate conditions specified by an .epw file. These might be sourced from observations or from a formatted predictions from an urban or larger scale climate model such as the HADCM3 of the ocean and atmosphere.

EnergyPlus is set up to manage linked zones, which are nodes in the capacitor network. Zones can comprise entire floors to single rooms. The basis of what should constitute a zone is that there should be relative homogeneity across the zone in terms of conditions. Multiple identical rooms in the middle of a floor would be better grouped, whereas two sides of a floor with radically different incident solar radiation should be modelled separately. Elements of EnergyPlus compromise:

- Air systems temperature and air mass are integrated finite difference solver for the change in the temperature of the zone matched with different sources: convective internal sources (electronic equipment for example) and sources from connected surfaces (such as walls), loss and gain from air mixing/inflow/outflow. Internal gains can be set to follow programs.

- Moisture levels are then calculated from the air mass balance at each time step.

- Other systems can be modelled and added as linked models. Several platforms that link EnergyPlus to other tools such as visualisation software (SkechUp) and design tools (Openstudio).

The advantages are that it is a well respected modelused across industry (Bienvenido-Huertas et al. 2019, Wood et al. 2015) that gives results credibility, it is efficient to run (in that is well coded)

and is highly flexible. One limitation is the assumed uniform conditions inside zones, for example this can be problematic for zones with high incident solar radiation as the side receiving this can be much warmer than the rest of the zone. This can be addressed by linking EnergyPlus into modelling software suites such as products offered by IES, as EnergyPlus input files can execute other programs and recover values from them; this allows them to run CDF simulations to inform zone boundary fluxes. Alternatively much more simply zones can be split to address heterogeneous boundaries.

2.11.2   Neural networks for modelling building energy

Neural networks are diverse machine learning tools that have been adapted for a large variety of tasks such as prediction or classification in diverse fields such as image recognition and energy use forecasting.

They consist of a network relay node that takes inputs from other nodes in a previous layer and combines them and then the output is relayed to the next part of the network, known as the activation function (Bishop 1994). The values being passed from the outputs of nodes to the inputs are adjusted by values that are learnt. This learning process can also be used to improve other parts of the network, such as parameters at the node, the number of nodes in a layer in the network or the number of layers. The learning of the features of the different elements of the neural network is where the incredible potential flexibility of networks comes from as well as the difficulty in successfully using networks, there are many poor choices for such values as well as potentially many successful choices.

2.11.3   Modelling Energy use for a university administration building

The results for modelling the predicted hourly energy consumption of the administration building were broadly successful for both models. The wider pattern of energy use is captured. The morning increase to daytime load is captured along with the end of the work day. EnergyPlus, in this example, underestimates the energy use in the morning active period. The author attributes this to poor modelling of 'adaptive thermal comfort': occupants opening windows to adjust the temperature to be more comfortable, which is not modelled in this study.

Both modelling methods, EnergyPlus and the neural network, in this study achieve similar levels of accuracy in terms of absolute error with around 80% of results within 13% of the measured value. There does seem to be strong linear correlation between the forecasts of both EnergyPlus and the neural network forecasts and the measured energy use, however the gradient has a very shallow slope over estimating lower values and underestimating higher values. The neural network did attribution more variation to the external weather factors such as sky cover which affects the impact of solar radiative heating.

In terms of uncertainty and sensitivity analysis Neto & Fiorelli (2008) uses percentage errors to measure divergence of methods from observed values, which are simple and easily understood and communicated. For example the uncertainty in dry bulb air temperature at $\pm 1.0°C$ was found to have a $\pm 1.2\%$ impact on energy consumption.

2.11.4   Summary

The neural network employed in this study is relatively primitive, the activation function used for the nodes in this work was a linear activation function. This activation function has the dis-

advantage that the neural network is equivalent to linear regression on the inputs which limits the complexity of the relationship the network can represent. The fact that essentially linear regression has similar predictive power to a physics based model which requires a degree of understanding of building design to implement, highlights the difficulty of integrating knowledge of physical principles with complex examples in the real world.

2.12   Case Example: ENVI-met Core model and Modelling transpiration and leaf temperature of urban trees

Simon (2016) is a thesis detailing the workings of the ENVI-met model and several developments of the core model to increase its capability to describe vegetation in urban environments. In Simon (2016), chapter four evaluates ENVI-met ability to model heat stress using real world observations.

2.12.1   The Core model

ENVI-met is a three dimensional model that models heat transfer, airflow and solar radiation along with other features. As a dynamical model on a grid structure the model depends on cuboid grid cells that each have defined physical parameters, every time step the cells' variables (such as temperature) are updated on the basis of formulae that mimic physical processes such as heat conduction or convection.

The boundary model is one dimensional, equivalent to assuming uniform boundary conditions, representing a height profile of the different boundary conditions such as wind speed, air pressure, temperature and humidity. CFD models generally use a nesting area around the zone of interest to try and allow the 'edge effects' associated with uniform boundary conditions and prevent instability.

To ensure stability, the upper limit of the simulation generally extends up to 2500 meters (sometimes this is cited as at least 10 times vertical scale of the physical model. Cell resolution of around 1-5 meters means that buildings and vegetation are modelled directly, rather than being summarily parameterised over areas such as built land cover proportion or proportion of vegetated surface.

In terms of the technical details of how the model is set up:

- airflow is modelled using the non-hydrostatic Navier-Stokes equations and assuming incompressibility of flow.

- air temperature and humidity are modelled using advection diffusion equation with each cell having values for internal sources and sinks.

- turbulence is modelled to approximate the impact of these effects estimating the accumulation and dissipation of turbulent energy

- pollutant concentrations are modelled using advection diffusion equations with sources and sinks incorporating modelled chemical reactions.

- solar radiation is broadly handled by adjusting the intensity for optical depth and angle of incidence. Additional adjustments are made for leaf shading in terms of incoming and outgoing radiation.

- ground and building surfaces are modelled using calculated heat flux across the boundary and then treated as thermal capacitors and heat sources and sinks. For permeable soils moisture exchange is modelled as well.

- plants can be simulated in detail: different plant types characteristics such as paucity of light and water, photosynthesis rate, CO_2 concentrations and leaf area. As indicators of heat stress on plants which mean plants will reduce the amount they transpire which has a cooling effect on the local environment.

2.13 Case Study: Modelling transpiration and leaf temperature of urban trees and evaluation using measurement data

Simon (2016) used the ENVI-met model to carry out a study of the health of a large and small tree in a confined courtyard in the urban environment of Mainz, Germany. The purpose was to evaluate the accuracy of the ENVI-met simulation of tree health.

Plant health is viewed as particularly important in the context of planning to mitigate heat stress on the local area as plants such as trees provide evaporative cooling (increasing latent heat flux) reducing air temperatures and improving thermal comfort for the local population. Poor plant health compromises these functions leading to limited heat stress mitigation from evaporative cooling.

Plant health was monitored in two ways. The first was through sap movement as a measure of plant metabolic activity. The second was leaf temperatures measured damaging conditions for the plant. Temperatures of $40°C$ considered harmful to the plant.

The results 2.1 show that the majority of behaviour is captured in the model with the differences between cloud conditions and the small and large trees. The model does struggle with some smaller scale variations such as variation between days. One likely source of uncertainty is identified due to the weather station being over two kilo-meters away and therefore not reporting weather at the location of the trees.

In terms of uncertainty analysis the work evaluates the degree of agreement between the two data-sets using visual comparison of predictions with observed values and evaluating summary statistics the R squared values and the Normalised Root Mean Squared Errors (NRMSE) calculated for degrees of cloudiness and the different trees. These measures give non specific measures of general agreement. Both summarise squared differences hence produce nearly identical analysis.

Both the visual inspection of the plot in 2.1 and R^2 and NRMSE values suggest that the fully cloudy days modelled were less well predicted in terms of the simulated data predicting the sap flow rate and leaf temperatures. It was identified that this is due to others coming into play when the main factor solar radiative flux is not present.

In terms of appraising uncertainty Simon2016 evaluates uncertainty in chapter four by visually appraising predictive accuracy in the results from 2.1.

2.13.1 Conclusions

ENVI-met and similar models enable the modelling of very detailed situations such as this case the heat stress of plants in urban environments. These models are good at capturing the larger
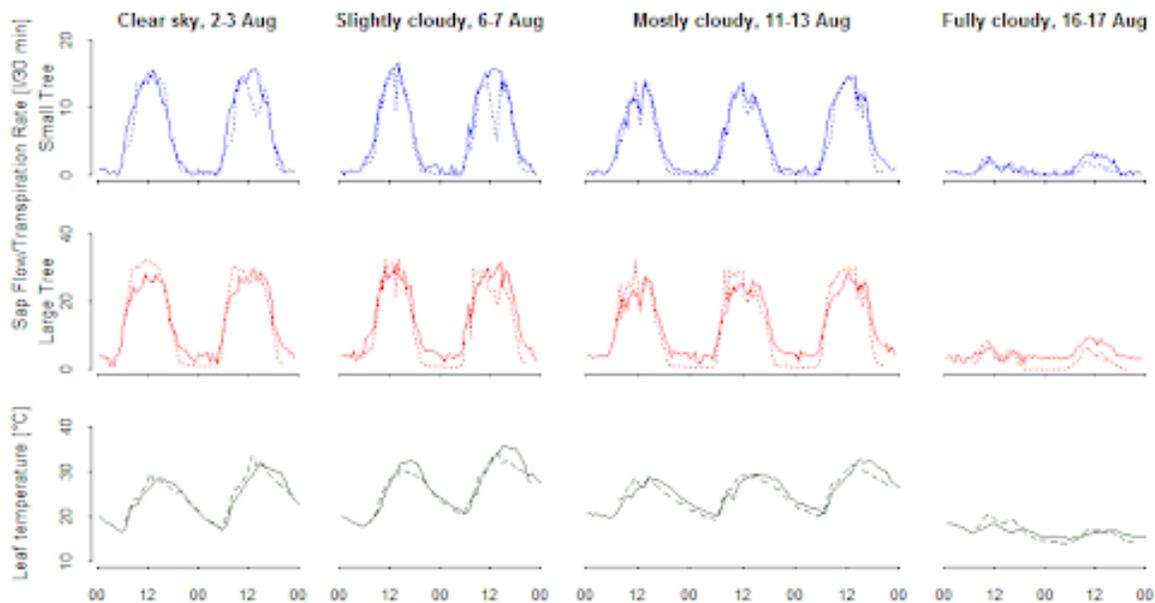
Figure 2.1: Results of ENVI-met simulation compared with measured values from chapter 4: Simon (2016)

behaviour of systems such as difference between different trees and broader diurnal patterns, however for reasons such as poor data or insufficient modelling they tend to not be able to provide high detail such as to within a degree Celsius prediction accuracy.

Urban weather can broadly be termed as meso-climate to micro-climate. Meso-climate effect for example might be the atmospheric drag provided by a city Kanda et al. (2013). Micro-climate is a term used to refer to effects such as airflow around a building or the shading of the urban canopy. Efforts to provide sample weather incorporating these effects for example the Urban Weather Generator (UWG) Bueno et al. (2013).

Chapter 3

Model selection for BEM in an uncertain climate

This chapter examines the problem of emulator model selection under uncertainty for a simple EnergyPlus model Energy (2010). This chapter makes use of a variety of measures to assess the success of three different models in emulating the behaviour of the building. This allows examination of some of the limitations of Gaussian processes (GPs) in real world situations and ways to carry out model selection.

This chapter does this by proposing a mixed expert model using a GP and two variants thereupon. ~~A method of~~ Maximum likelihood is used in evaluating model choice inherently taking into account the quantification of uncertainty is then evaluated to determine the best model. ~~I~~ It is conclude~~d~~ that maximum likelihood is a useful approach for model selection and results in the model best suited to the task of modelling energy use and highlight benefits of subdividing data into different behavioural patterns - in this case the heat systems being on and off due to cold and hot weather.

Keywords— Uncertainty Quantification - Building Energy Modelling - Gaussian Processes - Maximum Likelihood Estimation, MLE - Model Selection

3.1   Introduction

The emulator model selection in this chapter is for an EnergyPlus Energy (2010) model under climate variability. At present there are methods for using computer simulations (Dodoo & Gustavsson 2016) and methods for statistical analysis of empirical data (Christenson et al. 2006). In this chapter the methodology to select the emulator model of the building requires a sizeable data set to carry out cross-validation of the model-selection. As a result the model selection was evaluated on a simulated data-set.

For selection of the simulator, usability, fast computation and common practice were considered. One of the most widely-used ~~models~~ tools available is EnergyPlus (Energy 2010). EnergyPlus is used for a wide variety of purposes from agent based stock modelling Fcibse (2010) to integration with building design platforms such as IES Ltd. (n.d.). EnergyPlus has also been optimised to have good run times whilst retaining flexibility and accuracy Hong et al. (2008). DOE-2 Hirsch & Associates (2010) is the most comparable model to EnergyPlus. It has faster run times and in short is a slightly easier to use but less detailed model. A more detailed comparison can be found

in Rallapalli (2010), they are both commercial tools that differ by implementation. Agent based modelling is an alternative approach such as Zhao (2012)..

With climate change projections forecasting several degrees of global warming, cooling energy use is likely to be an increasing concern here in the UK. Uncertainty from a climate model is propagated through a simulation model such as Jentsch et al. (2015), in this case determining the weather years that are most extreme from the perspective of building performance. This propagation often uses Monte Carlo methods Ulam (1949) requiring many simulations for different weather files. Emulators are able to provide a partial solution to this problem in that they can provide faster estimates of the building behaviour than a simulator from past simulation runs or data.

The statistical process of emulation is when a computer simulator or real world response, f, has its behaviour modelled statistically producing an approximate model, $\tilde{f}$. This has been well established as a method (O'Hagan 2004, Bastos & O'Hagan 2009). In this case, the cooling energy use, y, is considered a response to the weather, x. As such $y = f(x)$ is the actual relationship whereas $y \approx \tilde{f}(x)$ describes replacing the simulator with the emulator, which is done in this chapter isbecause the emulator is much faster computationally. Generally $\tilde{f}$ will be a distribution of values due to the statistical nature of emulators. The benefits are cheap to evaluate estimates of the simulator's behaviour which can be useful for a number of reasons:

- Estimating behaviour of a computationally expensive simulator.

- Increasing the feasibility of other uncertainty quantification methods (such as parameter estimation and sensitivity analysis, discussed in O'Hagan & Oakley (2004)) which will typically require sampling the simulator 1000's of times and is often impractical. This will be particularly important here as the uncertainty estimates when applying the model in real world contexts such as chapter 4 does.

Here t The emulation is of the response of a building's interior environment to weather pattern changes , which when need to be passed on the BEM . Weather is highly varied. EnergyPlus uses .epw files that contain hour by hour weather data for 30 or so variables over the simulation period.

In this chapter, the statistical model has been set up to be comparable to real world data so a emulator can use the same structure in chapter 4 for investigating emulation of real world building performance, this is discussed in more detail in section 3.2.1. To this aim the inputs and outputs have been selected to have similar inputs and outputs to a proposed real world data set.

In this chapter, several emulators for modelling cooling energy consumption will be compared and the best will be chosen. The motivation for this is due to the unsuitability other energy use modelling in chapter 4, justified in section 4.2.1. Criteria that will be used to compare the model performance are likelihood, standardised residuals and visual inspection of results.

This chapter demonstrates how a simple adjustment to GP models can improve performance and how climate uncertainty can make drawing clear conclusions difficult.

3.2   Methodology

The methodology in this chapter lays out the sources of various data and how the models are constructed. The section starts discussing weather data and its source, followed by EnergyPlus, the raw distribution of the cooling energy data. Subdivision of the data into training and test data.

This section describes the elements of the statistical model and how this relates to uncertainty quantification.

3.2.1 Weather data

One of the key elements of uncertainty is for building energy use forecasts is the choice of weather data used. This chapter uses UKCP09 weather data (Murphy et al. 2010), comprised of the hourly time steps that are down-scaled Met-office forecasts to 5km grids for 30 year time series. These forecast consist of 10000 possible realisations for 30 year time periods. These 10000 samples are different represent different possibles climates drawn from the Met-office model.

This information is not in .epw format, the file format for EnergyPlus. There is sufficient information available to carry out the conversion. The Prometheus download files, (Eames 2016), are a set of files designed to summarise the UKCP09 files into 30 files for key locations across the UK, they are designed for energy use and overheating assessment, this makes them suitable for selecting representative files of wider weather patterns.

The Prometheus weather files pre-process the Met-office forecasts into a .epw format. The prometheus downloads provide 10, 33, 50, 66, 90 percentiles of the 10000 samples rated by estimated average building performance for heating and cooling, medium (a1b) and high (a1fi) climate scenarios for 2030s, 2050s and 2080s. This gives 30 weather files in .epw format for cooling. The 5km grid cell for London, Heathrow was used.

The weather in western Europe is strongly seasonal making the change in behaviour seen in different seasons apparent. Some different challenges may be observed when dealing with data from other countries closer to the equator.

Data available (from EnergyPlus simulation) for this study is 30 years with energy consumption for heating and cooling daily against weather values every hour. A common reporting method for real world weather station data is to report maximum and minimum, wet and dry bulb daily temperatures with average direct solar radiation, as done with the Birmingham Climate Laboratory data-set used in chapters 5 & 6.

The Weston Park weather station in Sheffield produces records weather data on daily intervals which motivates this study's choice to work with daily summaries of the .epw files rather than the much shorter intervals usually provided, this is taken advantage of in the next chapter 4. This data can be directly constructed from the .epw. Combined with the previous day's weather values we now have for every day in our 30 different years:

- Inputs:

    v1 Max current day dry bulb temperature ($°C$)

    v2 Max current day wet bulb temperature ($°C$)

    v3 Min current day dry bulb temperature ($°C$)

    v4 Min current day wet bulb temperature ($°C$)

    v5 Max previous day dry bulb temperature ($°C$)

    v6 Max previous day wet bulb temperature ($°C$)

    v7 Min previous day dry bulb temperature ($°C$)

    v8 Min previous day wet bulb temperature ($°C$)

    v9 Average direct normal radiation (W/m^2)

   v10 Day of the year

- Output:

   v11 Cooling Energy Used (J)

As discussed earlier the model is set up to allow for use of real world weather data. In this case chapter 4 using data from Weston Park weather station in Sheffield, to carry out similar modelling.

The combination of wet and dry bulb air temperature ensures that any humidity impact is included alongside temperature.

3.2.1.1   Rational for choosing previous days energy use

The previous day's temperatures permit some memory in the system (if the weather is hot all day for two days in a row presumably cooling is more of a problem). Figure 3.1 shows the correlation between different weather variables and the cooling energy. The strongest value is exhibited in the previous day. In order to avoid feeding variables that are strongly auto correlated it is advisable to look for each strong peak in the correlations.

After the first day most variables seem to have a peak or turning point in correlation at 8 to 9 days lag. However the strength of this peak is either a barely more than a turning point or in the case of $v9$, direct solar radiation, has a peak at less than 0.4 correlation suggesting limited information to exploit. In summary, this chapter chooses to use the highest correlated previous day data, however a more resource intensive approach might be interesting to investigate the value of the weather data at 8-9 days previous. This could represent resonance in the system of this time scale which would represent a response time of the building to external weather effects.

3.2.2   EnergyPlus simulations

The simulations carried out were of cooling energy consumption of an office block. This work has used the standard multi-story office block .idf (building simulation file) provided as part of the EnergyPlus package and weather files from the Prometheus (Eames et al. 2011) download files.

The Prometheus weather files are specified at $10, 33, 50, 66, 90$ percentiles, medium (ab1) and high (a1fi) climate scenarios for 2030s, 2050s and 2080s as mentioned earlier give 30 weather files in the .epw format. These 30 files can provide our cross validation set as discussed below in section 3.2.4.

Cooling energy use was set up to be recorded on daily intervals. The time step for simulation was set at 10 minute intervals. The building was modelled in 9 zone types each producing predicted heating and cooling energy use along with temperature values. As the model selection criteria, in this chapter was the focus of the results not the building model, a decision was made to select a single moderately complex building model. The set point temperatures were set up to be $20.0°C$ for heating and $24°C$ for cooling operational throughout the simulation period. These set point temperatures are broadly in line with Bonfigli et al. (2017), Humphreys & Nicol (2015).

The zones were top, middle and ground floors and east, centre and west side of the building. For this exercise I have taken the 9 zones added together to give total energy used for cooling. This building set up was then run, for all weather files, to give the output data sets.

Fig. 3.2 provides a visual representation of the EnergyPlus model. In a computer model textcolored, the discretisations of the laws of physics can obviously be adjusted for computational convenience. In this case it is done to avoid creating 20 or more zones that are nearly identical and

Figure 3.1: Correlation of previous days weather to cooling energy use. (v1-10 3.2.1: v1 - Max cur. dry bulb temp., v2 - Max cur. wet bulb temp., v3 - Min cur. dry bulb temp., v4 - Min cur. wet bulb temp., v5 - Max prev. dry bulb temp., v6 - Max prev. wet bulb temp., v7 - Min prev. dry bulb temp., v8 - Min prev. wet bulb temp., v9 Aver. direct normal radiation, v10 - Day of the year)

Figure 3.2: Visualisation of the multistory EnergyPlus model from AutoCAD, the model is a standard example from the EnergyPlus download Energy (2010) . The image is taken directly from AutoCAD, the different sketches are of different zones within the BEM model. The model was selected prior to knowledge of the availability of the Information Commons used in chapter 4.
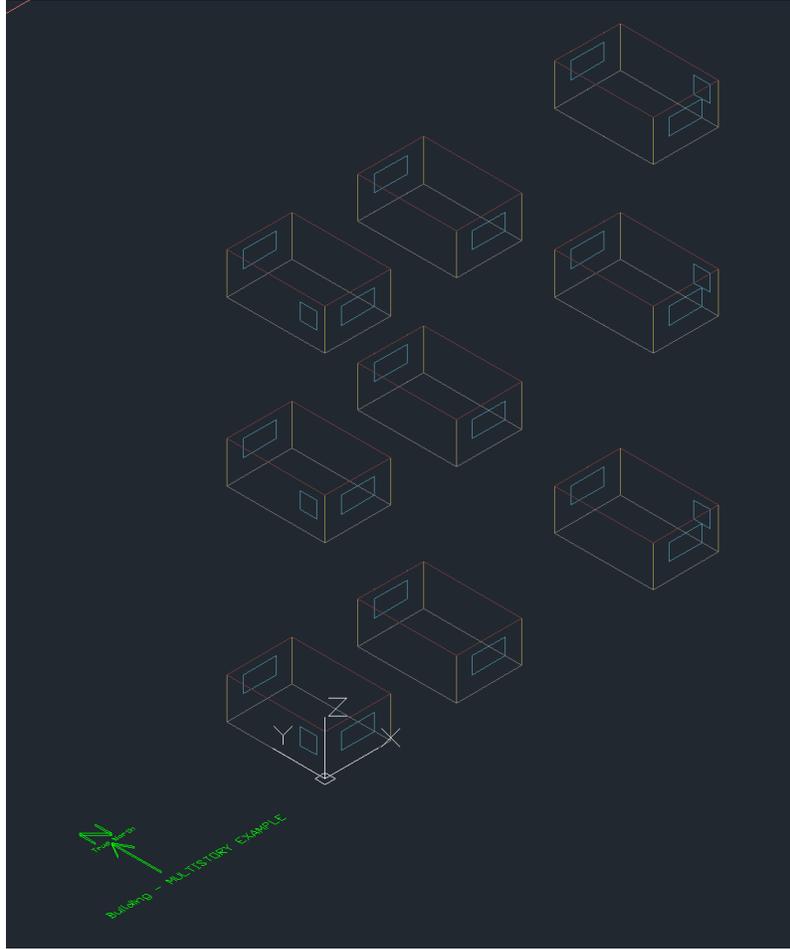
have very similar (possibly in some cases to near computational accuracy) results. This greatly speeds up model computation time and reduces memory usage, in the example above the run time comparison approximately be in order of 4 or more times as intensive due the increased number of connections between zones.

The middle floor is essentially designed to represent the many different levels in a multistory building symbolically. It assumes they are very similar so have the same inputs at the top and the bottom for each layer. In order to represent this in a symbolicfinite difference model, the heat exchange of the top of the middle zones is linked intoresolved simultaneously with the heat exchange at the bottom of the same zone. Once this is resolved, the heat flux is then linked to the bottom of the 'top' floor zones to provide the amount of heat gained. The same grouping is applied with the bottom of the middle floor zones and the ground floor zones. This means the middle zones are unaffected by the top and bottom floors, however the top and ground floors are linked to the middle zone, hence the AutoCAD interpreting the zones as physically dissociated, fig. 3.2.

3.2.3 Data distribution

When proposing models there are some quick checks that can be carried out to check the assumptions in the proposed models are reasonable. First I will visually inspect the data to check for linear and non linear behaviour between the input and output, then I will discuss the degree of non-stationarity. Finally I will consider if there are any particular values that are unusually common in the output data.

3.2.3.1 Linear behaviour

In general much statistical analysis relates to evaluating linear correlations. In statistical analysis linear correlations relate to the output Y having a relationship to a variable x of the form $y = f(x)$ where $f(x) = ax + b$. That is not to say there aren't non-linear dependencies on other variables, only that these variable are linear in respect to x.

The data here will need to have some relationship between the input and the output data. In this section I visually inspect the input and the outputs, particularly looking for any linear patterns. The GP used in this chapter has a linear-regression mean function allowing it to model this simple behaviour easily. In turn this allows the main process to account for more non-linear behaviour. The linear mean help the GP model account for limiting behaviour at the edge of the parameter space: for example as mean temperature increases outside of data points so does a GP with a fitted linear mean.

EnergyPlus, when running a cooling system to a set point temperature, will use any amount of energy required to maintain the temperature below the set point temperature. As the temperature of a gas increases the equivalent kinetic energy increases linearly within it, therefore the amount of energy (as simulated by EnergyPlus) required to cool an environment does as well.

The link between variables is assumed to take the form:

$$v11 = f(v1, ..., v10) \tag{3.1}$$

Figure 3.3 shows there is broadly an increase in cooling energy use linearly with inputs (see section 3.2.1), suggesting that a linear mean function is appropriate, when modelling f. Variables v9 (direct solar radiation) and v10 (day of the year) demonstrate different behaviour.

It should also be noted that real buildings are unlikely to have such a linear relationship due to variable cooling system efficiency ~~(most systems have a peak efficiency above which they become less efficient or hit peak capacity)~~. The behaviour of chilling energy use at zero, precisely zero due simulation not having a non zero inactive state, will be discussed later (see section 3.2.5).

The average direct normal radiation (v9), in relation to the cooling energy (v11), is less clearly linear. It also shows signs of striation (where the values take fixed values so form hozontal bands) suggesting that a rounding process was involved in the reported selection of these values. Given the general linear trend, it was decided to leave v9 in unaltered as a possible predictor.

The day of the year (v10), in relation to cooling energy (v11), demonstrates periodic behaviour. The Gaussian process should easily capture this behaviour, however other methods to handle this behaviour include fitting period function such as sine curves as linear models (models of the form $y = a\sin(x) + b$).

It should be noted, that away from their data points GPs are dominated by their mean functions in this case this linear model.
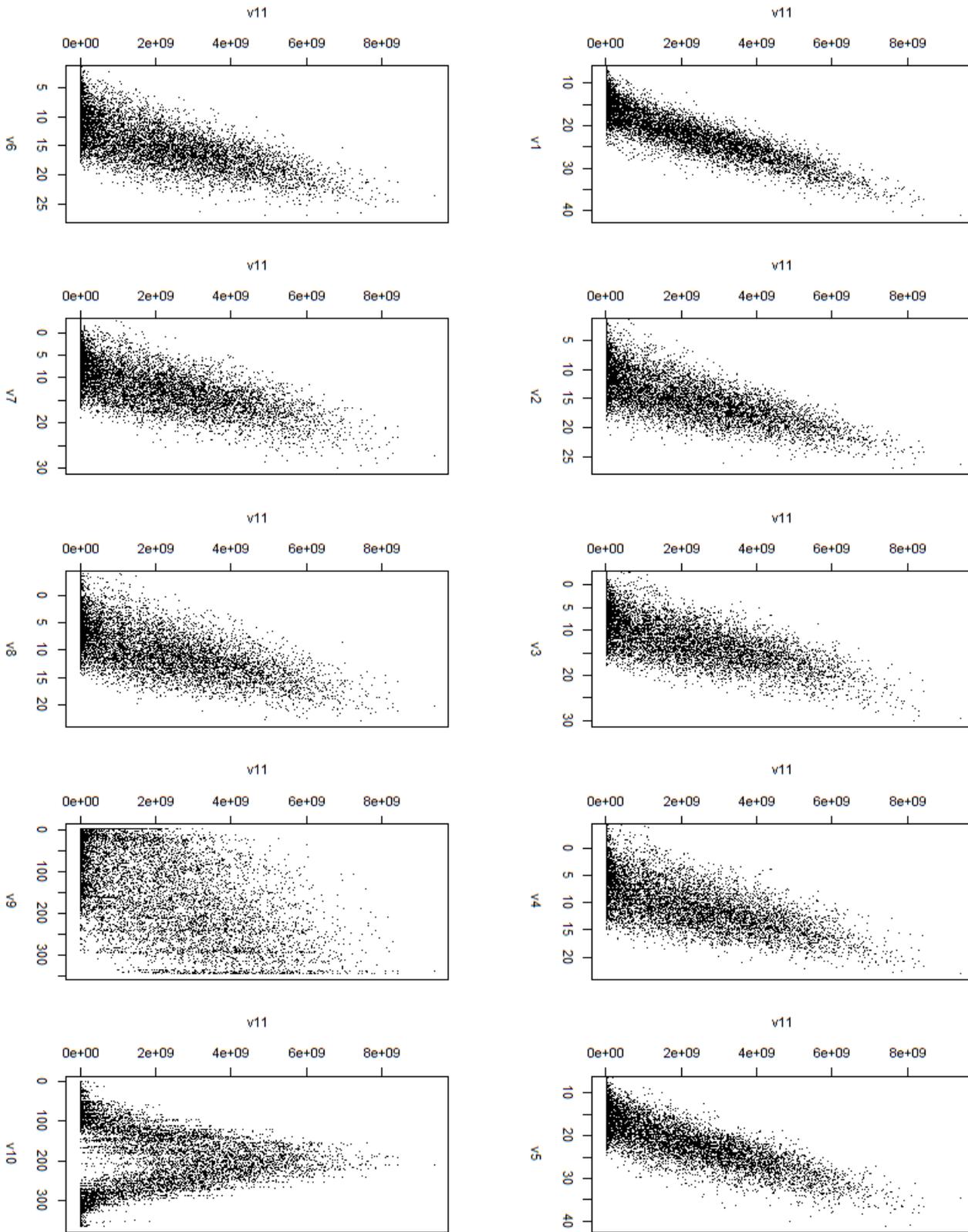
Figure 3.3: Plot of simulated cooling energy (v11 in J, Joules) against input variables (v1-10 3.2.1: v1 - Max cur. dry bulb temp., v2 - Max cur. wet bulb temp., v3 - Min cur. dry bulb temp., v4 - Min cur. wet bulb temp., v5 - Max prev. dry bulb temp., v6 - Max prev. wet bulb temp., v7 - Min prev. dry bulb temp., v8 - Min prev. wet bulb temp., v9 Aver. direct normal radiation, v10 - Day of the year). It should also be noted that there are tails of values that are clipped off the graphs at zero.

3.2.3.2   Non-stationarity

GPs are most commonly formulated with stationary covariance functions or kernels. Stationarity is defined as the variation two point in the function being solely dependant on distance between two points. More formally:

$$Var(x, x') = f(|x - x'|) \tag{3.2}$$

where x, x' are members of a complete metric space. More informally this mean that the function's variation is same in different places comparing the variation of output to input location. Non stationarity is the converse where variation is dependant on the values of x, x'.

There is indeed an issue with non-stationarity here. In figure 3.3 the large number of values at zero, compared to the variation in the non zero v11 values, is problematic. In this sense v11 has inflated zero values, however these zeros actually represent the off state of the chiller so are realistic if high problem for many statistical models. The GP used in this work is assumed to have no nugget term due to the deterministic nature of the simulator.

There have been several different attempts to manage non-statonarity for GPs. The most obvious in some cases is to subdivide the input space into partitions which contain different types of behaviour Chipman et al. (1998). This works relatively well for problems where the input space has relatively low number of dimension on which to carry out the partition. ~~Gramacy et al. 2008, outlining TREED GPs,~~Gramacy & Lee (2008) partitions the input space and fitting different GP on the partitions, separately is know as a TREED Gaussian Process. This approach however is more challenging when the partition of the input space is difficult to get right.

Another approach, more theoretical than the approach adopted here, is laid out far in Volodina & Williamson (2020). In Volodina & Williamson (2020) a new method is proposed for creating non-stationary GPs with full priors for the hyper-parameters in the models. In Volodina & Williamson (2020) they employ a mixture of Gaussian processes that smoothly transition between Gaussian processes with different hyper-parameters. The mixing and classification are fitted simultaneously by maximum likelihood. The mixture of experts models (1. and 3.) presented below is simpler than Volodina & Williamson (2020) in that it depends on supervised classification of value $\sin(\frac{1}{x})$'s the input space. The method presented below is also more friendly to higher sample sizes due the data sub-division in the method 1. in this chapter. This is discussed in more detail in subsection 3.2.6.

3.2.3.3   Spacing of output data

With data that has such a large variation in output values from 0 to 10^9, it would be normal to consider transforming the data to ensure the data is well spaced out over the output space. This has the advantage of reducing numerical problems in calculations.

In figure 3.4a, histograms of values and transformed values, it is clear that the data is fairly evenly spaced for nonzero values. A log transformation (all values were shifted $+10^{-15}$ before transformation, figure 3.4 (b)) exposes the fact that near zero values are precisely zero (as such the transformation returns them as -15). The log transformation clarifies there is in-fact two apparently distinct behaviours going on: values at zero and those above zero.

In short, transforming this data ~~doesn't~~ not have any apparent advantages that would assist the models selected here.

Figure 3.4: Histograms of frequency of simulated cooling energy use values.

(a)   (b)
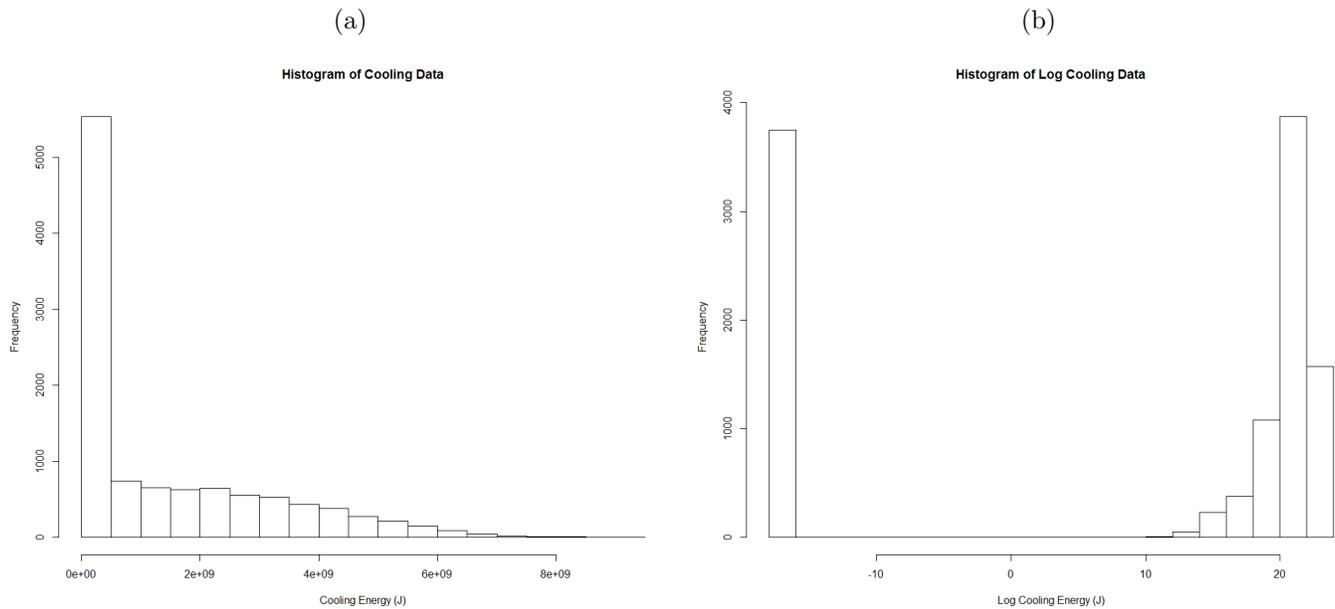
3.2.4   Cross validation

Cross validation is the process of sub-dividing the data-set for a problem into training and test sets multiple times in order to avoid over fitting of a model to a particular set of circumstances. In this ~~case I aim~~chapter, it is aims to ensure the robustness of the model choice to different weather years, given the variation of weather in years.

Of the 30 available weather years, as data points, on which to build an emulator this gives ~~us~~ around 30×365 data points (11000). GPs require a matrix inversion and storage of an approximately 11000×11000 matrix to use all available data textcolorredpoints, simultaneously. This is not very feasible and does not provide a test set on which to evaluate the accuracy of our statistical model.

Each of the 30 years provides a suitable subdivision for creating validation sets, given the winter/summer cycle yea~~d~~rs are are the obvious way of subdividing the data. This reinforces in fig. 3.3 day of the year v10 against the cooling energy showing the annual summer peak in cooling energy. In this chapter pairings of years are used to form training sets. Pairs were chosen as when considering the likely available real-world data for which there is unlikely to be much more than a few years worth of data from which to carry out modelling.

Pairings of years were picked as to also allow the study in this chapter to provide an estimate of the accuracy of the results in chapter 4. Providing the model in this chapter the same duration of information that was available in chapter 4.

A smaller number of years also makes the computation task much more tractable, whilst providing 28 test sets for the model, 30 minus the current training pair. ~~As such~~ For clarity: the model is trained on a pair of years then validated against each of the 28 remaining years. In ~~total~~summary, using each possible training set ($30 \times 29/2$, 425) and then predict the performance for each of the 28 remaining years giving 12180 unique validation results summarised in the results

section 3.3 fig. 3.5.

3.2.5   Dual behaviour GP

The data appears to exhibit two distinct behaviours:

- Low temperatures, unsurprisingly, lead to zero cooling use.

- Non zero energy use is linked to a warmer environment.

The overall trend is expected to be linear as discussed in 3.2.3.1. There is no guarantee local behaviour will be so. The difficulty of dealing with the different possible interactions lends the data to using a GP. Using a GP will also facilitate the quantification of uncertainty (or lack of knowledge) at points in the design space.

The two behaviours will cause problems for GPs. The GP is formulated with a stationary kernel as discussed earlier, so estimation of scale and variance parameters could be affected as will the mean function dragged down towards the zero values. This model was inspired by approaches such as Gramacy & Lee (2008), however the conception here is not known to be used elsewhere.

Three models are proposed, where Y^* is a vector of cooling energy on particular days. X_{tr} the inputs or regression variables. For different models I divide the data into sub data-sets for use in different models for use together to better inform the parts of the model. The first data-set is the observation where the Y^* values are zero associated with their X_{tr} values. The second being the data points with non-zero Y^* values. So the data-sets are denoted:

1. Zero values of Y^*:
$$(X_{tr}|Y_{tr} = 0) \tag{3.3}$$

2. Non zero Y^* values
$$(X_{tr}||Y_{tr} > 0) \tag{3.4}$$

3. The undivided data-set denoted:
$$(X_{tr}) \tag{3.5}$$

which can be used used to conditions posteriors such as in 3.6, 3.7 and 3.8.

The details of the Bernoulli switch implementation are discussed further below see equations 3.9, 3.10. Models are as follows:

1. GP with Bernoulli switch and data > 0 (using the standard logit link function, g and linear mean). Initially a binomial model is created to classify the value as zero or to use a GP. The GP is trained on data. The data points passed to the GP are only those that are non-zero energy consumption. If Y_{tr} is the training values of energy consumption matching up with X_{tr} then the following structure is used:

$$
\begin{aligned}
Z|X^*, X_{tr} &\sim Bernoulli(g(X^*)) \\
Y^*|X^*, X_{tr}, Z = 0 &\sim \delta_0(X^*) \\
Y^*|X^*, X_{tr}, Z = 1 &\sim \mathcal{GP}(\mu(X^*), \\
&\quad k(X^*, X^*)|(X_{tr}|Y_{tr} > 0))
\end{aligned} \tag{3.6}
$$

where Z is the Bernoulli random variable for the switch, X_{tr} is the two years training points, g is the link function, $\delta_0(\cdot)$ is the Dirac delta function with a point mass of 1 at zero treated as a probability distribution, X^* the new point to evaluate the posterior at, μ is the posterior mean function and k is the posterior covariance function.

2. Simple GP. Numerically optimised maximum likelihood spread and variance parameters and a maximum likelihood selected mean function. The GP is trained on two full years worth of weather against energy use data. The following structure is used:

$$Y^*|X^*, X_{tr} \sim \mathcal{GP}(\mu(X^*), k(X^*, X^*)|X_{tr}) \tag{3.7}$$

where X_{tr} is the two years training points, X^* the new point to evaluate the posterior at, μ is the posterior mean function and k is the posterior covariance function.

3. GP with Bernoulli switch and all data. This model is similar to model 1 except for the GP using all the data. This leads to the following structure:

$$\begin{aligned}
Z|X^*, X_{tr} &\sim Bernoulli(g(X^*)) \\
Y^*|X^*, X_{tr}, Z = 0 &\sim \delta_0(X^*) \\
Y^*|X^*, X_{tr}, Z = 1 &\sim \mathcal{GP}(\mu(X^*), k(X^*, X^*)|X_{tr})
\end{aligned} \tag{3.8}$$

where Z is the Bernoulli random variable for the switch, X_{tr} is the two years training points, g is the link function, $\delta_0(\cdot)$ is the Dirac delta function with a point mass of 1 at zero treated as a probability distribution, X^* the new point to evaluate the posterior at, μ is the posterior mean function and k is the posterior covariance function.

The proposed models 1. and 3. are mixture models. These models belong to a class of models developed initially developed in the early 2000's called Mixture of Expert models (Tresp 2000). The model is hierarchical, it is composed of a switch that turns a GP on and off, allowing for the energy consumption to be exactly 0 kWh.

The models with the Bernoulli switch is fitted into 2 stages:

1. Fit Bernoulli switch. The data is classified into $(X_{tr}|Y_{tr} = 0)$ and $(X_{tr}|Y_{tr} > 0)$ groups. Then the linear classifier is trained on those classifications.

2. Fit GP on the data-set either (X_{tr}) or $(X_{tr}|Y_{tr} > 0)$ depending on if it is model 1. or 3. .

The switch is a Bernoulli random variable with the canonical logit link function (see g below). The model is fitted as a generalised linear model (GLM) using maximum likelihood in the binomial family. To coerce the data into a Bernoulli format, the data was classified to be 0 represented $e = 0$ or $e > 0$ represented 1 which turns on the GP. To summarise:

$$Y = \begin{cases} \mathcal{GP}\left(\mu(X^*), k(X^*, X'^*)\right) & \text{probability} \quad p \\ 0 & \text{probability} \quad 1 - p \end{cases} \tag{3.9}$$

Where $\mathcal{GP}$ represents the GP, it should be noted that the top half is not quite non-zero as there is some density for a GP at zero and p is a function of x^*) via the logit link function to a linear predictor as follows:

$$p = \frac{e^{\beta \cdot X}}{e^{\beta \cdot X} + 1}(= g(X)) \tag{3.10}$$

Where $\beta \cdot X$ is the linear predictor for the GLM, X is the design data, Y the response, cooling energy used and g the link function. When $Y = 0$ it represents the chilling being off and therefore not consuming energy. The model distinguishes between the states on and off first rather directly estimating consumption. Distinguishing between the choice of model using likelihood will determine the effectiveness on this approach comparing model 2 without a switch against models 1 and 3.

The GPs are all fitted using variance parameters and mean function linear estimated using a numerical maximum likelihood method. A Matérn $\frac{5}{2}$ covariance function was used. The GP was evaluated using the DiceKriging package in R. No nugget was used in the GP.

This method has been adopted on the assumption that there are two distinct behaviours. Cooling periods have periods where energy use is zero. This corresponds to temperature being too cool to require cooling (sub-zero temperatures outside requires some unusual circumstances for residents to want cooling further) and the inverse for high temperatures leads to zero demand for heating.

The posterior of the GP is formulated as follows for predictions (Rasmussen et al. 2006):

$$Y^*|X^*, X_{tr}, Y_{tr} \sim \mathcal{N}(\mu(X^*), \sigma(X^*))$$

$$\begin{aligned}
\mu(X^*) =&\, k(X^*, X_{tr}) k(X_{tr}, X_{tr})^{-1} Y_{tr} \\
\sigma(X^*) =&\, k(X^*, X^*) \\
&- k(X^*, X_{tr}) k(X_{tr}, X_{tr})^{-1} k(X_{tr}, X^*)
\end{aligned} \tag{3.11}$$

where Y^* is the predicted values for energy cooling based on X^*, X_{tr} is the two years training points as inputs, X^* the new inputs to evaluate the posterior at, Y_{tr} is the two years of cooling outputs for training, $K(A, B)$ is the matrix of all possible values of the covariance function $k(\cdot, \cdot)$ (in this case the Matérn covariance function) between the sets of vectors $A = \{a_1, ..., a_n\}$ and $B = \{b_1, ..., b_n\}$.

3.2.6 Sparse GPs

For using the GP with n data points, the training is a $O(n^3)$ operation and storing requires $O(n^2)$ memory. This means for large data sets GPs become much less practical. Whilst sparse GPs haven't been used in this thesis here, if all the data were to be used simultaneously as discussed in section 3.2.4, simple GPs would be computationally challenging. This chapter does not make use of sparse GPs, however they be highly relevant to future research on the topic of GP working on subsets of data as they both solve problems inherent to 'big' data: namely diversity of behaviour, sub-setted GPs, and volume of data, sparse GPs.

The core principle is to use representative data points rather than the full data set. This means the GP is only as costly as the size of the representative set, opposed to using all the data. Methods such as in Snelson & Ghahramani (2006) use pseudo inputs to approximate the data, this sort of process of choosing such inputs is based on approximate maximum likelihood.

Sparse GPs offer a way of including large amounts of generated data into our emulator however are significantly more difficult to implement due being more temperamental when fitting hyper-

Figure 3.5: Plot of standardised residuals for estimated cooling energy use annually across all possible combinations of training sets and prediction years.
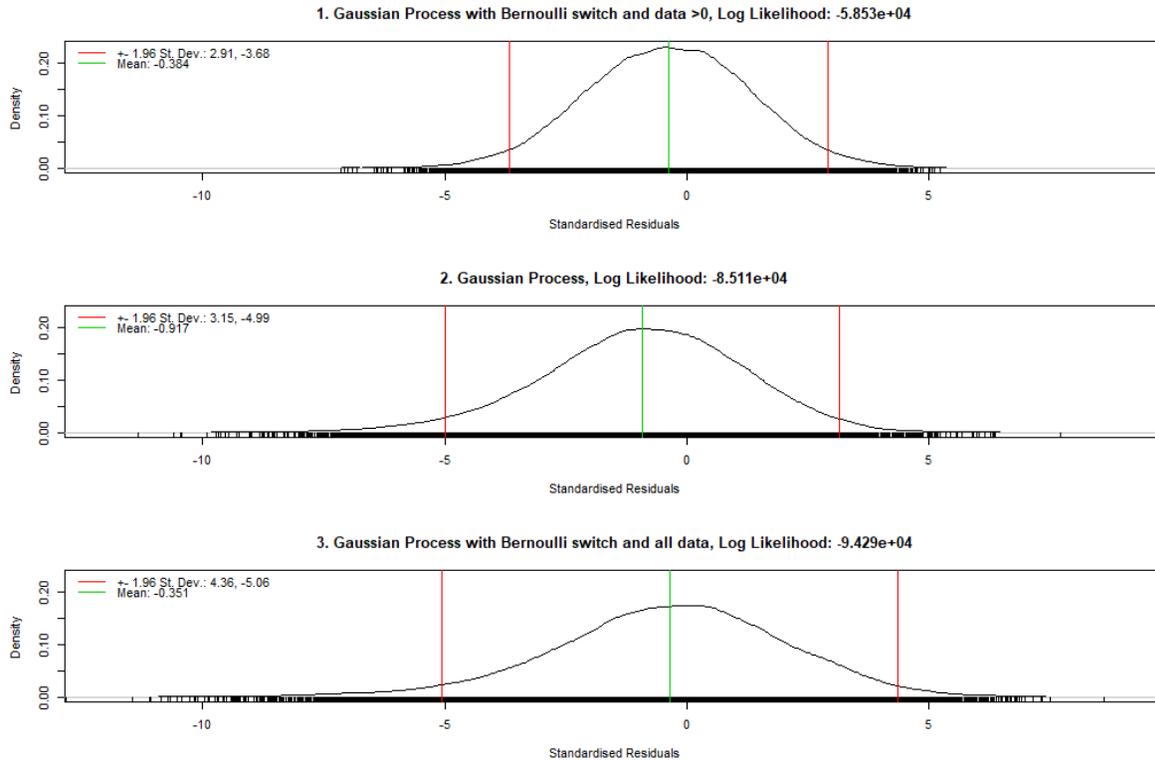
parameters.. The emulator here allows prediction of a single day without running an entire year's simulation.

3.2.7   Verification of uncertainty quantification

One of the goals is to evaluate how successful this methodology is at modelling the cooling energy data.

Ideally the predictions should give precisely correct values for energy consumption. However if the actual output is dependent on more information than is available to the model there will be unexplained variation. This is termed residual uncertainty in O'Hagan & Oakley (2004). The model cannot distinguish between the inputs but there is a discrepancy between outputs.

Stochastic models address this concern by having a range of values for each output and a relative weighting on their likelihood or more specifically a probability distribution.

These distributions are then used to make inference on the outputs. For example in (Andrianakis et al. 2015) inference is created on the input parameters of a HIV spread model. The inference here is much simpler to describe the behaviour of energy consumption under selected climatic conditions.

Figure 3.5 shows the standardised residuals estimated by each model. To estimate the distribution of annual cooling use the following method was used:

1. Generate each possible combination pairing, as discussed in subsection 3.2.4, of training years.

2. For each training set:

   a) Train model

   b) For each year not in training set:

      i. Sample 2000 realisations of the model output for the whole year (daily predictions). The 2000 samples was found sufficient to make the central limit theorem applicable.

      ii. For each realisation sum energy consumption over the year (annual prediction).

      iii. Calculate the mean and standard deviation of the 2000 annual realisations.

3. Evaluate the standard residuals for each training group's predictions:

$$Residual = \frac{Actual\ Usage - Mean\ Usage}{Standard\ Deviation} \tag{3.12}$$

It should be noted that the actual data here ~~the data~~ is from EnergyPlus. In total there are $12,180$ combinations ($training(30 \times 29/2) \times predictions(28)$). Residual were used of the mean squared error in order to qualify error in the confidence of the model.

The distribution of standardised residuals to be expected is not immediately obvious. Here it suggested that the distribution should be $\mathcal{N}(0, 1^2)$. Essentially there are 2 data sets with different associated certainty. The training set and the larger set. The goal is to use the training set to estimate behaviour in the larger data set.

This represents the real world scenario that there are two years of data and comparing predicted forecasts with then measured energy use. The modelled data should provide a mean of the realisations that is near the actual value. If the mean is less than the actual value energy use is underestimated, if above over estimated.

The variation of the realisation around the mean provides information about the uncertainty. It is summarised here by using the standard deviation. This summary is reasonable if it is assumed the data is mono-modal. It is arguable that the stronger statement of normality via the central limit theorem is not inappropriate.

Whilst the central limit theorem ~~is not directly applicable~~ conditions are poorly met, due to covariance structures and small number of items in the summation, normality could be assumed given that the models 1. to 3. all contain GPs (and hence are normal multivariate distributions, which sum to normal distributions) and that 365 is not small.

Model 2. has a normal posterior that could be directly calculated, however for consistency has been treated the same as the other two models. Here there are two data sets with different associated noise. The data produced from the forecast from our emulator are noisy. This noise is represented here by the standardised residuals.

They are the difference between two data sets both assumed approximately Gaussian and should ideally be equal. Ideally there would be zero difference between the two distributions or zero residuals. Distinguishing between the models requires some measure of model fit. Using the ideal distribution as a likelihood will enable creation of a posterior of model preference.

Here the prior is that the models have an equal probability as being most suitable. The question is how to update this prior.

The log likelihood can be considered to be:

$$\log\left(\mathcal{L}(\mathcal{M}_i|R)\right) = \Sigma_{R_i} \, \log\left(p\left(r_{\mathcal{M}_i}|\mathcal{N}(0,1^2)\right)\right) \tag{3.13}$$

where $\mathcal{L}$ is the likelihood function, $\mathcal{M}_i$ is the statistical model (models 1. to 3.), R is the set of residuals derived from all the models, R_i is the set of residuals for $\mathcal{M}_i$, $p(\cdot)$ is the probability density function, $r_{\mathcal{M}_i}$ are the residuals derived from the model $\mathcal{M}_i$.

This likelihood then allows consideration of a posterior weighted by the relative likelihoods. In the following form:

$$\mathcal{P}(\mathcal{M}_i) = \frac{\mathcal{L}(\mathcal{M}_i)}{\Sigma_{j\in I} \, \mathcal{L}(\mathcal{M}_j)} \tag{3.14}$$

where $\mathcal{P}(\cdot)$ is the posterior, I is the set of models.

3.3   Results

The results presented here are from evaluating the results of emulating the simulated data from a third weather year after having trained the emulator on two uniformly randomly selected others, in order to cross validate the model.

There are three main metrics used to evaluate the choice of model selection. The first is maximum likelihood which is a single value metric that summarises trade off of uncertainty against mean precision though this trade does rarely have some unusual and daft results. The second is the distribution of residuals from the predictions from the cross validation results. The final measure is a visual inspection anecdotally of two cross validation runs to look for any structural errors. The visual inspection does not supersede the numerical evaluations, it does however provide identification of anomalous behaviour and is important for assessing model fitness in section 3.6.

3.3.1   Evaluation of model choice using likelihood and standardised residuals

The results for the model choice can be most easily made using the likelihood based inference. In this case the posterior likelihood is clear. This can then be evaluated in the context of the residuals to evaluate whether the strength of the recommendation is appropriate.

The models log likelihood (figure 3.5) make it apparent that (to 3 s.f. in the log likelihood):

$$\begin{aligned}
\mathcal{P}(\mathcal{M}_{1.}) &= \frac{e^{-5.853e4}}{e^{-5.853e4} + e^{-8.511e4} + e^{-9.429e4}} \\
&= \frac{1}{1 + e^{-2.658e4} + e^{-3.575e4}} \approx 1
\end{aligned} \tag{3.15}$$

which clearly recommends model 1. as the most suitable model, the posterior (it recommends model 1. with a probability of 1 to machine precision).

Figure 3.5 shows the spread of residuals in each model. The bell shape of each set of residuals density plot suggests that the normal likelihood assumption, implicit in using a GP, was reasonable.

All three densities have standardised residuals that have disproportionate number of values for the $\mathcal{N}(0,1^2)$ distribution. Their occurrence suggests that the models are overconfident and underestimate the probability of extreme occurrences.

The leading factor in model 1. having higher likelihood is likely the lower frequency of such extreme standard residuals. The mean of the density is closest to 0 for model 3. followed closely by 1. . All are negative suggesting general underestimation of energy use.

The fact the mean of the residuals for all three models is within the body of the bell curve for the other two models suggests there is substantial similarity to suggest the models do not have much to recommend over one another in terms of mean prediction, however the 95% confidence intervals (based on 1.96 standard variations) of the residuals are substantially different suggesting this remains a distinguishing feature.

Model 1. has the lowest residuals closest to a $\mathcal{N}(0, 1^2)$ distribution in terms of having the lowest residuals variances.

3.3.2 Prediction of Daily Energy Use: Visual assessment of anecdotal years

Figures 3.6, 3.7 and 3.8 show for one training set (years 1 and 2, 2030 10% and 33% percentiles using medium projection) predicting three years (4, 2030 80% percentile medium; 17, 2050 33% percentile high; 28, 2080 50% percentile high; respectively).

3.3.2.1 Winter performance

The effect of the Bernoulli switch can be seen in both 1. and 3. in all of the figures 3.6, 3.7 & 3.8. The start and end of the year being the winter in the Northern hemisphere has zero cooling energy use in places. Model 2., the simple GP, can't cope with this hetro-stochastic behaviour suggesting uncertainty here.

It should also be noted that in places all three models can stray into having negative means in the winter. This is most pronounced in model 2. without a switch to suppress the non zero values.

3.3.2.2 Summer performance

It appears that in the middle of the year (Northern hemisphere summer) figure 3.6 that model 2. actually performs best. The mean is nearest to the actual values and uncertainty is suitably compact.

However it can be seen that in figure 3.8 the models have a tendency to underestimate the summer cooling energy use. The larger uncertainty in model 1. allows for this behavioural uncertainty, essentially as the model doesn't know the energy consumption precisely its uncertainty tries to account for that.

3.4 Discussion

The comparison of the modelling suggests that a large part of the full behaviour from the full EnergyPlus model can be captured with two years worth of weather/energy use data. This chapter strongly supports the idea of maximum likelihood being a good method for choosing between models, even if gives apparently overly confident results.

There still seem to be challenges as all models exhibit overconfidence in their predictions and have lower mean predicted energy consumption than the actual data.

3.4.1   Model 1. GP with Bernoulli switch and data > 0

Model 1. has been assessed to have the best performance in terms of having been heavily preferred by the posterior distribution. The mean is similarly close to zero as model 3. though the variance is better quantified accounting for the posterior choice (see 3.3.1).

It also has good winter performance due to the binomial switch though it does still occasionally have a negative mean energy use, which clearly has no basis in the simulation in EnergyPlus (cooling systems don't produce energy when it's cold). Generally model 1. is least certain in the summer. This is due to the dominance of the GP here which has different data than the GP in the other two models. It appears the zeros in the data affect the GPs spread and variance parameter estimation.

Model 1. also has the fastest training time and lowest memory usage of the models, this is due to the segregation on the near zero data from the non zero data. Whilst the saving is very modest as not much more than 30% of the points are non zero. However this speed up is much more than the time spent training the binomial model.

3.4.2   Model 2. Simple GP

The raw GP notably has non estimates for energy consumption which is problematic, there also seems to be challenges with estimating variance and spread parameters.

It should be noted the performance is still not poor. The very simplistic linear prior could possibly be improved which would improve the performance of all three models.

3.4.3   Model 3. GP with Bernoulli switch and all data

This model performs the poorest of all the models, despite having the closest estimate of the mean to the ideal value of 0. The likelihood function penalises issues with the variance very heavily which given values out at $+8$ and -11 standard deviations causes considerable problems.

3.4.4   Likelihood as a decision function

In this work the likelihood function has behaved as a effective discriminator between models. As maximum likelihood has historically used for fitting parameters across many different contexts including GPs in Rasmussen et al. (2006), this is a simple extension of theory With no other information used to decide on the best model for the task of a given limited amount of data.

The information evaluated here has been used to update a prior without preference for a model to a posterior with a very strong preference for model 1. As a decision function it is highly motivated in this context by the variation in the model's standardised residuals rather than the centring of the standardised residuals.

This is largely because all models comparatively are better at specifying the value of a function rather than the uncertainty. A challenge it might face would be that it would favour overconfident well centred models over appropriately confident models.

The model choice proposed here is overconfident. This is likely a result of the fact that when training on a subset of global available information the model doesn't have access to the full possible range of building behaviour in relation to the weather.

Essentially each weather year is not a collection of uncorrelated draws from a 'true' ~~random~~ random weather, but rather a selection of highly related and in complex fashion correlated draws.

~~This would not be surprising to those working with weather data.~~ However the impact is profound when considering that the a series of correlated draws has different variation relative to the 'true' global distribution than uncorrelated draws.

This makes model such as GPs which attach correlation structure to their predictions essential to try to estimate the true uncertainty as it will be higher than if the weather is treated as a series of uncorrelated draws. The models here do not perfectly capture all of this uncertainty but model 1. does the best job, by filtering out zeros in the data, this manages a form of heterogeneous behaviour that makes GP less effective prediction tools.

3.5 Conclusions

This chapter provides a template for doing model selection under uncertain climate in the case of a building energy model using adapted GPs. For a real world problem such a varied data set of different weather years would not be available to test the accuracy of the statistical model against different building energy responses. Discriminating easily between modelling choices, such as in this case determining the value of a Bernoulli switch, allowed the modeller to test models before working with real world data as in chapter 4. The number of training years was selected ~~to be comparable to real world scenarios where a building might have been operation for two or more years, but not operating comparably to at present for hundreds of years.~~ to be comparable to the information available for the Information Commons use as the basis of chapter 4.

Using two years of cooling energy use data, from EnergyPlus, to predict another year has the following challenges:

- A tendency to underestimate cooling energy usage during active periods and over estimate during limited inactivity in less active periods.

- A large tendency to underestimate the uncertainty with errors often being large compared to predicted variance.

The simple mixture of experts model proposed here improve the performance of a GP. The mixture of experts model has allowed for better management of non-stationarity in the data whilst using a GP.

Likelihood based model selection is effective selection criteria in this circumstance ~~effectively~~ trading off model confidence against mean accuracy. In the case of the handling the uncertain climate in this chapter likelihood handles the trade off well.

3.6 Model Fitness

3.6.1 User Requirements

The user group for this chapter are those will the incorporate the model design into their model architecture for modelling building energy use. User group as outlined in section 1.4 is the individuals constructing model such as those in chapter 3.

The user group for model selection of an emulator for BEM models are likely to have the following priorities:

- Emulators are often used to enable Bayesian techniques **?**, making use of emulator's faster prediction times and inherent uncertainty quantification

- Predictions to have characteristics matching the energy use values predicted and be accurate.

- Those working with the models need to understand how the model was selected.

3.6.2   Chapter Output

This chapter produces the following outputs:

- A model architecture recommendation.

- Visualisations of model biases.

- A Bayesian likelihood of recommended model.

3.6.3   Evaluation of outputs against user requirements

The user requirements are that the model or, formally, the emulator runs faster than EnergyPlus simulations, this is met by the model architecture. Whilst the valuation of the model selection requires computationally intensive Monte-Carlo Methods to evaluate the dual behaviour model produces predictions for very low computational cost. For discussion of computational costs of training and a possible solution see sections 3.2.4 and 3.2.6 for discussion of this further. Gaussian processes and other black box models are inherently low cost prediction and have higher training cost. White box models such as physical model tend to have higher information requirements, lower to zero training costs and long prediction costs.

The predictions of the the dual behaviour Gaussian process have known accuracy. Focusing on figure 3.5, a visualisation this chapter doesn't give real values of difference between the 'real data' and the predicted data from the emulator. The moderate negative biases do communicate that the emulator is under predicting energy use whilst the spread of the standard deviations suggest that the model is also over confident. Both of these elements are problematic for the user as being unable to communicate the real terms accuracy could have a impact on ability to get stakeholder buy for decisions dependant on the the model. This partly address by the figures 3.6,3.7 & 3.8 which are discussed in section 3.3.2 comparing relative performance. One issue that remains unaddressed is the negative energy use prediction shown in grey which are not possible the EnergyPLus model being emulated.

The accuracy of the model needs to be empirically linked to the model performance against weather, not all the different aspect of the BEM's capabilities. This should be measured against the amount of uncertainty that the model and the uncertainty introduced by uncertain weather introduce into the prediction.

The likelihood selection criteria is not accessible to those without a statistical or mathematical background. The discussion of the bias in section 3.3 does explain the least over confident model was picked on the basis of the likelihood criteria, the concept of over confidence can be discussed as the most accurate to stakeholders even if the logic behind this is unclear to those without a basic grounding in uncertainty analysis. The user would likely have this level of understanding as such this is not a high level of concern. This discussed in more detail in section 2.7

3.6.4 Recommendations

The two main concern raised in the evaluation above were:

- Lack of real term accuracy measurement, in particular evaluating the proportion of uncertainty that is contributed by weather opposed to the GP modelling.

- Negative energy use predictions not being possible the EnergyPlus model or the real world.

This chapter would recommend that potentially further work could be done to address the realism issue with the model predicting negative energy values. In section 3.2.3.3 a decision was made to not model the data under a log transform the data, this decision should be re-evaluated in the next chapter 4.

The second recommendation would be to provide actual estimates of accuracy in real terms of the energy use predictions for the cooling systems. This will make communicating model accuracy to stakeholders and ensuring stakeholders are able have confidence in the model.

Figure 3.6: Plots of estimated cooling energy use against day of the year. Training years listed alongside the predicted year.
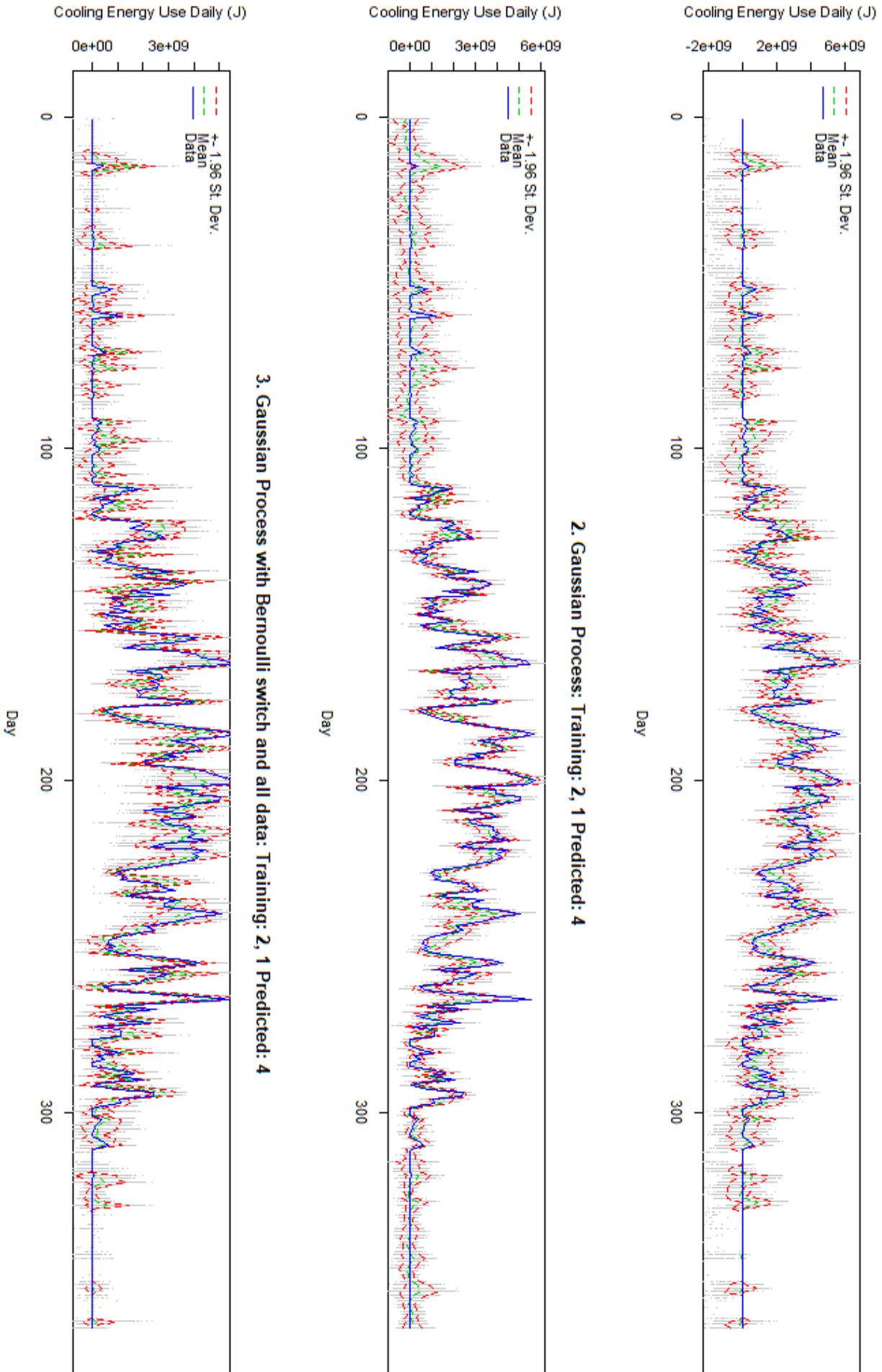
Figure 3.7: Plots of estimated cooling energy use against day of the year. Training years listed alongside the predicted year.
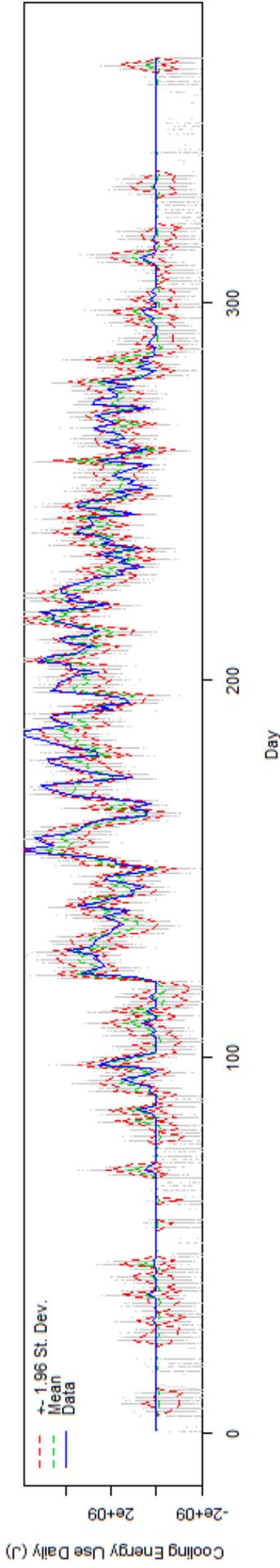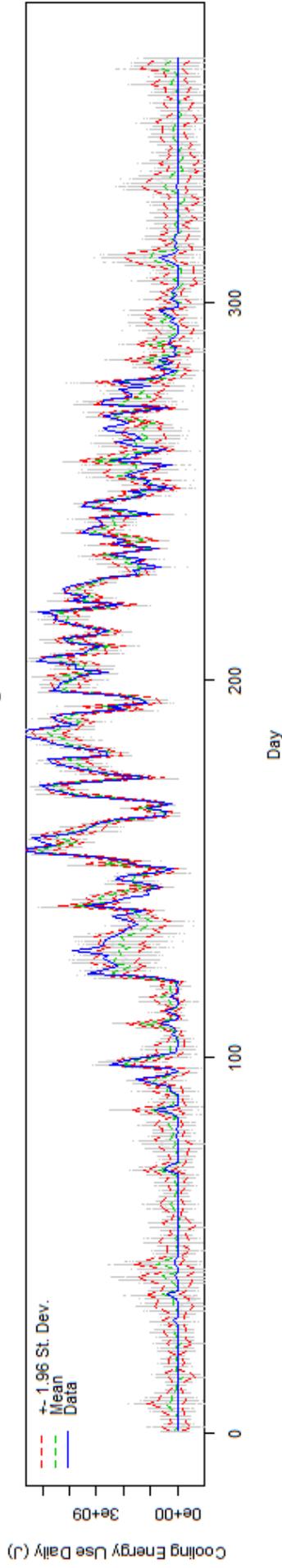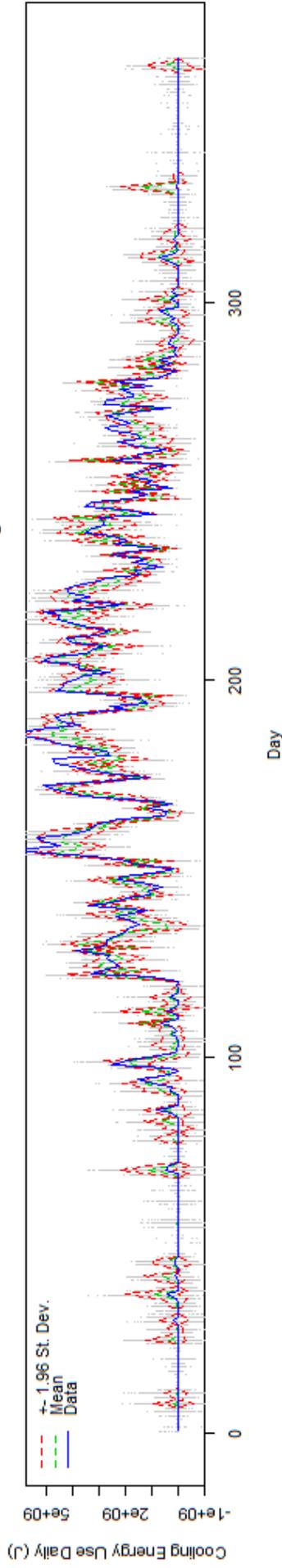
Figure 3.8: Plots of estimated cooling energy use against day of the year. Training years listed alongside the predicted year.
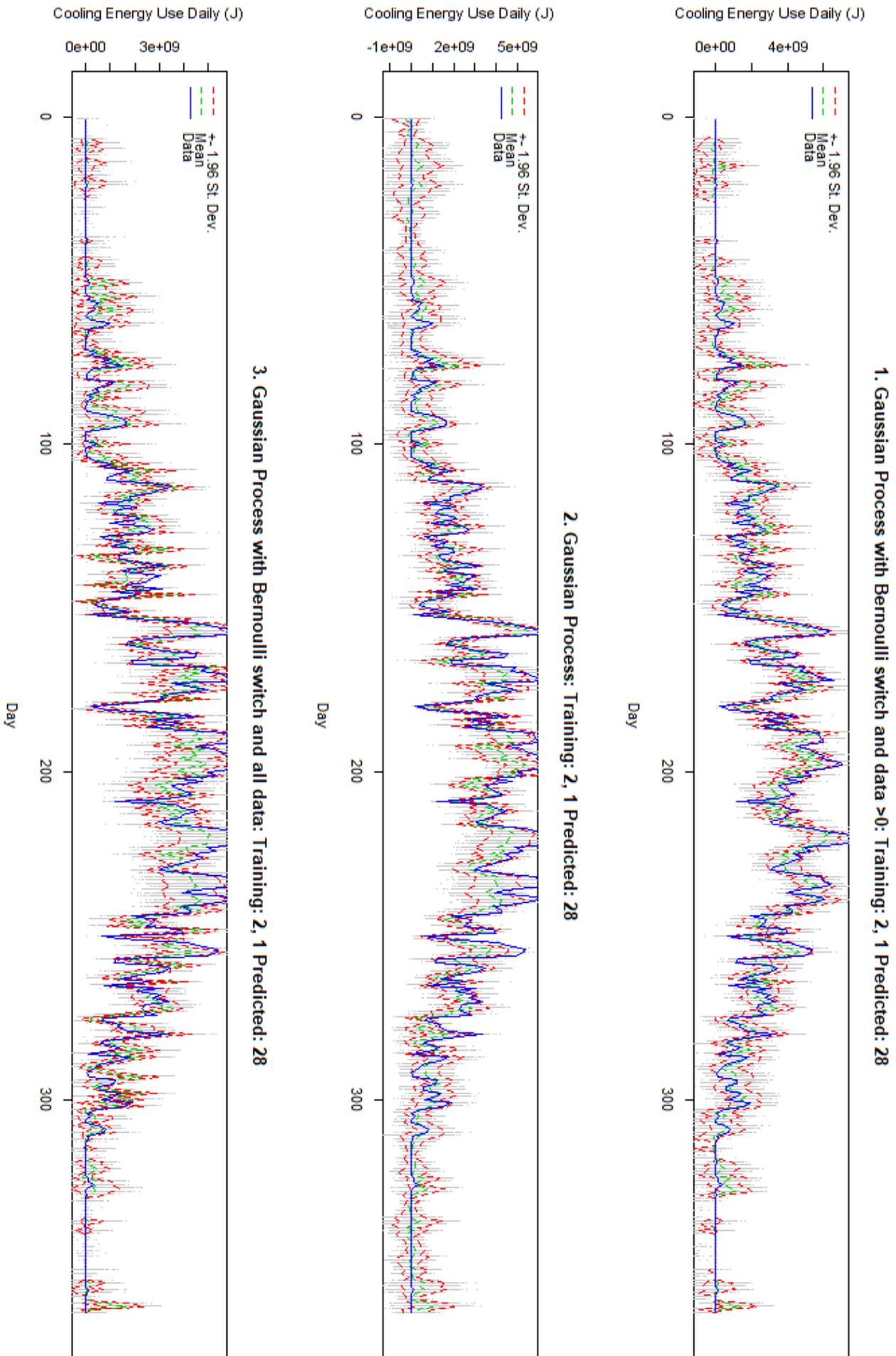
Chapter 4

Statistical modelling of climate change impact on chiller energy use

In this chapter I develop the methodology used in chapter 3 to emulate the behaviour of a modelled building energy system to generate climate change projection for an existing building, the Information Commons at the University of of Sheffield.

This chapter makes use of a nearby weather station to provide historical weather data to pair with the historical energy demand of the building which provides training set to train the model from the previous chapter. ~~This~~ The model ~~,~~ , proposed in this chapter is then used to emulate building behaviour for future climate scenario provided by down-scaling of the Met-office forecasts.

~~I conclude that~~The results of this chapter will provide insight into whether the cooling energy use is likely increase as a results of climate change.

Keywords— Building Energy Modelling - Climate Change - Climate Risk - Gaussian Processes - Model Selection - Uncertainty Quantification

4.1 Introduction

Uncertainty issues are common around climate change forecasting Tian & De Wilde (2011). Weather forecasts for future climates contain significant uncertainty that should be incorporated to facilitate decision making in light of this uncertainty. Propagating weather uncertainty through the behaviour of building is difficult. This stems from the costs, in terms of labour time and computer run time, of building and running a building model making propagating uncertainty. ~~This means a need exists for methods that are able to do this process with lower costs.~~

Upon being presented with a new building project and being asked to produce estimates for some internal variable of a building, building service engineers might have an instinct to power up their favourite desktop simulator familiar BEM modelling tool (such as EnergyPlus Energy (2010)). Their objectives are likely to be tasks such as to estimate thermal comfort inside a building or the cooling energy used in future for existing or new builds or by the need to reach building-related CO_2 emission targets (Committee on Climate Change (UK) 2015). However, due to ~~costs~~time and information access constraints this could be prohibitive in pilot studies, this chapter suggests that ~~statistical~~dual-behaviour GP models, from chapter 3, could be used to conduct pilot studies with the added benefit providing improved uncertainty quantification.

Many buildings built in the last decade have inbuilt sensors and automated metering that provide records of the building's past energy use. This has been a response to several different waves of legislation and guidelines in different countries encouraging uptake of automated metering Ahmad et al. (2016). The micro-climate around buildings are being increasingly monitored and surveyed more than before, an example of this sort of research is Munir et al. (2019). If local weather is needed, there might well be a public weather station in the local area.

Classically this data might be used to calibrate a Building Energy Model (BEM). However, an increasingly empirical or machine learning approach is used as an alternative Zhao & Magoulès (2012), Nguyen et al. (2014), Wang & Srinivasan (2017). In this chapter, a statistical model, utilising Gaussian processes, will be used to bypass the creation of a BEM simulator. This approach puts aside the modeller's opinions, replacing them solely with a data trained model. Whilst implementation of a Gaussian process is a recent addition to BEM, this paper attempts to take things back to basics and explain the reasoning and method in using machine learning.

In tThis chapter, I make use of the data from draws a distinction is drawn between the energy used by the roof top plant HVAC (Heating and ventilation Air Conditioning) and in this case refers solely to the air circulation pumps with the cooling provided by chillers and the heating provided by a distinct heating network.

Climate change is a phenomenon that is expected to have an increasing impact on society (IPCC 2014). The UK Committee on Climate Change has stated that building related CO_2 emissions are part of key emission targets (Change 2017). The ability to provide site specific climate change projections for buildings could help drive delivery of CO_2 emission targets. As a test case, this paperchapter presents an empirical approach for making energy use predictions for the Information Commons building, at the University of Sheffield, under future and past climate from the UK Met-Office Murphy et al. (2010).

These sorts of study can be enhanced by more detailed information from many sources. To make these projects viable, information has to be recorded before the conceptions of the study which means retaining as much data as possible.

4.2   Subject: Information Commons

The Information Commons (IC) at the University of Sheffield is a library and work space situated near the city centre (Lewis 2010). It was opened in 2007 and will provide support to students into the second half of this century.

As a study space it is open nearly all day every day of the year. This means there is continuous demand to maintain the environment.

The building was designed to have a low environmental impact. Measures taken to achieve this include energy efficient conditional air modules (CAM, utilising air distribution via underfloor air spaces and floor mounted fan tiles), climate control, high performance thermal insulation, intelligent lighting and participation in the Veolia district heating scheme.

While buildings like the IC are a flagship building, a group of buildings that are inherently unique combining new technologies built for clients, they often have striking visuals such as internal balconies. Dynamic BEM of such buildings can be expensive in terms of time (even more so if with limited access to detailed building information). This class of buildings also tends to be heavily metered, as a result of their proclivity for technological uptake, making it an ideal building type for the application of machine learning techniques.

Classically the modelling of the IC would require the blueprints and a detailed survey to provide the information for example an EnergyPlus (Energy 2010) simulation. The building's irregular interior would be difficult to produce simple zones from without creating many problems from estimations and approximations. Building designers have attempted to develop techniques to optimise parameters. For example using machine learning and other techniques with BEM to optimise window size and other parameters Wood et al. (2015).

Electricity Used 01/05/15 – 30/04/16: 8.449e+05 kWh   Electricity Used 01/05/16 – 01/05/17: 8.315e+05 kWh
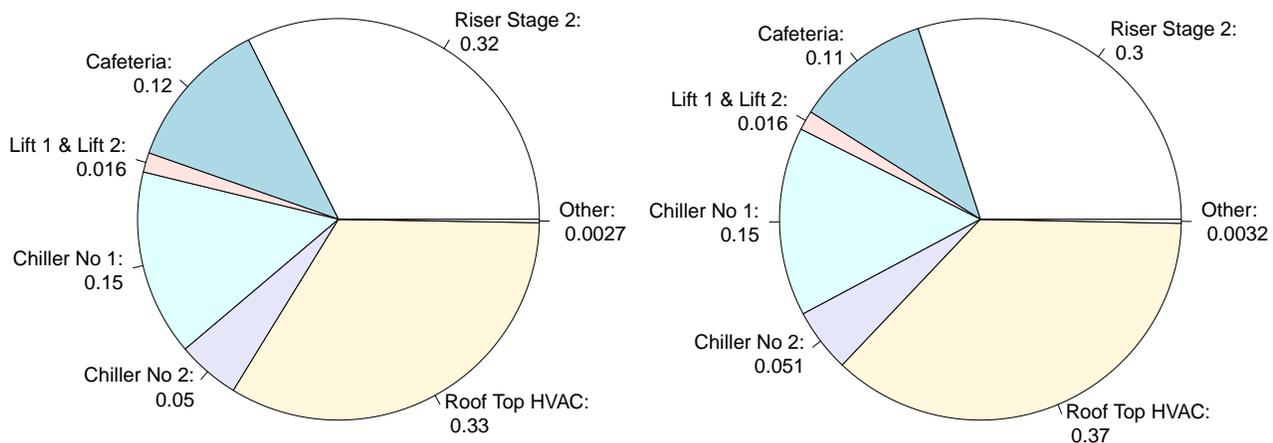
Figure 4.1: Information Commons energy use by system by year for years: 01/05/2015 to 30/04/2016 and 01/05/2016 to 01/05/2017

Figure 4.1 shows the IC's electrical energy used across the metered systems. The electrical energy use modelled here in this chapter will be the Chiller No 1 and 2, the systems that cools circulated air.

4.2.1   Choice of variable for modelling

The three systems with an apparent link to the maximum daily temperature were the two Chillers and the district heating use (figure 4.2 provides plots of energy use against temperature) were considered:.

- The heating energy has metering changes in the data, requiring additional modelling to rectify, these ~~presents~~ challenges ~~from changes in metering frequency,~~ make the heating energy use unnecessarily complex to model.

- The Roof Top Plant HVAC is a poor candidate due to the very weak relationship with the top maximum temperature. The relationship was found to even less apparent visual inspect with other variables as such the variable was rejected.

- The Chiller 1 & 2 energy however whilst, they retain significant variability that might require the use of a non-linear model such as a Gaussian Process, do show a linear relationship with temperature making this a good candidate for modelling in this chapter. (It should be noted the the combined values of chiller 1 & 2 were used due the general inactivity of chiller 2 and the lack of operator and designer knowledge available for the work as to why the two chiller with the same specifications were utilised so differently.

which means As such this paper chapter will focus on the sum of Chillers 1 & 2 energy use. Some oOther systems were dismissed as being not considered due to using very small amounts of total energy use.
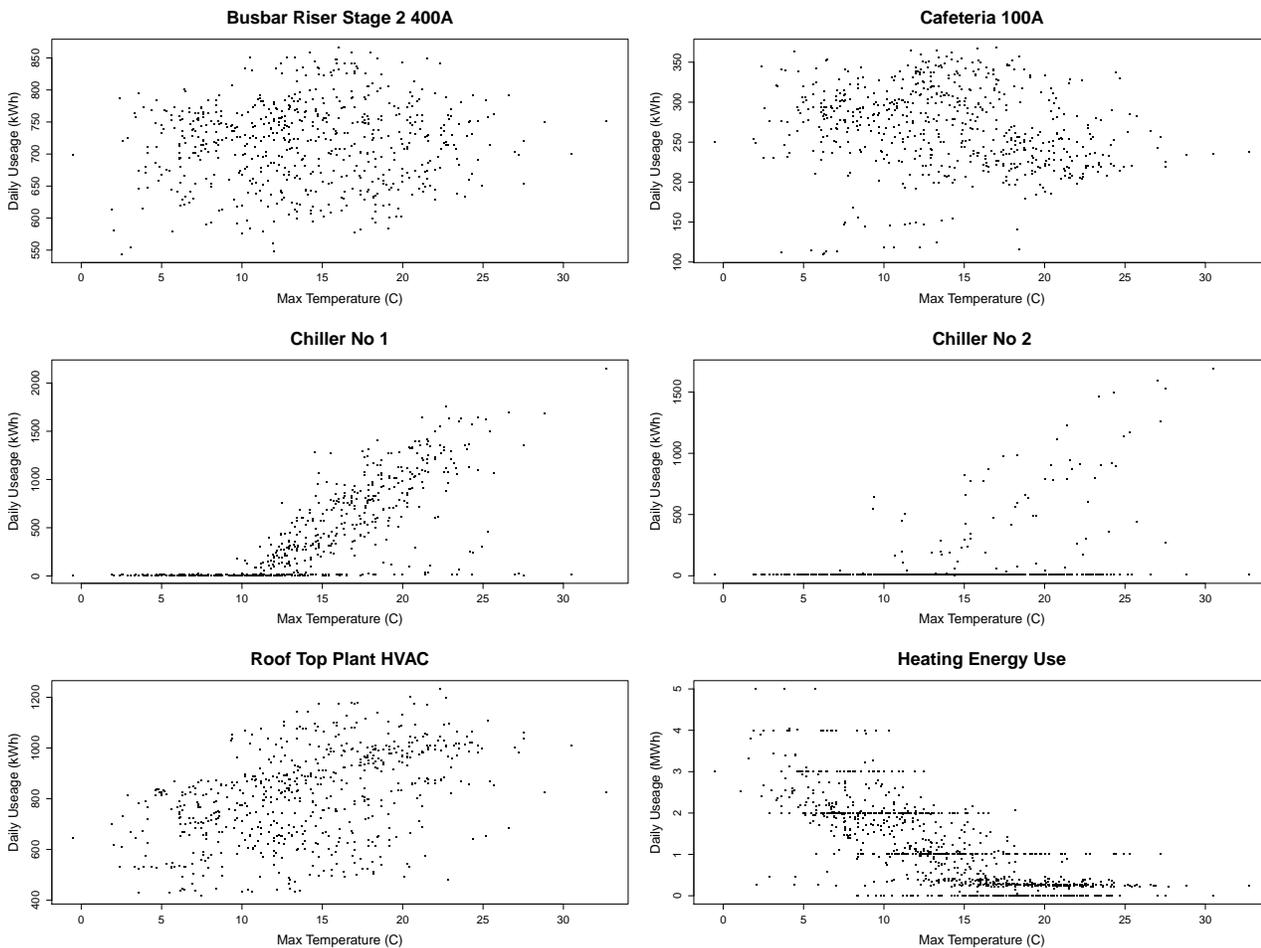
Figure 4.2: Selected Plots of Information Commons energy use against maximum daily temperature in Weston Park. Weston Park is around 600 meters away from the building.

4.3 Methodology

Many building parameters are uncertain, which necessitates the use of probabilistic models. This chapter combines Gaussian processes with a sampling approach to evaluate multiple sources

of uncertainty. The ability of a GP to handle multiple sources of uncertainty, as has been discussed in previous chapters, is the main motivation for using it as an emulator.

There are two main sources of uncertainty. The first is the inability to produce a perfect building energy model, the second is imperfection in weather forecasting. This chapter makes use of the Met-Office climate forecasts (Murphy et al. 2010), which contain multiple forecasts. By using 100 of these forecast samples, this chapter allows for variation in the weather.

A recommended text for the the justification of the detailed theory of Gaussian processes is Rasmussen and Williams (Rasmussen et al. 2006).

4.3.1   Data

There are three main data-sets used in this work: weather station data from Weston Park Museum, present local climate around Weston Park area; meter readings from systems of the Information Commons from the University of Sheffield Estates; forecast weather for Sheffield from 2021 to 2050 from the UK Met-Office UKCP09 data-set using the high emission scenario (Murphy et al. 2010).

The Weston Park data-set was provided as daily readings with values for maximum and minimum daily dry bulb temperature, average relative humidity, sunshine hours, wind speed and direction and rainfall as inputs for the empirical model. The Met-office data was formatted into the same format. The energy use data from Chiller one and two were added together, anomalous readings were discarded.

4.3.2   Modelling

~~This study uses Gaussian processes (GPs).~~ GPs are a particularly interesting class of non-parametric models. They can perfectly reproduce data given to them whilst retaining a smooth interpolation between values that extends easily into many dimensional input spaces. Their hyper-parameters control the way in which they interpolate between data points. An overview can be found in section 2.6. The GP model as used in this chapter is detailed below.

The relationship modelled can be described (the green box in figure 4.3):

$$\mathcal{F}(\mathbf{w}) = \mathbf{e} \tag{4.1}$$

where $\mathbf{w}$ is weather (max and min temperatures, sunshine hours, day of the year), the output is the energy use on the day is $\mathbf{e}$, with some function, $\mathcal{F}$, which describes the relationship between inputs and outputs. Gaussian processes are particularly useful in describing this $\mathcal{F}$ when the modeller doesn't believe $\mathcal{F}$ should not change ~~much~~by a large margin between similar values of $\mathbf{w}$ (or be smooth). More specifically it assumes these changes in $\mathcal{F}$ are Gaussian.

Comparable models comprise models such as auto-regressive models (or AR models), splines, linear regression with Gaussian noise and radial basis function networks, all of which are in some cases special cases of the GPs and vice versa. The amount of flexibility and inherent uncertainty quantification make GPs a powerful tool in modern machine learning methods. ~~More detail can be found in (Rasmussen et al. 2006).~~ There are several excellent packages implemented in R and Python for utilising GPs. This chapter uses the DiceKriging package in R (Roustant et al. 2012), another good package in R is the kernlab package more focused on classification (Karatzoglou et al. 2004) and in Python the GPy project (GPy 2012).

The standard GP has homogeneous noise across the input space and realisations are smooth and continuous using an Gaussian covariance function. This is counter exemplified with the inactive

winter period, when the Chillers are not working due to cooler air brought in from outside, and the active summer period, where demand is varied. The inactive period can be seen in figure 4.2, middle row, with the flat near zero values at the start and end of the year. For a more detailed discussion of uncertainty quantification using GPs see Vernon et al. (2010), O'Hagan & Oakley (2004).

The model used here has two states, one active, one inactive. The more complex behaviour is captured by the active state. To cope with these two states, a second layer or classifier is added to the model. This model can be summarised:

$$\mathcal{F}(\mathbf{w}^*|\mathbf{e},\mathbf{w}) = \begin{cases} \mathcal{N}(\eta_1) & \text{(inactive state)} & \text{with probability} & \chi(\mathbf{w}^*|\eta_3) \\ \mathcal{GP}(\mathbf{w}^*|\mathbf{e}_a,\mathbf{w}_a;\eta_3) & \text{(active state)} & \text{with probability} & 1 - \chi(\mathbf{w}^*|\eta_3) \end{cases} \tag{4.2}$$

where $\mathbf{w}^*$ is the value of weather parameters to predict energy use, $\mathbf{e},\mathbf{w}$ are the training data, $\mathcal{N}(\eta_1)$ represent the normal distribution with a non zero mean (both mean and variance parameters), η_1 are the parameters of the normal distribution calculated from inactive training data both mean and variance, $\mathbf{e}_i,\mathbf{w}_i$, $\chi(\mathbf{w}^*|\eta_3)$ is the classifier's probability of inactive Chiller at $\mathbf{w}^*$ given parameter η_3 trained using $\mathbf{e},\mathbf{w}$, $\mathcal{GP}(\mathbf{w}^*|\mathbf{e}_a,\mathbf{w}_a,\eta_2)$ is the posterior GP at $\mathbf{w}^*$ given the active training data $\mathbf{e}_a,\mathbf{w}_a$ with hyper-parameters η_2.

$\chi(\cdot|\cdot)$ is the classifier's probability that distinguishes between active and inactive states. This chapter uses a binomial classifier with a logit link function in linear regression against the input parameters. This is a special case of the generalised linear model. Formulated as follows:

$$\chi(\mathbf{w}^*|\eta_3) = \frac{\exp(\gamma_0 + \boldsymbol{\gamma}\cdot\mathbf{w}^*)}{1+\exp(\gamma_0 + \boldsymbol{\gamma}\cdot\mathbf{w}^*)} \tag{4.3}$$

where $\{\gamma_0, \boldsymbol{\gamma}\}$ are the linear coefficients (d representing the dimension of $\mathbf{w}$) and comprise η_3. ~~These parameters are selected by maximum likelihood.~~

The parameters (η_3) were selected using maximum likelihood using all the training data. $\chi(\cdot|\cdot)$ acts to mix the inactive and active distributions. This is in a similar vein to the work done by Tresp (Tresp 2000), Rausmussen (Rasmussen & Ghahramani 2002) et al.; Park et al.Park & Wang (2017).

The active state is modelled using a GP, $\mathcal{GP}(\mathbf{w}^*|\mathbf{e}_a,\mathbf{w}_a,\eta_2)$. This follows from it containing the most complex behaviour and the GP being able to cope with this behaviour. GPs are defined in terms of mean and covariance functions for the prior. The Gaussian process used here uses a mean function linear in only the raw inputs and can be written:

$$\mu(\mathbf{w}) = \beta_0 + \boldsymbol{\beta}\cdot\mathbf{w} \tag{4.4}$$

where μ is the mean function, $\{\beta_0, \boldsymbol{\beta}\}$ are linear coefficients and $\mathbf{w}$ is the weather inputs. Only linear basis functions are considered as they are include to remove linear trend, the non-linear response in the model have no know form. Placing additional basis functions into the linear prior could introduce error from the inherent assumptions. The posterior (which is multivariate normal) mean for the Gaussian process can be written:

$$K(\mathbf{w}^*, W)K(W, W)^{-1}(\mathbf{e} - \mu(W)) \; + \; \mu(\mathbf{w}^*) \tag{4.5}$$

where $K(\cdot,\cdot)$ is the covariance function, $\mathbf{w}^*$ the weather values to evaluate the posterior at, W matrix of training locations and $\mathbf{e}$ the energy use at training locations. The posterior covariance matrix can be written:

$$K(\mathbf{w}^*, \mathbf{w}^*) \; - \; K(\mathbf{w}^*, W) K(W, W)^{-1} K(W, \mathbf{w}^*) \tag{4.6}$$

The GP in this chapter uses a linear model as a mean function with a Matérn covariance function with shape parameter $\nu = \frac{5}{2}$. The Matérn covariance function incorporates a measure of distance with scale hyper-parameters. Using the data deemed 'active' data, above a manually set threshold, the scale hyper-parameters were optimised using maximum likelihood. Predictions are then drawn from the posterior. The Matérn covariance function can be written:

$$K(r) = \sigma \cdot \left(1 + \sqrt{5}r + \frac{5r^2}{3} \right) \exp\left(-\sqrt{5}r \right) \tag{4.7}$$

where r is the distance between $\mathbf{w}$ and $\mathbf{w}'$ defined $(\mathbf{w} - \mathbf{w}')^T \mathbf{l}^{-1} (\mathbf{w} - \mathbf{w}')$, σ the variance parameter and $\mathbf{l}$ the length scale parameters. So η_2 is comprised: $\{\beta_0, \boldsymbol{\beta}, \sigma, \mathbf{l}\}$

Further formulae (Matérn function and posterior) can be found in Rasmussen et al. (2006).

The inactive state is modelled using a normal distribution, $\mathcal{N}(\eta_1)$. The mean and variance (in η_1) are directly calculated from the 'inactive' data below the set threshold.

The possibility of log transforming the data was considered during the modelling stage. As such small samples of the data were modelled with the data transformed using a log transform. The prediction accuracy, using mean squared error, was lower as a result of the space deformation. This was evaluated during the modelling process as not beneficial considering only this aspect. For discussion of this in relations to the recommendation on using the log transform in section 3.6.4, please see section 4.7.3 at the end this chapter.

4.3.3 Predictions

The trained model was used to create predictions between the weather experienced external climate and the energy used by the Information Commons Chillers.

As a surrogate to estimate the weather experienced in Weston Park over the time period from 2021 to 2050 this chapter takes the Met-Office Climate Projections from 2009 (Murphy et al. 2010) for the 25 km square grid cell containing Weston Park. The climate projections are from the Met-Office high emissions scenario; this is a highly-possible scenario due to the very high level of CO_2 observed recently. The likelihood of this scenario is taken from the extreme changes in the global economy required by the emission in (IPCC 2014) and Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor & (eds.), as of the IPCC report in 2014 the emissions had been exceeding the high emissions track. The control scenario is the Met-Office estimate of weather from 1961 to 1990 and is a baseline from which climate change can be measured.

These projections are probabilistic and are downloadable in 100 sample sets. This work uses one download set or 100 samples and makes projections for each of these samples. Our model is similarly probabilistic and as such it will also need to be repeated sampled.

These samples are often described as realisations of the distribution. In order to make predictions from the probabilistic model, in this chapter the distribution is sampled and the median and 2.5% and the 97.5% quantiles were calculated (a 95% interval). The 95% confidence interval was picked due it's prevalence in statistical education, to make communication of the result easier.
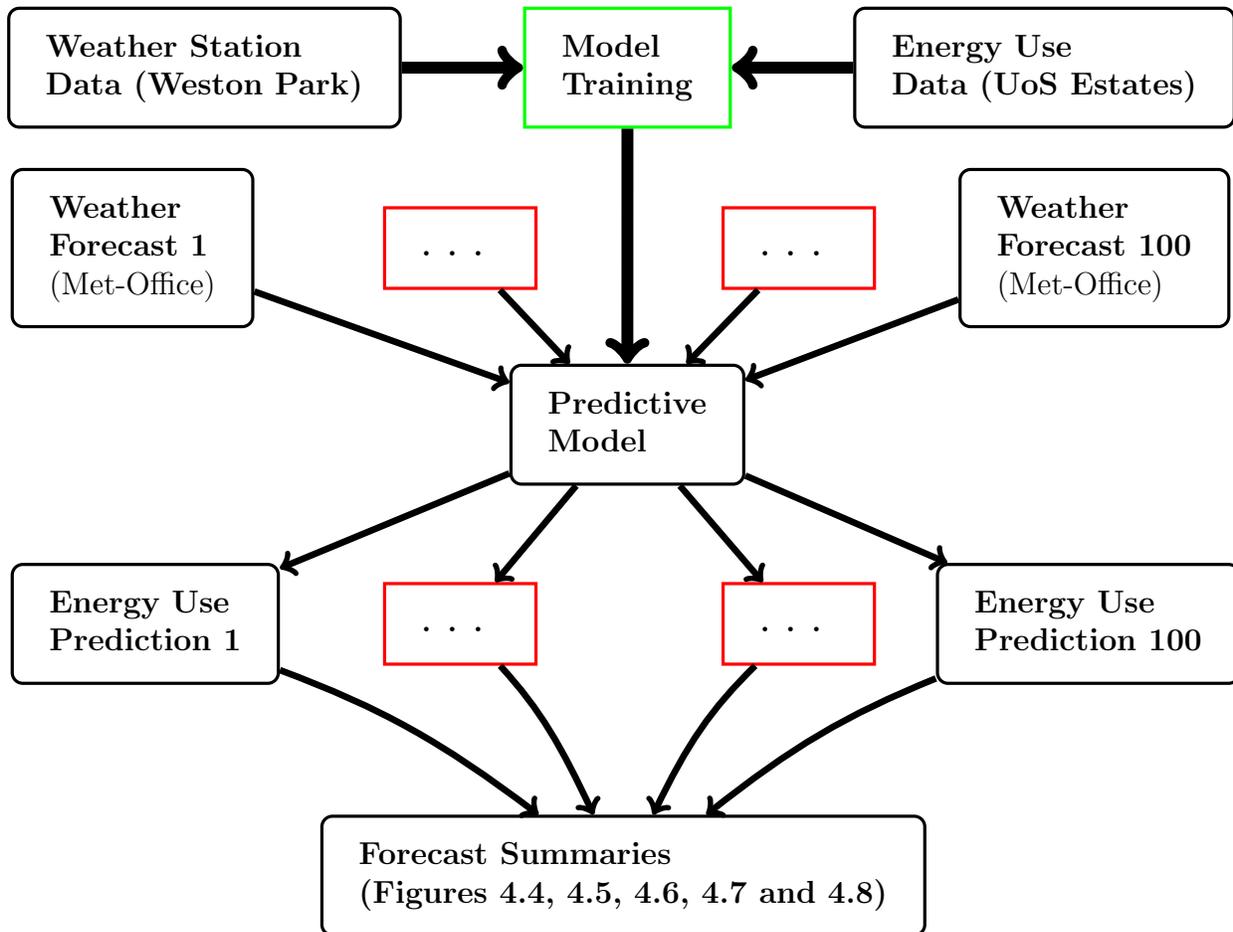
Figure 4.3: Work Flow Diagram: Sequentially the model i trained using the known energy use and recorded weather data, the resultant model is then used as predictive model for the weather forecasts to produce predictions, which are then summarised.

The structure used is by forecasts to go through each year and produce 2000 samples for the produced model for each weather sample for each year at once.

This scheme can be summarised as follows:

- For each year of the period 2021-2050:

    1. For each year's 100 weather samples:
        a) Draw 2000 samples from each class (active or inactive) predicted for each day for the whole year using forecast $\mathbf{w}$ values.
        b) Calculate the predicted probabilities from the classifier model for each day.
        c) Draw the proportion of the 2000 samples from each class according to the predicted probabilities for each day.
        d) The 2000 mixed class samples from the previous step are used as predictions for this weather sample for each day.

    2. Group and save the year's predictions. This gives 100 weather year samples of 365 days with 2000 samples ($100 \times 365 \times 2000$).

- Total predictions are 30 years of sample groups for each system predicted ($30 \times 100 \times 365 \times 2000 = 2.19$ billion).

4.3.4 Variance Decomposition

Variance decomposition is a technique used to attribute how much variation can be attributed to different sources. Informally the variance attributed to the weather is distilled by calculating the variance in the expectation of each weather day's samples. Take out the variation in the samples grouped for a single value of the weather, then calculate the variance. To analyse the remaining variance the same process is carried out in reverse. The expectation for each group of sample associated with a weather day is then subtracted from that group. The variance is then calculated across all the modified groups. Further detail and discussion can be found in Saltelli et al. (2010).

The variance decomposition more formally would be evaluated:

$$\begin{aligned}
\text{Var}_{\mathbf{w}}(\mathbf{e}) &= \text{Var}\left(\text{E}(\mathbf{e}|\mathbf{w})\right) \\
\text{Var}_{-\mathbf{w}}(\mathbf{e}) &= \text{Var}\left(\mathbf{e} - \text{E}(\mathbf{e}|\mathbf{w})\right)
\end{aligned} \tag{4.8}$$

where $\text{Var}_{\mathbf{w}}(\mathbf{e})$ is the variance of the building energy use prediction attributed to the weather input variance, $\textbf{Var}_{-\mathbf{w}}(\mathbf{e})$ the rest of the variance, $E(\mathbf{e}|\mathbf{w})$ is the expectation of predictions on a day given the weather.

4.4 Results

This section presents the results of the study. In section 4.5, ~~I~~this chapter discusses accuracy and possible validation of the results. The purpose of this chapter was to provide a case study of the application of the dual-behaviour GP from chapter 3 and characterise the uncertainty of the output. A simple model would not have variable uncertainty therefore the results would not have been comparable.

The modelling of the energy use suggests that the energy use from cooling systems in the Information Commons will increase in response to climate change. In addition a summary of future climate forecasts values show how peak usage occurs throughout the year and the likelihood of the different maximal values. Figure 4.4 supports this statement this in two ways:

- Firstly, the current energy use, horizontal line in red, is lower than the mean predicted energy use in the future for 25/30 years between 2021 and 2050. In addition only one of these is significantly lower than this mean the rest being very close to this mean value for that year. This suggested future year will be likely warmer than the current climate experienced. The Information Commons will cost more to cool in the next 30 years compared to the present.

- Secondly, the baseline scenario examples, in light grey, show that the energy use in future climate opposed to the climate against which level of global warming are measured will be significantly higher. The Information Commons will experience significant impact as a result of climate change on the same scale as global temperature changes between 33% and 50% of chiller energy use.

This ~~sort~~ information has potential to be used to review the systems in the light of changes climate will make to their operating environments. Suggesting that the University of Sheffield

should consider accounting for the increased cost of running the Information Commons chilling unit into the next 30 years.

4.4.1   Discussion of uncertainty in ~~E~~energy ~~U~~ucse forecasts

The predictions for the energy use between 2021 and 2050 are presented in figure 4.4 alongside a breakdown of uncertainty. The significant year to year variation is the result of the variation in the Met-Office weather data. Year to year weather can be varied as a result of multiple factors such as solar cycles and El-Nino effects to cite two examples of inter-annual weather effect that contribute to annual variation of weather.

The horizontal line, in the top plot, is the baseline energy use. It shows the location of the total energy used per year used over the training data period. This can also be considered the current Chiller energy use of the Information Commons. The predictions, in black error bars, are for the climate change scenario under the high emissions scenario from the Met-Office. The fainter smaller grey error bars represent an estimated energy used under the control scenario, based on weather data from 1961 to 1990.

The uncertainty, shown in the lower plot in figure 4.4, is larger for future years mainly due to the lack of knowledge about future weather. This can be deduced from the decomposition of uncertainty plot (bottom figure 4.4). The uncertainty in the climate change scenario is largely dependent on the uncertainty in future weather (blue or dashed-dot part) opposed to the uncertainty in the GP model (red dashed part). The uncertainty of the control scenario's energy use predictions is lower, likely because the weather in the past has already been observed opposed to the future weather, which has yet to be measured.

The predictions, in the top part of figure 4.4, for the climate change scenario is a general increase in total Chiller energy use. Whilst the uncertainty of the predictions do allow for 8 years being more likely to have lower energy use than current energy use. Most of these are marginal, having nearly the same chance of having more or less energy use than the present. Whilst many of the other predicted years (up to 2050) represent significantly higher chances of higher energy use.

The contrast between the predictions for the control scenario and the climate change scenario is marked. The fact there is no overlap between the predictions for pre-climate change and the present is incidental support for the climate change happening in changes detected in buildings' Chiller energy use, such as the Information Commons observed in this study.

4.4.2   Peak Energy Use

Figure 4.5 shows the probability distribution peak time (top) and value (bottom) for energy use in the climate change scenario. This was carried out as a demonstration of application empirical predictions of estimating system Chiller stress.

The peak value for the two years in the training data was extracted to provide comparison. The 2017 peak was noticeably lower than for 2016. The highest values from the two training data years have are a good fit for the black distribution curve in figure 4.5, both are well within the majority of the distribution's mass. In terms of the predicted distributions, the 2017 peak use is in the bottom 5% of future peaks and 2016 is more severe than 67% of future peaks.

The distribution for the top 1 peak value (of the year) is quite likely an outlier. This means the model in places relies heavily on individual values. The top 1 value distribution has a peak
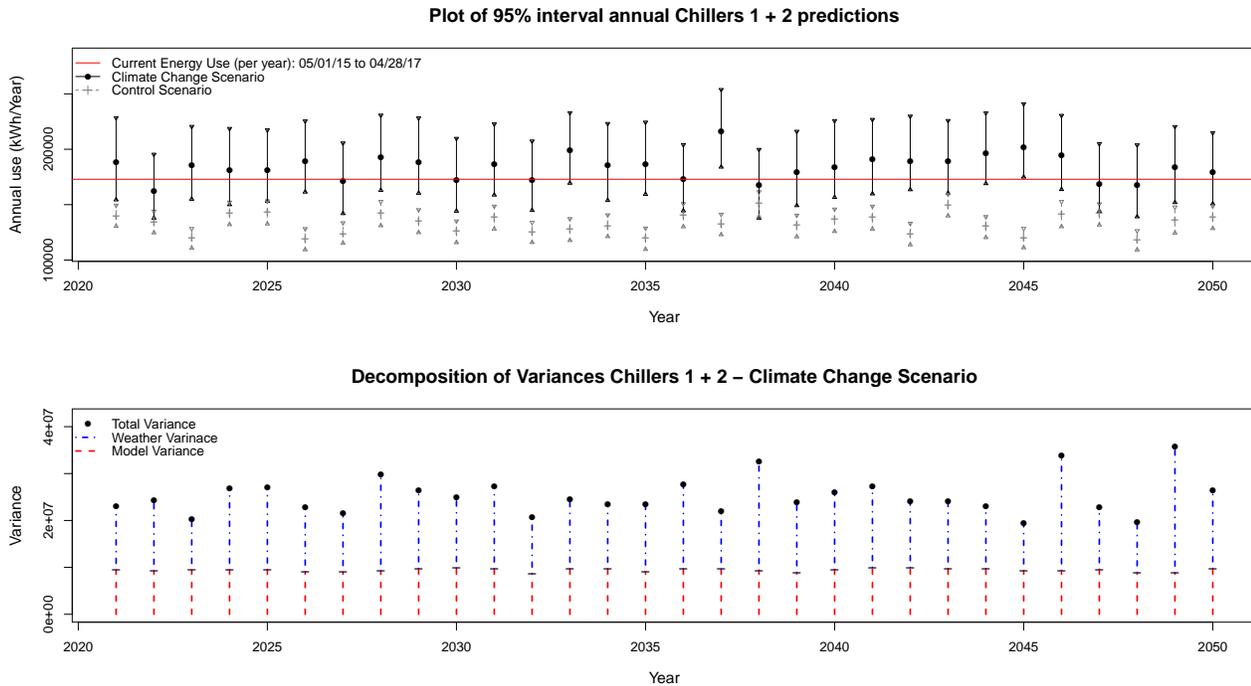
Figure 4.4: Predicted energy use for Chillers 1 and 2 combined. The 95% confidence plot has the international recognised baseline scenario for climate change Information Commons performance in light grey and the forecast energy use for 2021 to 2050 in black. The variance decomposition plot breaks down the sources of uncertainty into the variance from model performance, bottom, and weather uncertainty, top.

exactly at the point of the 2016 peak (the blue vertical line). The model is attributing high values here, solely because it has seen high values at this time of year, not because of a modelled relationship with the weather. The fact that the top 1 peak value peak in density is not present in the distribution of the 10 highest values, suggests the model requires further improvement.

The fact that peaks are only likely to occur in the summer period, serves positively in a sanity check. The conclusions that can be reached reliably about the time are of no great surprise, high Chiller use occurs in the summer. This discussion, presented in the last few paragraphs, however has revealed that whilst the model is sensible under scrutiny, it does have limits.

4.5 Discussion

4.5.1 Evaluation of modelling choices

The usual approach with models is to verify the model by checking the predictions against known data. There are several key challenges here with this approach, which this chapter will address, however it is import to communicate the limitations of the approach as explained below:

- The chapter is about forecasts/projections for energy use. There doesn't exist energy data for the values that this ~~paper~~chapter predicts. The climate change happening now isn't comparable to past changes. ~~The future's actual data is likely to be quite different~~It is very
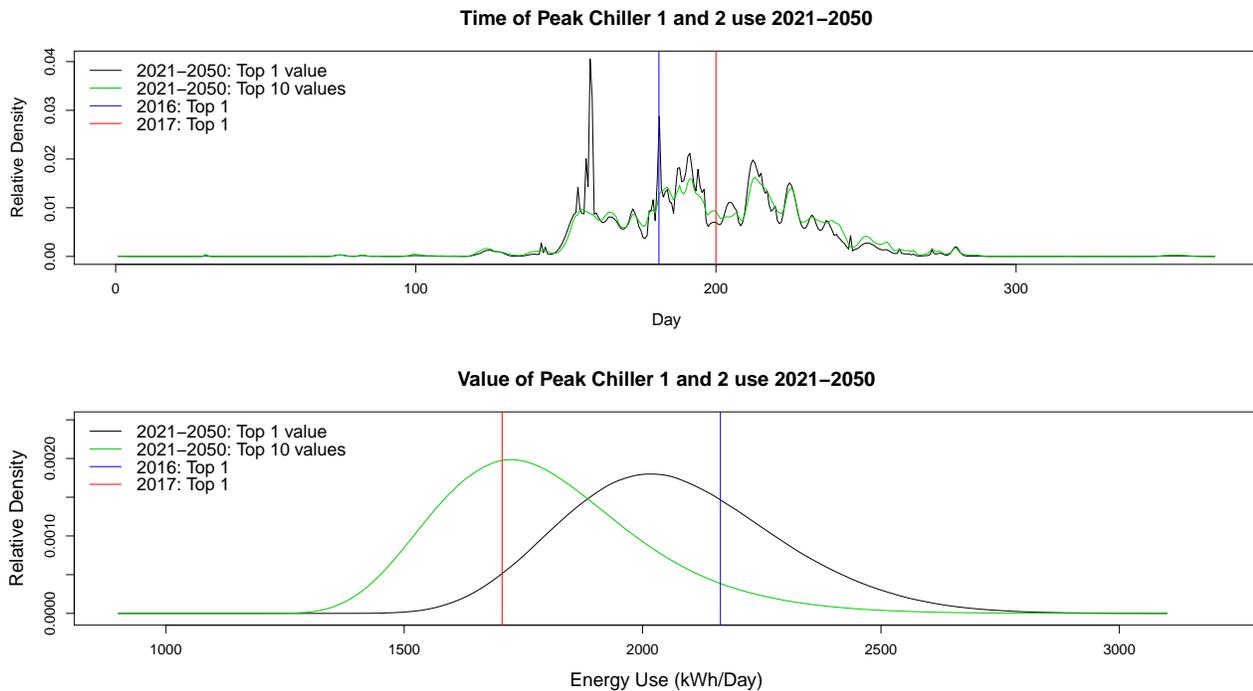
Figure 4.5: Plots describing the characteristics of predicted peaks

difficult to verify if the changes seen in past will similar to the present, as such accuracy will be poorer than validation on current data sets would indicate.

- The model in the paper uses two years ~~worth~~ ofchiller energy use data. Sub-dividing the data into training and test data-sets would mean that the model has less predictive power due to a smaller quantity of data, which limits verification for the actual model in use.

- Cross validation can't be used to evaluate accuracy of forecasts. ~~Some unpublished work by the author~~The result of the previous chapter suggests this method could have under-dispersed predictions. Bench-marking for this problem is hard to establish without much more extensive work.

In light of these challenges, this paper provides several plots as visual diagnostics. This will demonstrate the general model performance and justify the use of modelling both Chillers combined.

Figures 4.6,4.7,4.8 show the data used to create the model by day against the day of the year with a plot of prediction for 2050. Figure 4.8 shows the data that was chosen for the final model. Figures 4.6,4.7 show the Chillers data and modelled individually. These plots allow viewing of the data structure relative to the day of the year.

The basic behaviour on display in all plots is the summer cooling period with increased cooling energy use. The more complex behaviour lies in the degree of uncertainty around the median value. The uncertainty around the median has to communicate uncertainty due to weather and uncertainty in the output due to unexplained behaviour by the input data. The input data is $\mathbf{w}$ from section 4.3.2.

In figure 4.7, the data and predictions for Chiller No 2 have a near zero mean for the entire year however the uncertainty increases in the summer representing higher probability of higher energy

Training Data Values Chiller 1

Chiller 1 use Prediction by day in 2050
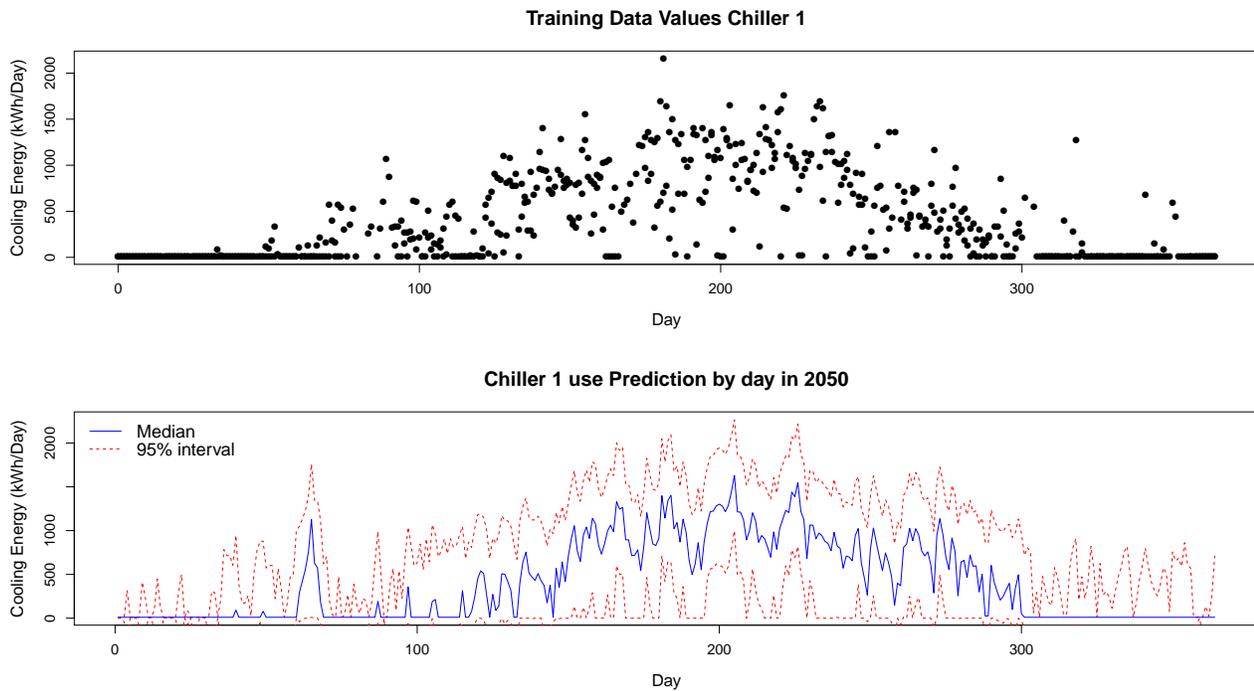
Figure 4.6: Comparison of predicted daily data and actual data: Chiller No 1

Training Data Values Chiller 2

Chiller 2 use Prediction by day in 2050
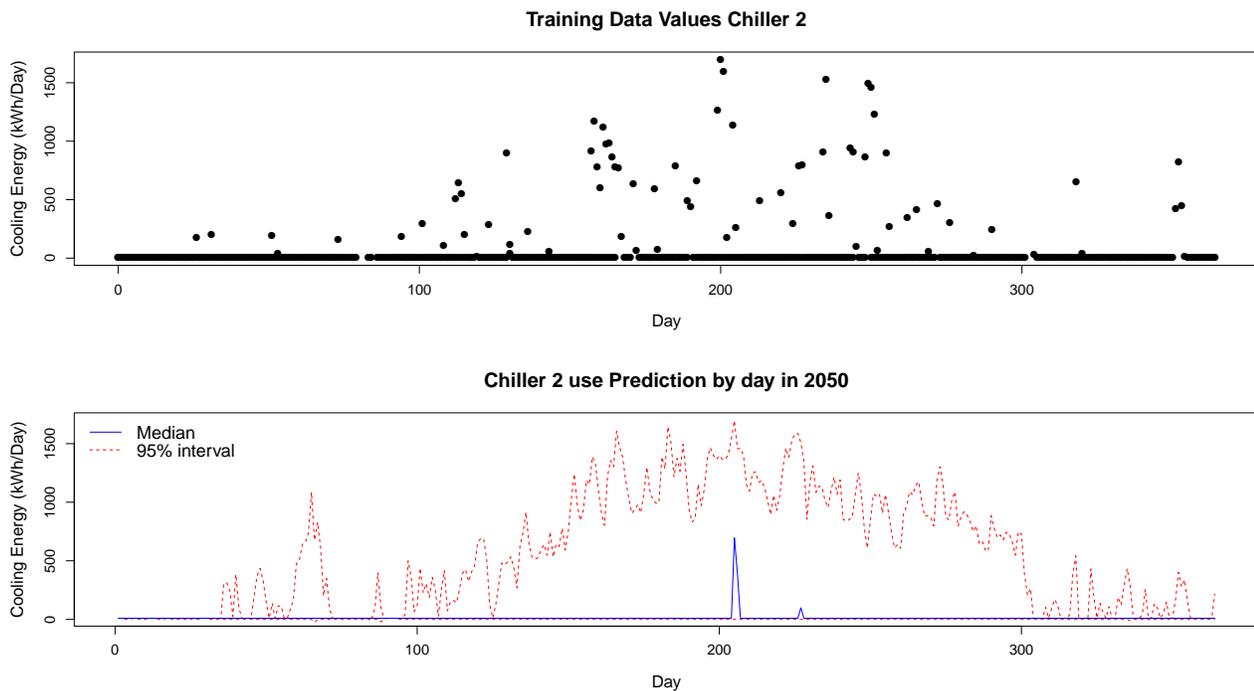
Figure 4.7: Comparison of predicted daily data and actual data: Chiller No 2

use values. This is similar to the training data for Chiller No 2. The training data is largely near zero with more frequent, larger non-zero values in the summer period. This represents a challenge for the model as it never abandons the inactive state and considers it most likely throughout the
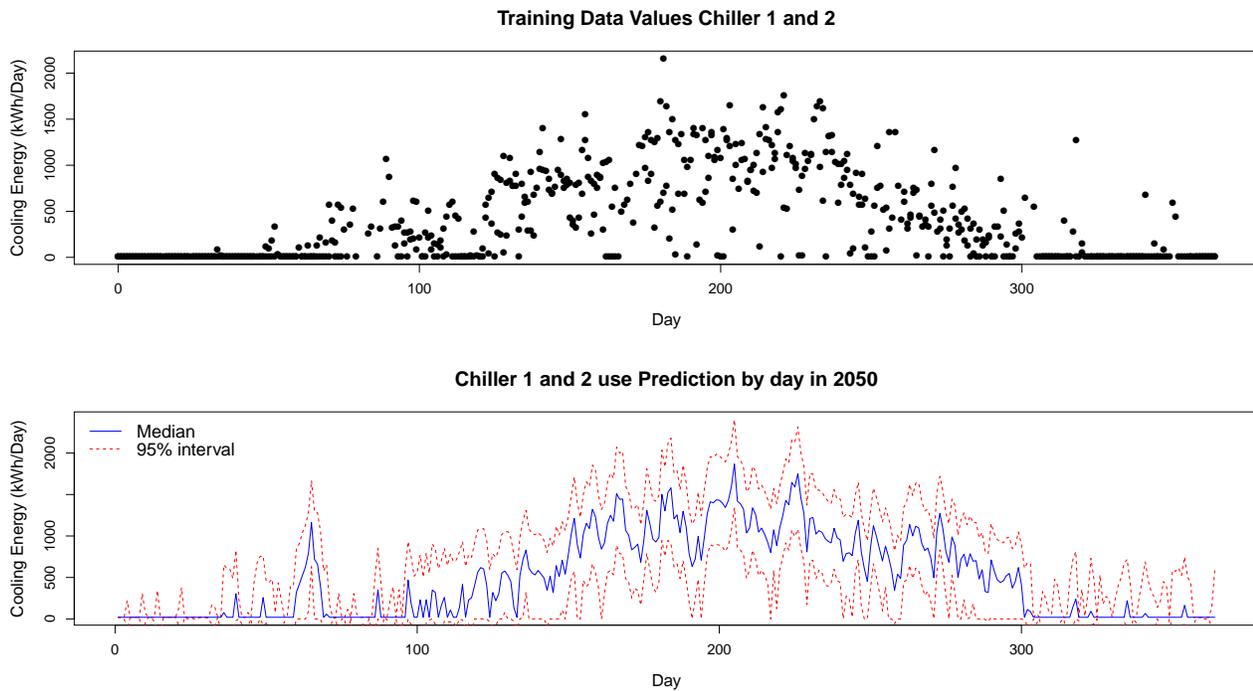
Figure 4.8: Comparison of predicted daily data and actual data: Chillers No 1 & 2 combined

year. The uncertainty remains large due to the lack of clearly identifiable features.

In figure 4.6, the data and predictions for Chiller No 1 has considerably more interesting features identified. Whereas figure 4.7 has only one departure from a near zero mean, figure 4.6 has identified periods with non-zeros throughout the year. These features correspond to conditions in the forecast weather that are similar to the weather in the training data with high energy use (the model has successfully identified 'hot' weather).

The challenge is that the two Chillers do not operate entirely independently. If one is on the other is much less likely to be so. This leads to confusing data where the exterior weather is warm, a Chiller doesn't respond and makes the model unable to decide whether the active or inactive state is a better model for a Chiller and draws inconclusive inference. Whilst this is a sensible conclusion (not drawing clear conclusions from unclear data), clear conclusions are more useful.

Figure 4.8 shows the data and predictions for the combined Chillers. This improves the situation by removing the values where warm weather is met with increased energy use response or vice versa, the dual-behaviour GP is able to adapt to handle the non-stationarity more effectively as result. Uncertainty estimates are significantly lower, indicating that this approach is more successful. There is an important example here in that information may be required from outside the data-set to improve predictions: there is a limit of accuracy that can be achieved with any particular data set.

4.5.2 Wider Applications

The methods used here are quite different to the traditional desktop study, where a model is produced to using the discussed BEM simulators and basic building information such as floor-plans, into climate impact on buildings, such as Jentsch et al. (2013). In the past, similar things

might have been attempted with linear regression and neural networks (see Zhao & Magoulès (2012) sections 2.2 and 2.3). The success will have been limited by constriction of computational speed and memory for statistical methods. They might also have been challenged by the lack of significant available data for this kind of study: significant amounts of automated metering and weather stations were not available 10 years ago. The 2.19 billion samples for this paper's predictions would not have been manageable 20 years ago, sampling approaches of this nature are now possible.

Traditional desktop studies depend heavily on physics based models. These are able to provide flexible models that can be adjusted to describe nearly any possible design proposal.

The cost is that you need to have good detail of the proposed or existing structures. This leads to costly on-site visits and often being challenged to do better than an educated guess as to values in the model, for example installed U-values for construction materials. Even calibrating the building models for these values likely will be subject to identifiability issues (many possible 'good calibrations').

Emulation of building energy use needs to solve a calibration problem. The emulation in this paper doesn't require estimation of specific details of occupancy behaviour, geometry or physical construction of the building in question. This is a quintessentially black box approach, where the inner workings of the approach do not relate to physical elements in reality.

Uncertainty quantification is clearly a major problem with the desktop physical models. This is epitomised by the growth of use of emulation for computer models (Caiado & Goldstein (2015) and Vernon et al. (2010)). There are not many easy ways to accurately specify the uncertainty as to the U-value of a wall. The desktop physical model doesn't include uncertainty explicitly in the model. Attempts to express uncertainty, for example sampling many possible U-values, is limited by again the 'educated' guesses used. Indeed the approach of desktop modelling is a way of avoiding prohibitively expensive physical testing

The presented approach in this chapter is much clearer about the degree of uncertainty (maximum likelihood will penalise overconfident models, subject to over-fitting the training data) and the limitations of the flexibility of the model. In terms of evaluating the errors of the model we can look back at the results in chapter 3 and see that model does perform quite well. There is some evidence of net underestimation of energy use per day though on the whole the fit remains good.

4.6 Conclusions

The building energy modelling in this chapter suggests that climate is likely to cause a noticeable increase in the cooling energy use of the University of Sheffield building: the Information Commons. The main source of uncertainty was found to be from the variation from the weather inputs, this was in addition to substantial uncertainty exists from the Gaussian process modelling of the data. By providing uncertainty analysis of key information such as size and time peak system usage.

The emulation approach, adopted in this chapter, suggests that the Information Commons is likely to require additional cooling as a result of forecast climate change in UK weather patterns. The uncertainty propagation allows inference to be drawn about the distribution of these peak values such as the likely future peak energy use and the times of year this should occur.

The approach outlined presented in this chapter is a technique that can be used to draw conclusions from building energy use data. In terms of working with information across cities in the UK. The heating and cooling demand in the Information Commons in Sheffield are weather

sensitive, some of the other systems such as the HVAC system doesn't. This case study illustrates the benefits doing uncertainty analysis on key building systems. It highlights future risks in the form of the on average Chiller energy use increase and the distribution of future peak load values. This can facilitate decision making under uncertainty allowing building operators to tailor risk of building systems being overloaded against capital spending commitments for example.

4.7 Model Fitness

4.7.1 User Requirements

The users of output of this chapter are those making decisions about how to construct and manage buildings. Those doing parallel work the research in this chapter constructing models to assess weather's impact on building energy use.

The users of the output of modelling building responses to future climate change are:

- To understand how the Information Commons is likely to experience climate change.

- To be able to understand the limitations of this modelling such as the degree of uncertainty in the model output.

The chapter should also address the choice to ensure the predictions are realistic in relation the second recommendation from subsection 3.6.4 to consider log transforming of the data for modelling.

4.7.2 Chapter Output

The output from this chapter has the following outputs that communicate how the Information Commons responds to climate change:

- Predictions of the energy use per annum with confidence intervals. Figure 4.4

- The breakdown of the sources of uncertainty in the model predictions. Figure 4.4

- Predictions of the changes in peak energy use for the system under climate change. Figure 4.5

4.7.3 Evaluation of outputs against user requirements

The focus of chapter 4 compared to chapter 3 is more focused on output for non-subject matter experts (SMEs), at least within the field of BEM. This mean the output is graphic display of information rather than technical recommendation for how to construct models.

The main requirement is open in terms of how it can be met explain the impact of climate change on the Information Commons. The figure 4.4 is not immediately intuitive. The language could be more accessible as 'decomposition of variance' is phrase not used much outside of statistical communities. Talking about the source of uncertainty might be a better way of communicating the concept. There is also a lack of summary statistic to go with the plot such as the average amount energy used increasing. This is partly addressed by figure 4.5, where the amount of the peak energy use is displayed stat current and predicted. However summary statistics contain less information so are easier to communicate and engage the user community.

The uncertainty quantification in this chapter, is one the stronger elements. The graphical summaries do express the variation in the predictions. Exploring the peak energy use was a useful demonstration of the way yearly performance can be uncertain. The uncertainty decomposition can covey the dependence on the lack knowledge about future weather. However these again would be easier to communicate with summary statistics to engage stakeholders.

Figure 4.4 can be used to discuss the accuracy recommendation made in section 3.6.4 that the proportion of the variation contributed by the model and by the weather uncertainty. This plot demonstrates that majority of the uncertainty is contributed by the weather. Suggests that even a perfect model would still have a similar magnitude of uncertainty, suggesting that appropriate amount of variable were include in the model.

The log transform recommendation in section 3.6.4 for consideration was rejected at on the basis of prioritising model accuracy. This was probably poorly conceived, as the model retains the issues with the negative cooling energy use of the previous chapter which would be likely be an issue for establishing stakeholder confidence.

4.7.4 Recommendations

The main recommendations are as follows:

- That future work should have summary statistics as key results to present these will make it easier to communicate to stakeholders. This highlights that over dependence graphical information displays is to be avoided.

- That when considering the modelling decisions that involve trading off accuracy against realistic world features, that realism should be prioritised. This could help assure stakeholders of the model accuracy.

Acknowledgements

Chapter 5

Significance of roughness length on wind speed over Birmingham

Air movement in urban spaces plays a vital role in pedestrian comfort and in building energy demand, ventilation and air quality. This chapter will not address pedestrian comfort directly, as empirical validation of thermal comfort is complex topic relating to both indoor and outdoor thermal comfort: Haldi & Robinson (2010) & Taleghani et al. (2019) The complexity of urban geometry makes prediction of air movement in built-up areas highly challenging. In physical wind tunnel experiments and computational environmental modelling, the construct of (aerodynamic) roughness length has been proposed as one of the parameters for predicting wind speed in a given urban area. Within the modelling of the urban fabric the boundary layer is defined to be highly variable depending on urban geometry, Pisello et al. (2015) & Kanda et al. (2013). However, the significance of roughness length in predicting near-ground urban wind speed has yet to be evaluated with large-scale observational data.

I This thesis presents a statistical methodology that allows for examination of the significance of roughness length on urban near-ground wind speed prediction with real world observational data. This methodology has been developed and tested using the Birmingham Urban Climate Laboratory (BUCL) data-set and the UK Environment Agency Geomatics Survey (LIDAR) data-set over the city of Birmingham.

In this chapter, it is I first demonstrated how frontal area index (λ_F) and plan area index (λ_P) can be calculated using pixels in the LIDAR data-sets. WeThis chapter then introduces a new likelihood based model assessment (Gaussian processes) method for modelling non-linear responses with significant covariances. In order to evaluate the significance of roughness length in predicting wind speed within the area covered by the BUCL measurements, IThis chapter useds both linear models and Gaussian process models. The results show that whilst there is some linear correlation between wind speed and roughness, linear modelling here indicates too much confidence in it's own inference. In contrast, theoretical work on the topic suggests roughness length is highly important for wind speeds.

In conclusion, ~~my work suggests~~ this chapter supports the current understanding that the gap between our theoretical understanding of the drivers of wind speed and empirical observations remains large. This relates to the wider discussion of the thesis of model fitness and weather/urban fabric interactions suggesting that neither the models provided in this chapter or to a less extent work in the wider community are well suited.

Keywords— Uncertainty Quantification - Urban Climate - Wind Speed - Roughness Length - LIDAR - Gaussian Processes - Sensitivity Analysis

Urban morphology summarises the complex layout of cities, often used for the purpose of modelling air movement such as wind speed Mo & Liu (2018), CO_2 levels Makido et al. (2012), particulates Li et al. (2017) and wind energy estimation Wang et al. (2017), Lukač et al. (2017). In this chapter, empirical evidence is used to evaluate the strength and significance of the impact ~~of~~on roughness length on wind speed in Birmingham. This chapter uses statistical and machine learning techniques designed to assess the sensitivity of complex systems (Strong & Oakley 2013).

This approach is necessitated by wind speed being affected by many factors (for larger scales see Rose & Apt (2016)), such as local terrain, turbulence effects and weather at higher altitudes Lee (1979), on which data is unavailable which affect the wind speed over cities. This chapter supports the current understanding that roughness length negatively correlates with wind speed at low altitude, whilst handling the poorly understood behaviour of wind speed observations which does not relate to large scale roughness lengths.

This chapter will work with these low level wind speed observations and roughness lengths as a potential indicators on the basis that the theoretically boundary layer thickness should contain significant information as to the wind speed at a location, on a smaller scale, in this chapter of a kilometre, than Kent et al. (2018a). Two methods for evaluating the significance of roughness length will be compared: linear models as a way of providing significance tests, and a more robust method, additive modelling, that has potential ability to distinguish between false associations with better quantified certainty. As will be discussed later, linear models in this context have a tendency to be overly confident about significance.

Summaries of urban morphology, such as roughness length which is the focus of this chapter, need to be able to contrast open farmland from suburbs from city centres. Other phenomena such as nearby air flow off hills and drought/heavy rainfall affecting transpiration rates will affect this local wind speed directly or indirectly. This chapter aims to empirically evaluate the benefits of this information, in contrast with other papers such as Mo & Liu (2018), Wang & Srinivasan (2017), Vonlanthen et al. (2017) which base their analysis on ~~fluid dynamics~~CFD simulations.

Current physics based summaries of urban morphology such as Adolphe (2001) and more recently Wang & Srinivasan (2017) make strong use of our physical understanding of the surrounding world to find ventilation corridors. They use concepts such as known heights of buildings and street width to estimate average sky view factor, roughness length or relative rugosity.

The concepts absolute and relative rugosity, within the literature of micro-climate modelling, are based on assumed discrete categorisations between buildings and other objects and that such classifications, which may or may not be the case. Absolute rugosity is directly correlated to the average built height of an area. Relative rugosity takes into account the effective boundary layer. Neither measure can distinguish between a few large buildings with large gaps and many small buildings of the same height and building footprint. The distinction does matter as air speed close to the ground comprises a boundary layer and therefore a viscous flow Young (1989). Smaller buildings with the same footprint and height as larger buildings will result in greater drag on the air flow leading to lower velocities closer to the ground. Our approach makes use of the LIDAR height map without attempting terrain classification.

These other methodologies are typified by work such as Moussa (2010), Mahmoudabadi et al. (2016), Höfle & Hollaus (2010). These works cover techniques such as classifiers, filters and clustering. These techniques treat the knowledge of what is and isn't a building as unknown and/or the target of the learning exercise. This area of image processing has valuable applications such as producing information for services such Google maps company. (n.d.), as exemplified in Swarup S. Medasani, (Thousand Oaks et al. (2013). This chapter follows a similar approach to Dougherty &

Lotufo (2003) though the terminology is somewhat different.

Figure 5.1: Diagram with data flow plan for BUCL/LIDAR height map data
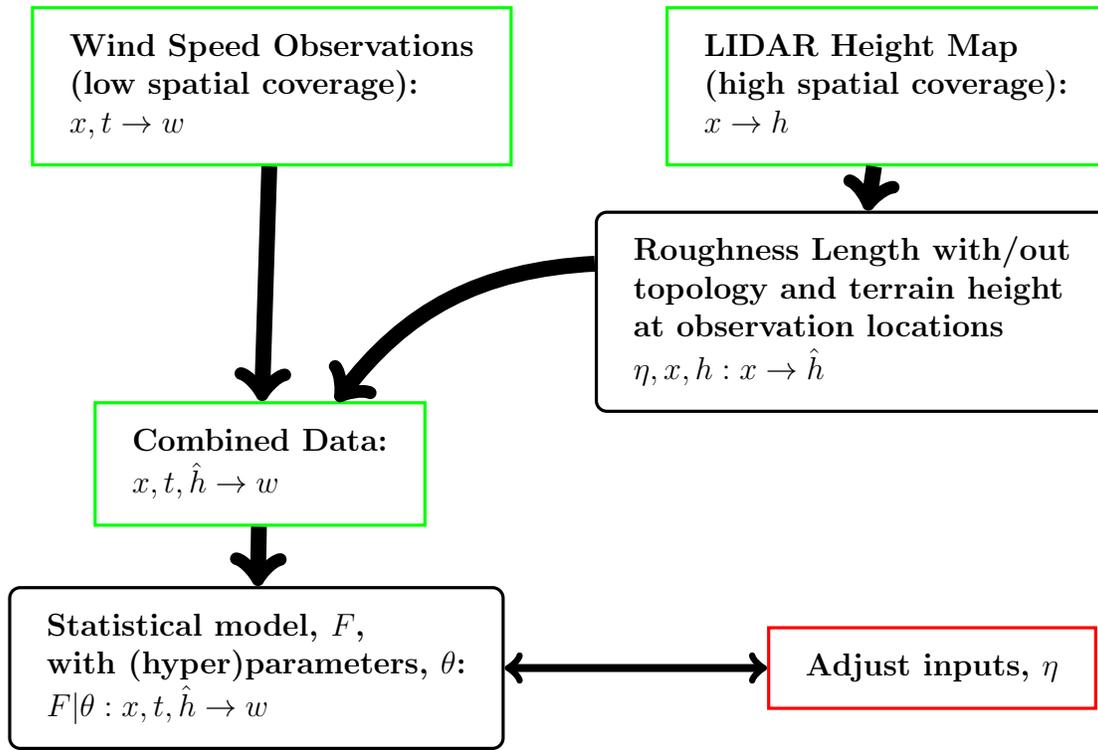
Figure 5.1 provides an overview of data use in this chapter.

- Two data-sets are used in this study, the Birmingham University Climate Laboratory data-set Warren, Young, Chapman, Muller, Grimmond & Cai (2016) wind speeds and the height maps from the Environment Agency Geomatic Survey Data: LIDAR composite DSM U.K._Environment_Agenc (n.d.).

- The LIDAR data is reprocessed to calculate average terrain height and roughness length (termed $\hat{h}$ in figure 5.1).

- This is then combined with the wind speed observation to form a data-set on which to test the significance of roughness length.

- A statistical model of the data is then built and fitted to the data. Summaries of this model are then used to assess the goodness of fit to see if the roughness length improves the goodness of fit of the model.

The purpose of this chapter is to demonstrate two techniques. The first, is the adjustment to calculation of roughness length using a Gaussian filter to define the area of relevance opposed to various catchment techniques used in other approaches such as Guan et al. (2015), the second is the use of non parametric methods (in this case a Gaussian Process) to model the output.

5.1   Data-sets

5.1.1   Birmingham University Climate Laboratory (BUCL)

This study makes use of the Birmingham University Climate Laboratory (BUCL) data-set Warren, Young, Chapman, Muller, Grimmond & Cai (2016), which covers most of Birmingham. There are 26 active weather stations ~~over a period of~~ over a period of 6 months, this time period was selected on the basis on when most sensors were active between 2012 and 2014. A more generic discussion of the BUCL data-set can be found in appendix A including a plot of the location of the weather stations.

Of particular pertinence to this chapter is the behaviour of the wind speed at different locations. Figure 5.2 shows 4 sensors, with GoogleMapscompany. (n.d.) captures of the surrounding locations (industrial, inner city, farm land and sub-urban) and a plot of their corresponding wind speed over 3 days. The selection of the types of example location cover the sub-types of urban degree examined in Pisello et al. (2015). The plots illustrate the minute by minute measurements and whilst the broad behaviour of the wind speed is similar, there are noticeable differences between sensors, it is this behaviour that this chapter aims to investigate.

5.1.2   Environment Agency Geomatic Survey Data: LIDAR composite DSM

The LIDAR data used in this study was sourced from the web archive of the UK government Environment Agency Geomatic Survey Data U.K._Environment_Agency (n.d.). The composite Dynamic Surface Model (composite DSM) at 1 m resolution was selected on the basis of providing good coverage across the study region. Example plots of this data-set can be seen in figures 5.9a, 5.9b and 5.8, with the data-set covering the vast majority of Birmingham whilst providing high resolution information of the areas of interest.

Figures 5.9 and 5.8 highlight the pale grey areas where there is missing data, possibly from environnmental issues during data collection.

5.2   Methodology

5.2.1   Morphology Definitions

This section provides detail on the calculation of the numerical approximation equivalent objects for computation and how this links to a smooth surface. There are dedicated libraries to this type of machine learning in most high level programming ~~languages~~packages such as numpy.convolve.

5.2.1.1   Introduction to Morphology as a differentiable surface and terminology

A height map, I, can described in the following way:

$$z = I(x,y) \quad \text{where} \quad (x \in [x_{min}, x_{max}], \ y \in [y_{min}, y_{max}]) \tag{5.1}$$
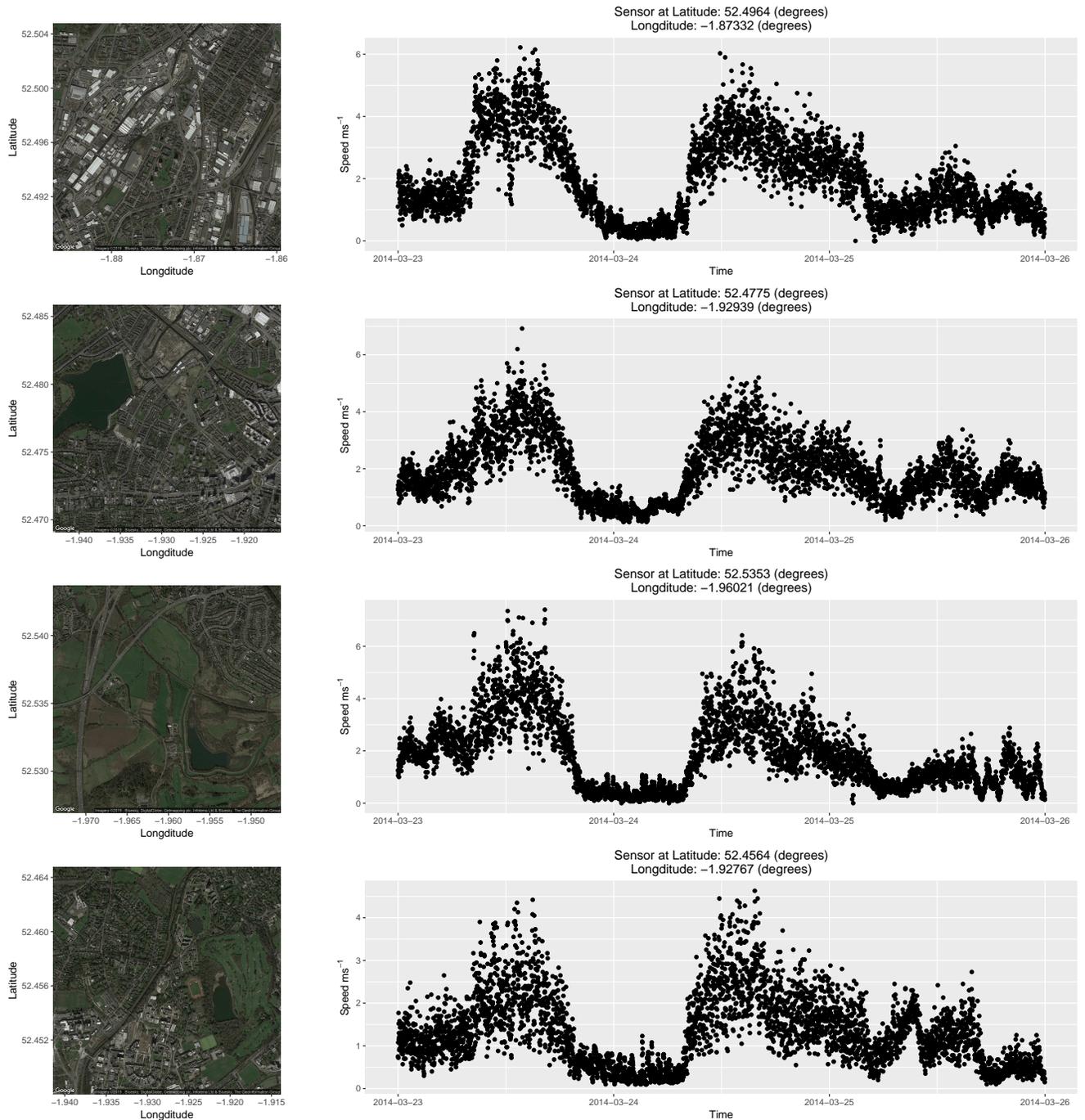
Figure 5.2: Plots of 4 example locations of sensor in the BUCL data-set Warren, Young, Chapman, Muller, Grimmond & Cai (2016) and 3 example days of wind speed at these 3 locations. Aerial photography: Google Maps company. (n.d.). The methodology for placement of sensors was subject the hosting organisations, only the placement methodology and height above sea level were provided in the data-set.

where x, y represent horizontal coordinates and $[x_{min}, x_{max}]$, $[y_{min}, y_{max}]$ their respective limits. ~~It is the goal of this chapter to provide summaries of the LIDAR surface.~~ Basic summaries can be created by simply taking the mean of the surface height and the standard deviation of the surface similarly.

The mean can be written:

$$\text{mean}(I) = \frac{\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} I(x,y) \ dy \ dx}{\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} 1 \ dy \ dx} \tag{5.2}$$

For an approximation to this function, ~~I~~the author generated a look up table of known values. In practice the function is only evaluated at a finite number of locations. In our case the grid of x_i and y_i values is spaced at 1m intervals. The mean for f are then be approximated:

$$\text{mean}(I) \approx \frac{\sum_{x_i=x_1}^{x_n} \sum_{y_i=y_1}^{y_n} I(x_i, y_i)}{n^2} \tag{5.3}$$

where n is the total number of pixels.

This chapter will require the definition of the partial derivatives of the surface. These represent the gradient of the surfaceor the frontal area of the building present to oncoming wind. This chapter will use the notation $\partial_x I(x,y)$ to denote the partial derivative of I respect to x at x,y, the approximations are calculated as follows:

$$\begin{aligned}
\partial_x I(x_i, y_i) &\approx \frac{I(x_{i+1}, y_i) - f(x_i, y_i)}{2} \\
\partial_y I(x_i, y_i) &\approx \frac{I(x_i, y_{i+1}) - I(x_i, y_i)}{2}
\end{aligned} \tag{5.4}$$

The last element required is the concept of a weighted integral or filter. This chapter only makes use of a single measure in the form of a Gaussian filter or radial basis function. The mean calculated using the Gaussian filter gives most weight to values close to the centre of the Gaussian density and very little, but non zero, weight to those farther away. Let $\text{mean}_\sigma(x^*, y^*)(I)$ be the mean of f using a Gaussian filter centred at (x^*, y^*) defined:

$$\begin{aligned}
\text{mean}_\sigma(x^*, y^*)(I) &= \frac{\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} I(x,y) \times \rho((x,y),(x^*,y^*)) \ dy \ dx}{\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} \rho((x,y),(x^*,y^*)) \ dy \ dx} \\
\text{where} \quad \rho((x,y),(x^*,y^*)) &= \exp\left(-\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x^* \\ y^* \end{bmatrix} \right)^T \Sigma^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x^* \\ y^* \end{bmatrix} \right) \right)
\end{aligned} \tag{5.5}$$

$$\text{where} \quad \Sigma^{-1} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \tag{5.6}$$

The $\text{mean}_\sigma(\cdot, \cdot)(\cdot)$ is a isotropic filter, C_ρ, with kernel, $\rho(\cdot, \cdot)$:$\rho(x,y)$ is a weighting, filter or $\frac{\rho}{\int \int \rho \ dx \ dy}$ is the measure, typically the scaling parameters $\{\sigma_1, \sigma_2\}$ will be set on the basis of how far you want to include information, here it can be assumed that $\sigma_1 = \sigma_2$. There is a broad selection of choices that could be made here, and a detailed discussion of such processes is beyond the scope of this chapter, an example of the literature: Egmont-petersen et al. (2002).

5.2.1.2   High dimensional techniques and images

Figure 5.3: Height map as grey-scale image, key on the right indicates height above sea-level.
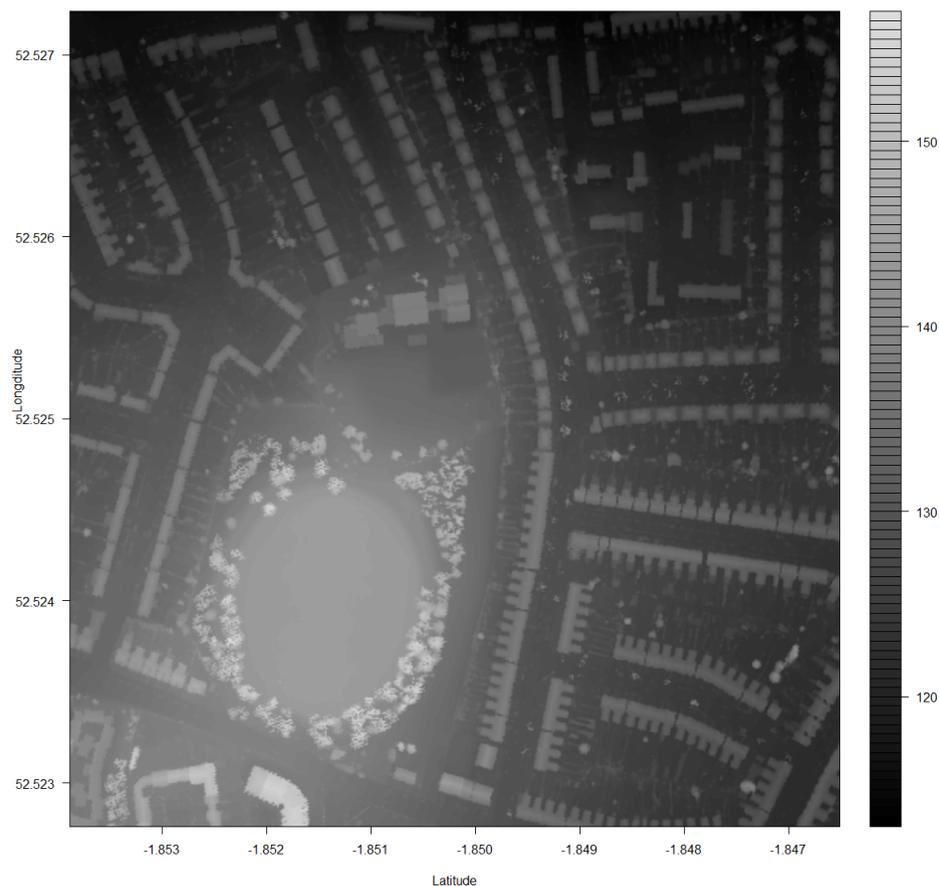
The techniques used here for image processing have computational cost to the number of pixels, compared to some statistical dimensional reduction techniques such as principal component analysis which are much more costly. The data can be visualised as a grey-scale image (fig. 5.3).

Image processing often makes heavy use of convolution methods (in this case using a Gaussian filter) to identify patterns in images. Edge detection is largely about finding the locations where the image suddenly changes. This is done by performing analysis on the gradients of the filtered image. ~~Street direction can be thought to be the average direction of the edges.~~

The footprint of buildings operates on the scale of around 5-10 meters. Hills' length scales, σ, are roughly around 5+ times larger than cars and small trees. As such the first operation performed is to separate the height map image into different length scales (see figure 5.4).

To do this a Gaussian filter with a length scale of 30 meters is used to identify large scale features (figure 5.4b) in the image then subtracted element-wise from the image. This process is ~~then~~ then repeated with another Gaussian filter at length scale of 5 meters to produce features that are of correct length scale to be buildings (figure 5.4c). It should be noted that ~~at either feature level features are not exclusive~~multiple features could impact wind speed: large scale features take into account building clusters and building scale features include as 'small' bumps in the ground. From the prospect of atmospheric modelling this is not likely to be problematic as wind of a large scale is obstructed by clusters of buildings in the same way as hills. See fig. 5.5, the chapter's methodology uses the images from the central column.

Figure 5.4: Cross-sections through different smoothings LIDAR Data
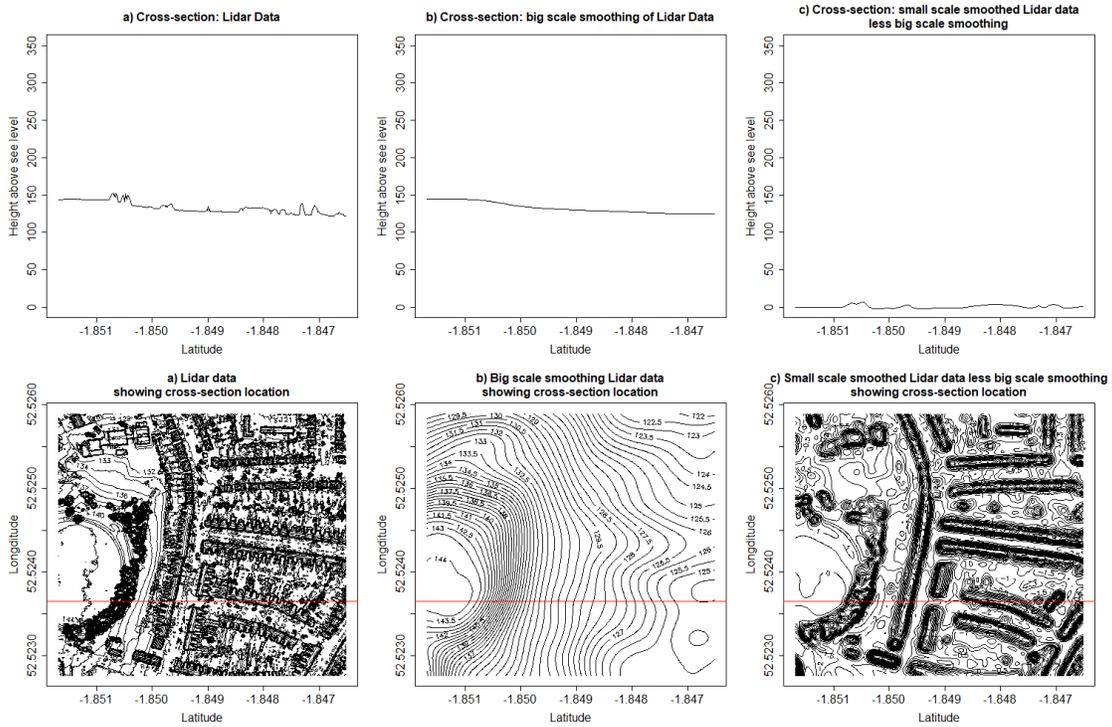
More formally if there is a LIDAR height map, I, with values on the grid with x or latitude values at $x_1, ..., x_n$ and y or longitude values at $y_1, ..., y_n$ spaced at 1 meter intervals. Or:

$$I_{i,j} = I(x_i, y_j) \tag{5.7}$$

So the larger scale features, I^L, are found:

$$I_{i,j}^L = \text{mean}_{30}(x_i, y_i)(I) \quad \forall \; i, j \in \{1, ..., n\} \tag{5.8}$$

The building features, I^B, are found:

$$I_{i,j}^B = \text{mean}_5(x_i, y_i)(I - I^L) \quad \forall \; i, j \in \{1, ..., n\} \tag{5.9}$$

5.2.1.3 Calculating roughness length by pixel

Up to this point the traditional roughness length which calculated such as Wicht & Wicht (2018) and Bottema (1997), ~~are based on well-studied~~representing the established theory. They are also calculated on concretely defined catchments of pixels, termed 'rugoxels' in Bottema (1997). The calculation of these 'rugoxels' is where some methods diverge in terms of how these catchments are defined, others make different assumptions in the underlying approximations of fluid flow over a rough surface. We also take into account vegetation and buildings equally in a departure from other methods to take advantage of the high resolution of LIDAR data. This leads to higher roughness length values than might typically seen in these studies.

The work in this chapter outlines an alternative approach that defines a blurred catchment around an arbitrary point. So calculating this roughness length and zero-displacement height,

Figure 5.5: Contour plot of Sensor W20 with the roughness length for wind directions with/without topology, 0 radians corresponds to North
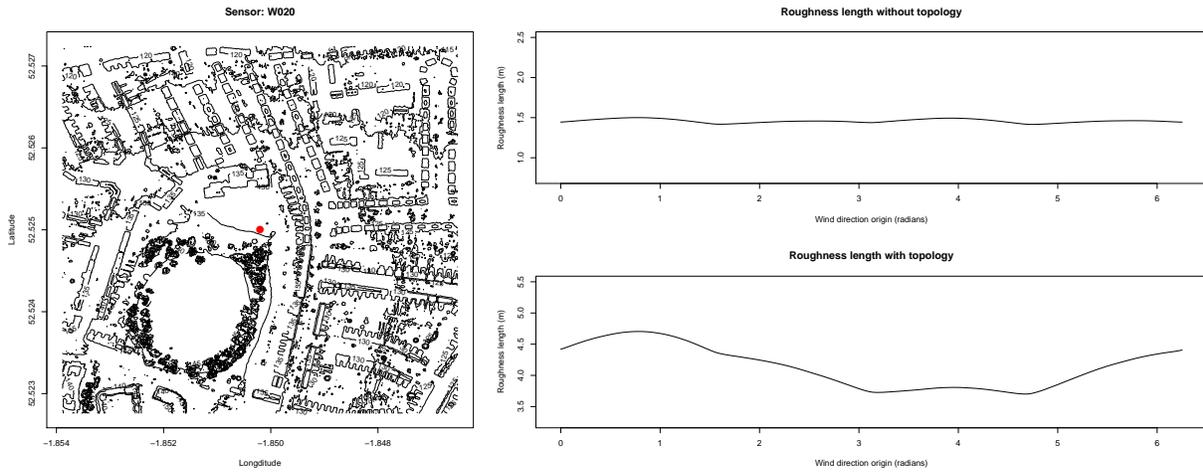
example formulas defined below, as termed in (Wicht & Wicht 2018) with interest to see if the urban corridors can be found in the Birmingham data set.

The definition of, z_0, roughness length or the 'scale' of the surface roughness from (Wicht & Wicht 2018), is defined:

$$z_0 = (h - z_d)exp\left(-\frac{\kappa}{\sqrt{0.5C_D\lambda_F(\mathbf{w})}}\right) \tag{5.10}$$

where h, is the volumetrically averaged building height, z_d the zero plane displacement height, κ the von Karman's constant (0.4, (Högström 1996)), C_D approximate drag coefficient (more discussion 0.2 for veg, 0.8 for buildings), $\lambda_F(\mathbf{w})$ frontal area index.

The zero-plane displacement height, z_d, the height at which the average wind speed should be zero is:

$$z_d = h(\lambda_P)^{0.6} \tag{5.11}$$

where λ_P is the area plan ratio.

Neither of the above formulae need adjusting for the work done here, however in calculating h, $\lambda_F(\mathbf{w})$ and λ_P some adjustments are necessary. Figure 5.6 illustrates how these calculations are carried out.

- For h each pixel represents a 1×1 meter square meaning volumetrically average height reduces to:

$$h = \frac{\int^{\text{all pixels}} I_{i,j}^2}{\int^{\text{all pixels}} I_{i,j}} \tag{5.12}$$

where f is the height above the minimum pixel value (or height) in the image less the large scale smoothing, which would remove the effect of topological height change on the data.

Figure 5.6: Plots demonstrating the calculation of frontal area of a building using pixels (each representing a 1×1 meter square area
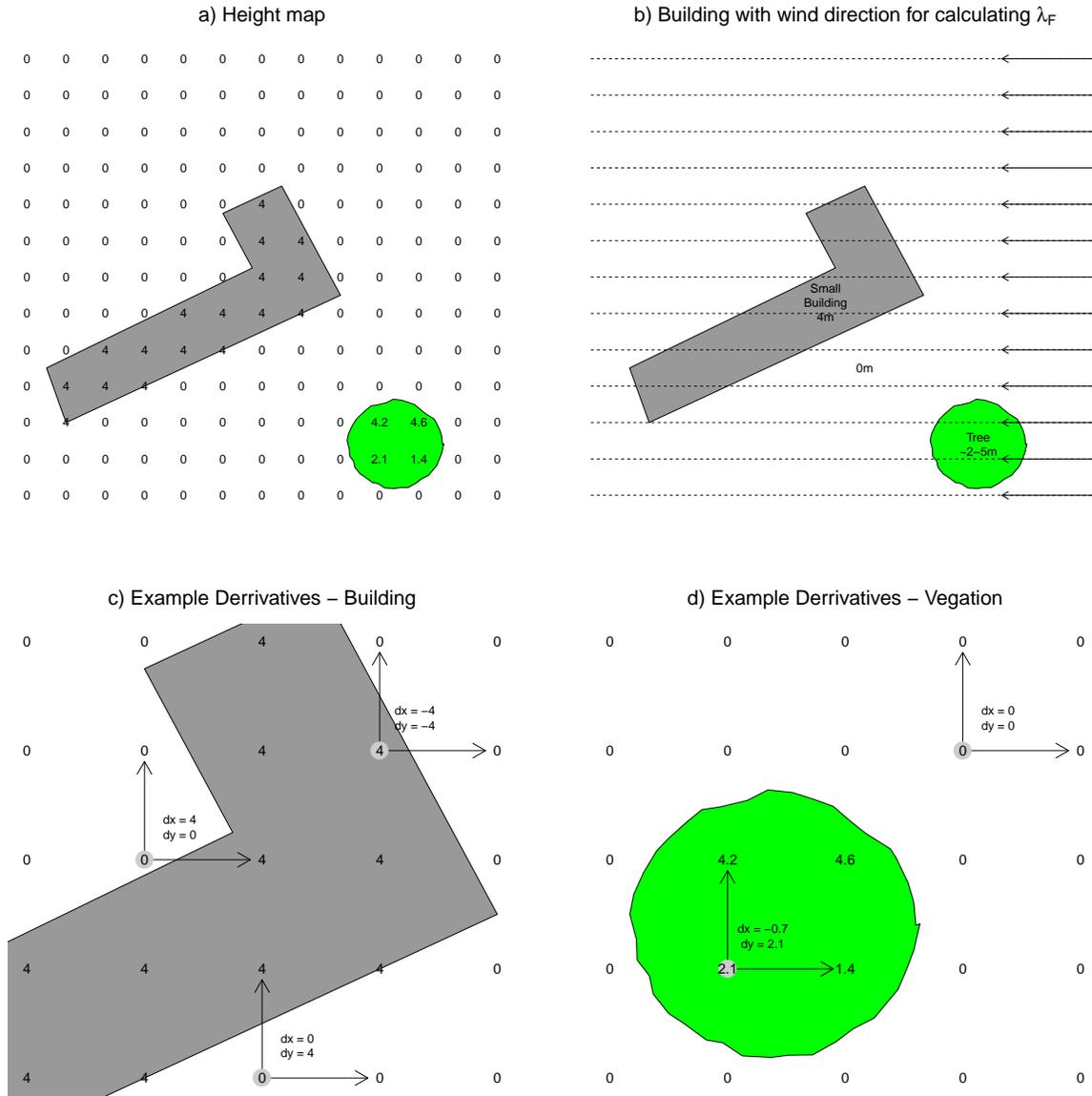
- For $\lambda_F(\mathbf{w})$, the frontal area index which depends on some wind direction, $\mathbf{w}$:

$$\lambda_F(\mathbf{w}) = \frac{\int^{\text{all pixels}} \max\left(\mathbf{w} \cdot \begin{bmatrix} \partial_x \\ \partial_y \end{bmatrix} I_{i,j}, \ 0\right)}{\int^{\text{all pixels}} 1} \tag{5.13}$$

$$\text{where } \mathbf{w} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ such that } |\mathbf{w}| = 1$$

- For λ_P, the area plan ratio, this chapter will consider vegetation and other obstructions as part of the 'plan area' as they obstruct air flow. Having defined f earlier to have large scale

feature removed hills and as such are of less concern.

$$\lambda_P = \frac{\int^{\text{all pixels}} \begin{cases} 1 & \text{where } I_{i,j} > t_P \\ 0 & \text{where } I_{i,j} < t_P \end{cases}}{\int^{\text{all pixels}} 1} \tag{5.14}$$

where t_P is a threshold used to decide if a pixel is 'plan area' or not.

By $\int^{\text{all pixels}}(*)$ in the context of this chapter, $\text{mean}_\sigma(\mathbf{x})$ is used as defined in equation 5.5. This defines an area around $\mathbf{x}$ and how to sum and the values.

This means for a given $\mathbf{x}$, $\mathbf{w}$ and σ, the method presented above allows us to calculate z_0 and z_d. Arbitrarily setting $\sigma = 30$, $\mathbf{x}$ will correspond to the latitude and longitude of each sensor and setting $\mathbf{w}$ to the prevailing wind across the city at a moment in time. This gives values that are relevant to the wind conditions at any one moment in time.

5.2.2 Statistical methodology - likelihood based model assessment for non-linear responses

Wind is highly variable in terms of its response patterns near rough surfaces, urban landscape being a prime example. It is desirable to have a statistical model that can capture as much of this behaviour as possible. In order to carry out fitting a model to wind data, 3 main tools/concepts have been used: Gaussian processes; handling expectation of stationary behaviour across space (expecting more than the Spatial location to matter for prediction); cross validation of model selection.

5.2.2.1 Gaussian Processes

In this chapter the key responses are parameterisations of wind across Birmingham. Gaussian processes are used to directly model these responses. Gaussian processes are used for many spatial interpolation problems such as handling galaxy formation simulations (Vernon et al. 2010) to infectious disease modelling (Andrianakis et al. 2015).

These models produce probabilistic predictions for for spatial interpolations of wind speed. An example interpolation is provided in figure 5.7. This chapter aims to produce predictions for wind speed between sensors whilst quantifying the accuracy of these predictions. The inherent probabilistic predictions produced by Gaussian processes provide a way of rating the accuracy of the modelling. A recommended textbook on Gaussian processes and their applications is Rasmussen et al. (2006).

Intuitively it would be expected that locations that are in a sense similar, will have similar local weather conditions. This similarity will obviously include physical location, two locations a short distance apart are more likely to be similar than two far apart, time and similarity between types of area, two suburban areas for example might well respond similarly.

Gaussian processes work on the basis of this sort of reasoning, like places have similar values. Gaussian processes are defined as multivariate normal distributions with covariance matrices defined to encode spatial association. The posterior distribution of a Gaussian process, which is a multivariate normal distribution, written in terms of the observations $\mathbf{y}$, X can be written at $\mathbf{x}^*$ the locations to evaluate the predictions of the Gaussian process:

Figure 5.7: Plot of wind velocity interpolation across time, for illustration at single time point. Arrow size indicates wind speed. Blue arrows are data points at this point in time, red are 95% confidence intervals on direction, green are the mean wind direction.
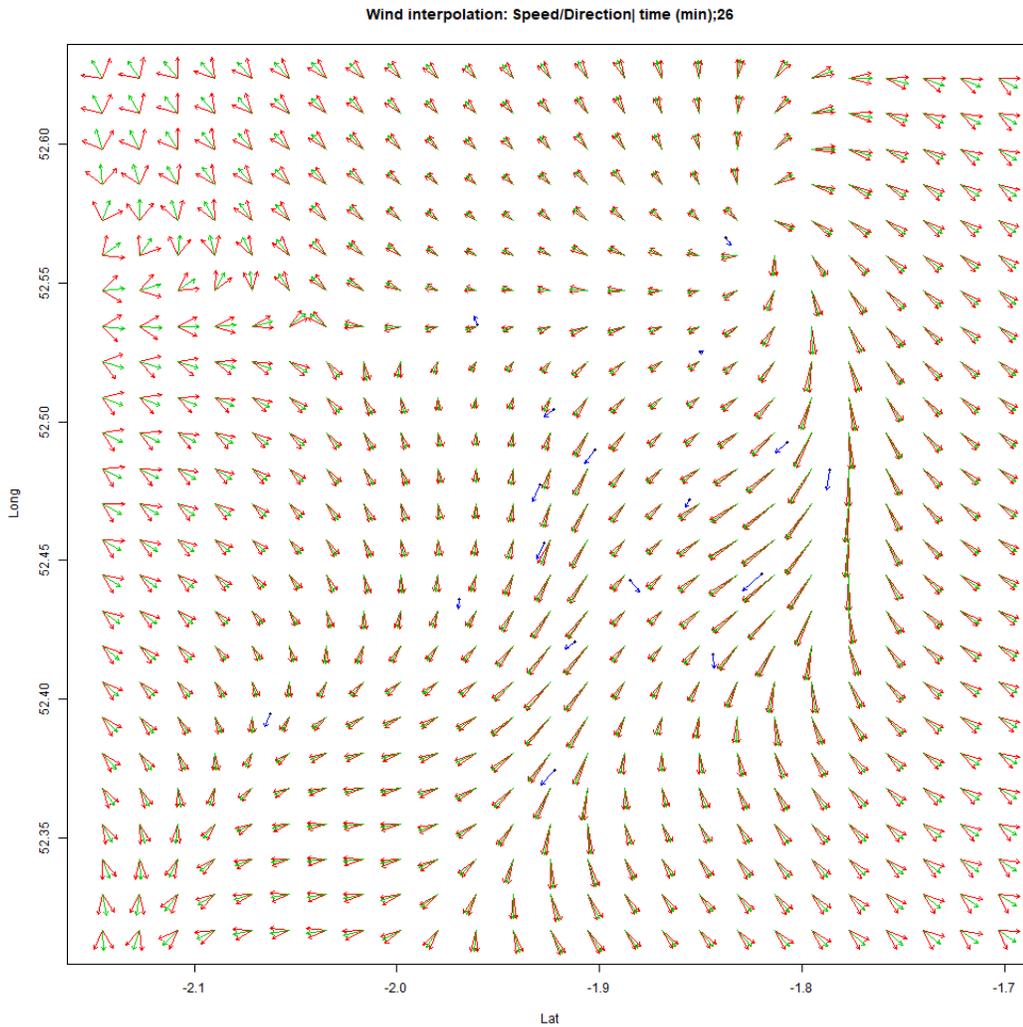
$$
\begin{aligned}
\text{mean}(\mathbf{x}^*) &= K(\mathbf{x}^*, X)K(X, X)^{-1}(\mathbf{y} - \mu(X)) + \mu(\mathbf{x}^*) \\
\text{cov}(\mathbf{x}^*) &= K(\mathbf{x}^*, \mathbf{x}^*) - K(\mathbf{x}^*, X)K(X, X)^{-1}K(X, \mathbf{x}^*)
\end{aligned}
\tag{5.15}
$$

where $\text{mean}(\mathbf{x}^*)$ is the posterior mean at the locations $\mathbf{x}^*$, $K(\cdot, \cdot)$ is the covariance function associated with the Gaussian process, X is the matrix of the locations of the observations, $\mathbf{y}$ are the values of the observations, $\mu(\mathbf{x}^*)$ is prior mean function of the Gaussian process at the locations $\mathbf{x}^*$ and $\text{cov}(\mathbf{x}^*)$ is the covariance matrix between prediction locations $\mathbf{x}^*$.

In this case the latitude and longitude, the spatial coordinates, along with time are thought to provide substantial information for providing these interpolations of wind speed and direction. However theoretical intuition provides strong evidence to suggest that physically derived values such as roughness length Wicht & Wicht (2018), deriving the extent of which is a focus of this chapter. To this end additional information based on the morphology can be provided as inputs to allow Gaussian processes to improve their predictive power. Section 5.2.1.3 details how this chapter processes LIDAR information to evaluate roughness length (figure 5.6 provides an overview of the

data).

The basic formulation of the Gaussian process interpolation problem is that there should be an interpolation in space and time of the wind speed and direction. From this perspective, roughness length is auxiliary in that it there to assist the process rather than being an item interest in and of itself. As such it should be only included if it improves predictive power. It can allow a Gaussian process to do this by having similar roughness lengths for locations further apart, giving it the ability to associate two flat hill tops of opposite sides of a valley.

5.2.2.2 Model Definition

The Birmingham University Climate Laboratory (BUCL, Warren, Young, Chapman, Muller, Grimmond & Cai (2016)) data-set contains minute by minute readings of weather information, including wind speed direction from the North, air temperature, humidity and solar radiation at around 25 weather stations. At any one moment around 20 were active. This work uses short windows in this data. Results are presented for a 40 minute window of the wind speed and direction using every observation in the window.

Each window has the whole procedure carried out on it independently of the other windows. Preliminary work identified that at different time points different Gaussian process model hyper-parameters were required. This suggests the data heterogeneous in time. The focus of this study is the strength of the association between roughness and prediction of the wind parameters.

These windows are selected at random from a time period from the 1st of August 2013 to 20th December 2013. This period was selected on the basis of being the height of maximum activity of the BUCL sensor network so maximising the spatial coverage for the purposes of this work.

Wind speed ($\mathcal{S}$) and direction (θ) were modelled without change of coordinate system. Speed was log transformed to the $[-\infty, +\infty]$ domain which is then modelled more suitably by a Gaussian process. Direction (evaluated in radians), however, remained problematic and the results are not presented in this chapter. With GP representing a Gaussian process, the model can be expressed:

$$\begin{aligned} \theta &\sim GP \qquad (\text{modulo } 2\pi) \\ \mathcal{S} &\sim \log(GP) \end{aligned} \tag{5.16}$$

The Gaussian processes used in this chapter have Mattérn $\frac{5}{2}$ covariance function and a constant mean prior. Matttérn covaraince functions are often used with real world data as they handle the less perfect relationships better than the infinitely differentiable squared exponential covariance function. For each time window the roughness parameter is re-calculated for each LIDAR tile to correspond to the average wind direction for the corresponding time window. Figures 5.9 show maps of Birmingham overlaid with the roughness length with and without the topology with the wind coming from the South. For comparison a map of the average height of each of the LIDAR tile in figure 5.8

5.2.2.3 Model Selection: Cross Validation and Additive modelling

Statistical models are capable of over-fitting (over learning) information in the data, in the sense learning details particular to the data-set rather than features that could be generalised beyond the data set. Gaussian processes are no exception. Considering adding variables to the

Figure 5.8: Map of Birmingham overlaid with average height above sea level (m) of each tile, dark grey areas represent a lack data in the LIDAR data-set
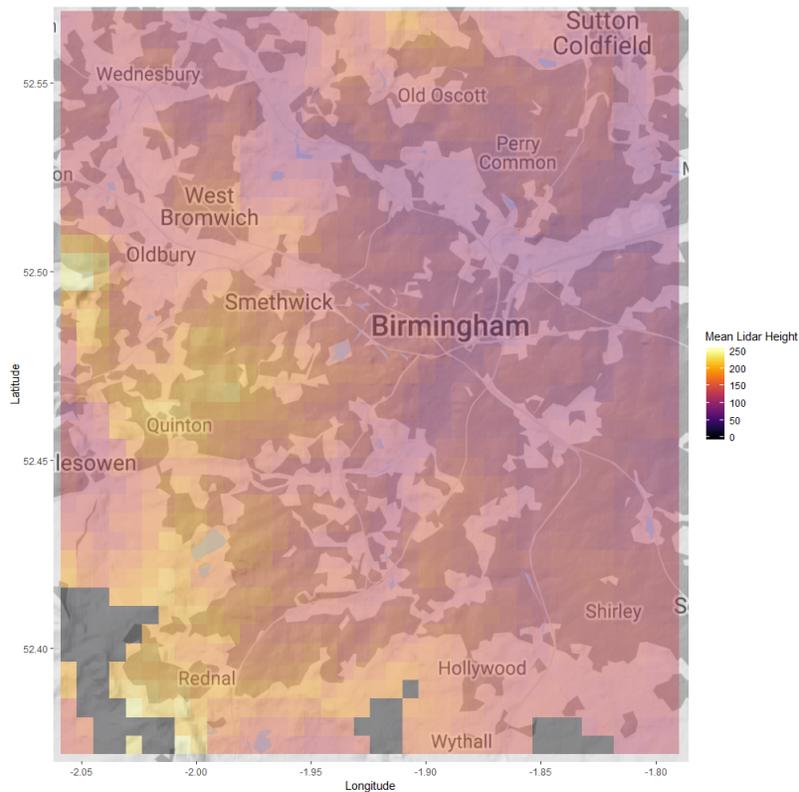
predictive process such as roughness length, means that an observer would like to know if this additional information does actually improve the model, rather than cause overfitting.

In order to consider alternative models within a quasi-Bayesian framework this chapter adopts predictive likelihoods as a form of cross validation. The posterior of a Gaussian process has a multivariate normal distribution across the location of the predictions (X_*). For a Gaussian process there is a shortcut that can be used to provide a leave one out cross validation for the predictions from the posterior that can be found on pages 116 to 118 Rasmussen et al. (2006).

For our purposes this isn't as much help as might be hoped for as the data lies in 'stacks' at each of the sensor locations. Removing one reading from one of these stacks is relatively uninformative as there are always other readings in close proximity making very limited changes in the available information meaning predictions of left out data are still very good.

In the case of the BUCL data-set, observations belonging to sensors, the obvious 'left out' group is the observations sensor by sensor. Essentially clustering the data then leaving out clusters in turn. In this case the clusters correspond to sensors. In this chapter the term leave class out (LCO) will be used to describe this approach. This approach will mean that the likelihood is robust to features that are common to all sensors.

Here a class class consists of a list of indices corresponding to observations in the data-set. So following the format of section 5.4.2. in Rasmussen et al. (2006) the log probability of LCO is:

$$L_{LCO}(\mathbf{y}, X, \theta) = \Sigma_{i=1}^{n} \log p(y_i | X, y_{-i}, \theta) \tag{5.17}$$

where $\mathbf{y}$ is the predicted responses, X prediction locations, θ the Gaussian process hyperparameters, y_i are the predictions for the ith class responses, y_{-i} are the response values excepting those in the class. The predictive mean and variance are:

$$\begin{aligned} \mu_i &= y_i - [K^{-1}\mathbf{y}]_i [[K^{-1}]_{ii}]^{-1} \\ \sigma_i^2 &= [[K^{-1}]_{ii}]^{-1} \end{aligned} \tag{5.18}$$

where μ_i are the mean predictions of the ith class observations, σ_i^2 are the variance predictions for the ith class observations, K is the covariance matrix $k(X, X)$ where k is the covariance function of the Gaussian process, $[\cdot]_i$ is the ith class entries of a vector, $[\cdot]_{ii}$ is the ith class rows and i^{th} class columns of a matrix.

Having generated a proper score rule (see Gneiting & Raftery (2007)) in this case the predictive likelihood functions are proper scoring rules. Therefore they can be used for model selection. In order to carry out this model selection process a step-wise approach was adopted:

1. Start with latitude, longitude and time as model inputs and baseline score at $-$Inf.

2. Try adding each morphological parameter one at a time.

3. Pick the model with best score as a result of the added morphological parameter as the current model, compared against score for the current model.

4. Consider models that have each of the inputs removed. Accept the best model if better than baseline.

in which all possible input variables were considered for adding to the model one at a time then the best added. Then all included variables were considered for removal

The model starts in a state (η) where latitude, longitude and time are predictions for each of the other input variables that are excluded at the start of the stepwise process. As such the state of the model can be described by the baseline log likelihood and a string of booleans representing the state of inclusion of variables. At the start of the modelling process with latitude, longitude, time, roughness length with topology, roughness length without topology and simple mean LIDAR height:

baseline is:

$$\zeta = -\text{Inf} \tag{5.19}$$

state of model:

$$\eta = \{TRUE, \ TRUE, \ TRUE, \ FALSE, \ FALSE, \ FALSE\} \tag{5.20}$$

The algorithm is as follows:

1. Calculate new score for the models which have one of the excluded variables, η_i is FALSE.

2. Compare the scores to the baseline, if a score is better change the corresponding η_i to TRUE and assign ζ to the new score.

3. Calculate new scores for the models which exclude each input individually.

4. Compare new scores from the new models to ζ, if a score is better update ζ and change the corresponding η_i to FALSE.

5. Repeat steps 1. to 4. until neither steps 2. or 4. result in a change in η.

A more full discussion of Bayesian model selection can be found in Gelman et al. (2014).

5.3   Results

In order to consider the significance of roughness length a number of methods were considered. First a linear model was implemented using a six month period of data that directly attempts to consider the slope parameters. This model does not represent the correlation between data points.

Secondly, a process of cross validation and additive modelling, detailed in subsection 5.2.2.3, was carried out, using sub-samples of the data. Relative improvement of methods was assessed by using actual roughness length against randomly sampled replacements. This approach has potential to be more robust to overconfidence at the cost of being less clear as to the sort of behaviour it is searching for compared to clear linear correlation linear models evaluate.

5.3.1   Evaluation of Roughness lengths in Birmingham

Figure 5.9 displays the roughness length evaluated across Birmingham both with and without accounting for topography. Figure 5.9a showing the map ~~for~~of roughness length without accounting topography displays a smoother transition between rural areas with low roughness and urban areas, than is displayed in figure 5.9b, which shows the map using the methodology used to account for topography. This ~~possibly~~ could suggest a result of the mean over 30 meters being too small and actually ends up removing the buildings entirely for larger buildings. The sharp edges of the building being insufficient to be noticed.

The other notable feature in both maps in figure 5.9 is the complete ~~flatness of~~lack of features in the centre of Birmingham. Anecdotal investigation of the LIDAR data was not able to determine the reason for this. It is possible a data anomaly exist for this data, either in data formatting or absence of data.

5.3.2   Linear Modelling

The standard linear regression modelling approach assumes that the relationship between regression variables and outputs is linear, and assuming Gaussian noise of observations allows assessment of significance.

The regression models considered were:

- Roughness lengths with and without topology directly regressed against wind speed.

- Roughness length, against local wind speed less the average wind speed (across the known information local wind speed anomaly).

To compare the ability of the linear model to discriminate between noise and the useful information contained in the roughness length, the linear regression was carried out on with the calculated roughness lengths and with randomly generated roughness lengths (uniformly sampled for each location from the range of actual values).

Tables 5.1, 5.2, 5.3 & 5.4 show the coefficients and significance levels of the linear models. The linear modelling intercepts and coefficients using actual roughness length values are significant to the nearly machine precision. There is an overall negative coefficient for the roughness length

Figure 5.9: Maps of Birmingham overlaid with roughness lengths, pale grey areas represent areas with lack of data in the LIDAR data-set

(a)

(b)
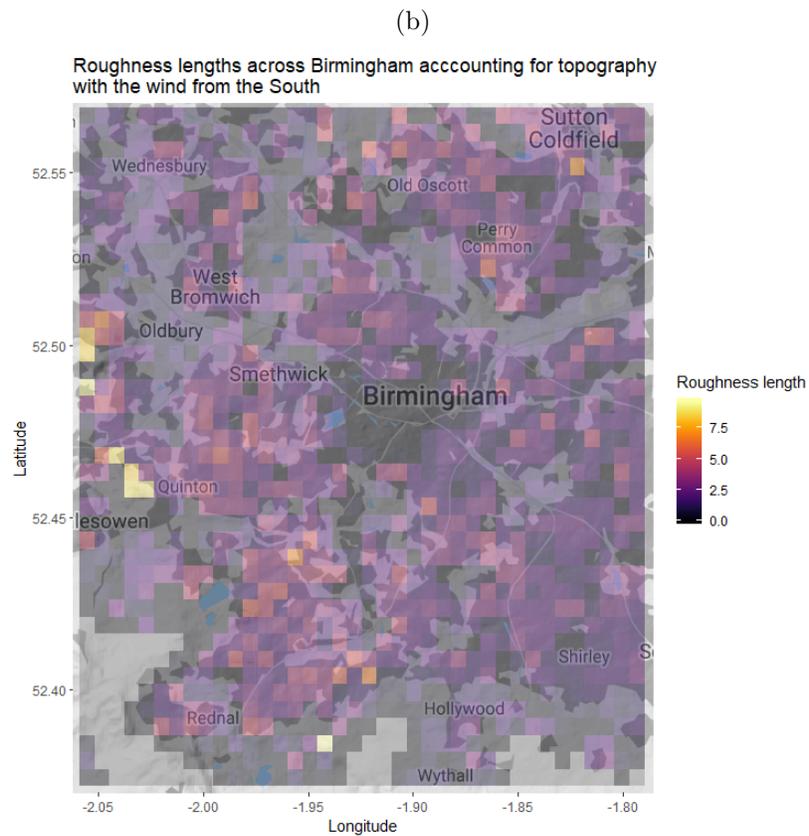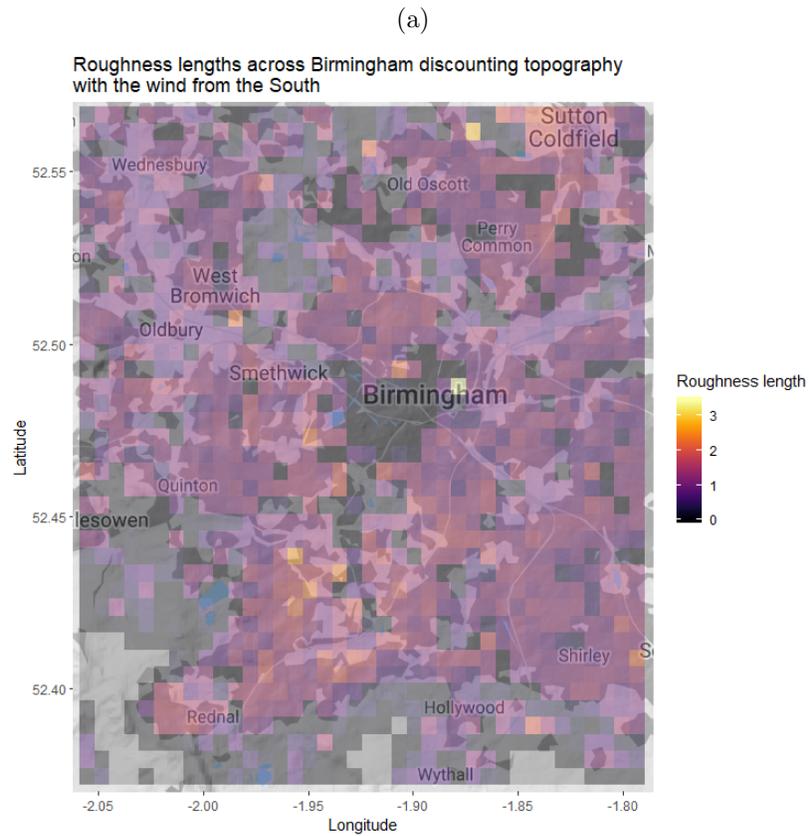
	Estimate	Std. Error	Pr(>|t|)
(Intercept)	1.4627	0.0010	0.0000
Roughness length without topology	-0.1426	0.0014	0.0000
Roughness length with topology	0.0469	0.0007	0.0000

Table 5.1: Linear model of local wind speed against roughness length

	Estimate	Std. Error	Pr(>|t|)
(Intercept)	-0.0066	0.0008	0.0000
Roughness length without topology	-0.1262	0.0012	0.0000
Roughness length with topology	0.0451	0.0005	0.0000

Table 5.2: Linear model of local wind speed anomaly against roughness length

	Estimate	Std. Error	Pr(>|t|)
(Intercept)	1.4191	0.0014	0.0000
Roughness length without topology randomised	-0.0006	0.0008	0.4806
Roughness length with topology randomised	-0.0022	0.0004	0.0000

Table 5.3: Linear model of local wind speed against random sample for roughness length

without topology for both the linear model of wind speed and wind speed anomaly with a small positive gradient for the roughness length with topology.

The model was also evaluated with only roughness length without topology which also had a similar slope coefficient to the first slope coefficient in tables 5.1 and 5.2, strongly suggesting that neither is statistically much different as a predictor.

The p-values are likely overstated. This is because the assumption of no correlation between observations, being poor. Adding more regression variables would not address this situation. Techniques such as sub-sampling depend on having sufficient information to select sub-samples to ensure independence, which broadly speaking corresponds to building a covariance matrix estimates model similar to that which the additive model below generates, the results of which are discussed below.

5.3.3 Additive Modelling Comparison

5.3.3.1 Uptake of Roughness Length as an informative parameter

In order to deal with the non-homogeneous nature of wind speed data over time as discussed earlier the experiment was repeated on uniformly randomly selected time intervals between 1st August and the 20th December.

The simple test is whether the model does uptake the roughness length parameters as an input to predict wind speed. The proportion of models that take up the roughness parameter give us an indicator as the strength of the association.

It is worth considering that this may not be 'true' association in the sense the roughness length is always equally important between different time slices. Even if roughness length is rejected, that is not to say it doesn't contain useful information, just that there is no apparent additional information the model can detect.

	Estimate	Std. Error	Pr($>$|t|)
(Intercept)	-0.0396	0.0012	0.0000
Roughness length without topology randomised	-0.0031	0.0007	0.0000
Roughness length with topology randomised	-0.0006	0.0003	0.0663

Table 5.4: Linear model of local wind speed anomaly against random sample for roughness length

As such, the experiment was run with two other groups of summaries of the LIDAR height maps around the sensors. The algorithm is attempting to find the best possible model. It is likely that there are many summaries of the LIDAR height map which contain some information. In this section it attempted to resolve the magnitude of these relative amounts.

These proportions can be seen in table 5.5. The proportions for the uptake either roughness length and each individually are very similar.

The similarity between these proportions in table 5.5 suggest that the difference between randomly assigned values and calculated roughness lengths is limited meaning this modelling doesn't provide evidence to support the significance of the roughness length, with or without topology.

Type of rough. len. used	Calculated	Randomised
Proportion of all data-sets for which model adopts feature		
Either rough. len.	0.817	0.817
Rough. len. (m) without topology	0.633	0.6
Rough. len. (m) with topology	0.583	0.6
Sensitivity indices for anecdotal time slice model		
Latitude	0.906	0.749
Longitude	0.906	0.765
Time	0.906	0.770
Rough. len. (m)without topology	0.978	0.764
Rough. len. (m) with topology	0.906	0.763
Sensor Attitude (m above sea lvl.)		0.905

Table 5.5: Comparison between models with calculated roughness length and those with random values

5.3.3.2 Relative importance of Roughness length in models

To judge the relative importance of the roughness parameters in this chapter two indices have been evaluated to the first order.

Both relate the output distribution Y to the inputs $\mathbf{x}$ with distribution X. Uniform distributions were used as priors for each input variable used in this analysis.

The first is the mean squared error. A comparison between the mean squared error of the basic model with latitude, longitude and time compared with the models of the comparable model that uses roughness a roughness parameter. The mean squared error is calculated from the leave class method discussed in section 5.2.2.3. Here a comparison can suggest how much an improvement adding the roughness length improves the model. Figure 5.10 shows the improvement in the mean squared error. Essentially, the randomised roughness length forms a baseline to compare the calculated roughness length with.
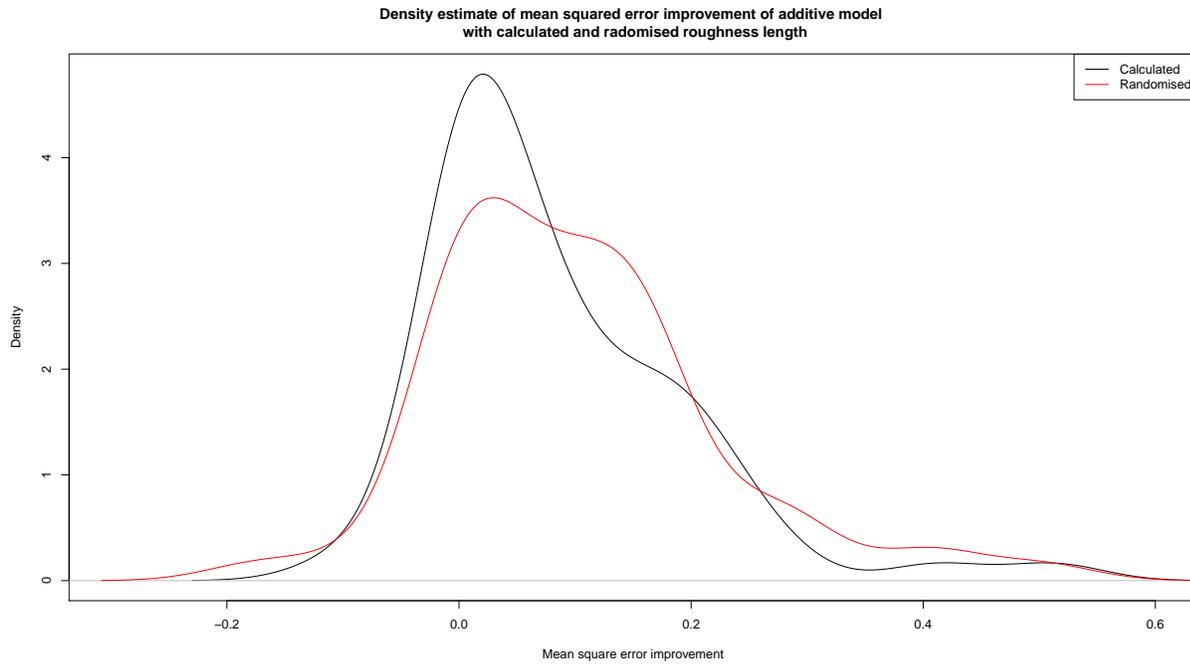
Figure 5.10: Density improvement plots for additive models. Randomised roughness length mean squared error improvement against calculate roughness length mean square error improvement

The second is using sensitivity analysis measures largely popularised by Saltelli et al. (n.d.) to more robustly assess the contribution of the variables to the variation in the model. The lower part of table 5.5 shows the sensitivity of our model according to the variance of the main effects as detailed in Oakley & O'Hagen (2004) for two anecdotal time slice models. To illustrate the main effect these two time slices have also had their main effect plotted in figure 5.11.

It should be remarked that there is relatively little apparent structure in the main effects. Combined with the complete lack of contrast between the distribution of mean square errors in figure 5.10, this suggests that the additive modelling does not identify clear benefit for the calculated roughness length of the data.

The wider conclusions that are reasonable to draw from this is that the p-values are indeed very overstated.

5.4   Conclusions

In conclusion, my key results are as follows:

- The calculation of roughness length using pre-processed lidar data was highly feasible. The challenges with the this approach appear to be incomplete data and some anomalies such as the apparently low roughness of the centre of Birmingham.

- The linear modelling of the wind speed data from the BUCL data-set Warren, Young, Chapman, Muller, Grimmond & Cai (2016) over Birmingham suggests higher aerodynamic roughness length decreases wind speed over the same terrain.

Figure 5.11: Main effect plots of additive non-parametric models of calculated and randomised roughness lengths along each axis around a point in space: mean value in black, data points as blue circles, uncertainty bounds in red

(a) Calculated roughness length
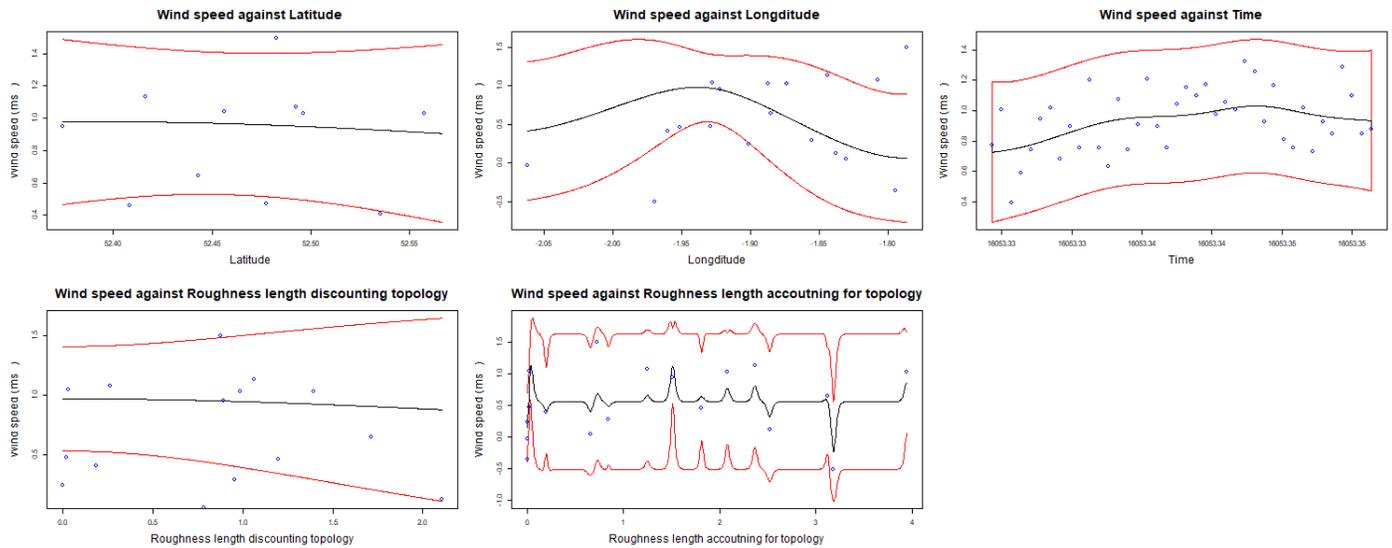
(b) Randomised roughness length
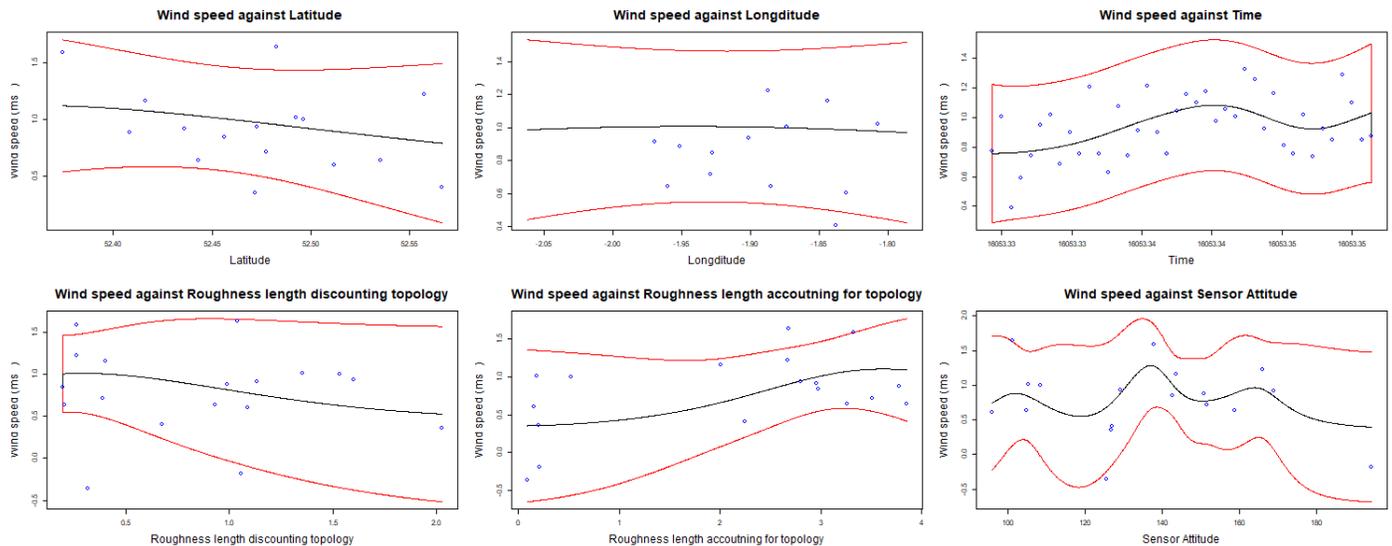
- The additive model using Gaussian processes, suggests that the linear modelling is over-confident in the strength of this relationship.

- Initial analysis (linear modelling) of the wind speed data from the BUCL data-set Warren, Young, Chapman, Muller, Grimmond & Cai (2016) over Birmingham suggests higher aerodynamic roughness length decreases wind speed over the same terrain.

- Further analysis (additive model) using Gaussian processes, suggests that the initial analysis using linear modelling is over-confident in the strength of this relationship.

The conclusion is that the gap between theoretical modelling of wind speed and the impact of roughness length, such as Mo & Liu (2018), observed at sensor locations remains large. This is evidenced by the strong theoretical evidence basis for the importance of roughness length compared to the relatively weak basis for empirical evidence for its importance.

5.5   Model Fitness

5.5.1   User Requirements

The user requirement of this chapter focuse on need of those designing fluid dynamic simulation models:

- They need to understand the empirical relationship in the wind speed and the roughness length in Birmingham.

- They need to understand the methodology and any uncertainty associated with the conclusions of this chapter.

5.5.2   Chapter Output

The output of this chapter is:

- A methodology for calculating roughness length or rugosity from LIDAR height map data. Section 5.2.1.3 contains the definition of roughness length and figure 5.9 maps this roughness length across Birmingham.

- Tables 5.1, 5.3, 5.2 & 5.4 display the results of the linear regression of roughness length against topology variables.

- Table 5.5 and figure 5.11 display the results of additive modelling with Gaussian Process regression.

5.5.3   Evaluation of outputs against user requirements

The conclusions of this chapter are that wind speed does decrease with roughness length and that linear modelling is over confident. In more specific terms:

- Whilst the conclusions were clear and concise qualitative descriptions that can easily be understood, the methodology is opaque. The additive Gaussian process methodology is partly unclear in its results. The lack of a clear behaviour to be seen in figures 5.10 or 5.11 make it difficult to interpret even for those familiar with Gaussian processes.

- The mapping of the results of the roughness length methodology provides a visual confirmation that the results relate to the city topography. Users are able to engage with properties of the methodology such as the city coverage and the variation of the values produced across the city.

In summary, the need to understand how the wind speed empirically relates to roughness length is communicated. The communication in this chapter meets the user requirement to be clear and concise. The fact that the format is qualitative is a result of composite conclusions that form the linear and additive modelling. The two methodologies contradictory results make the methodology more difficult to understand.

Within this chapter figures in 5.2 provides insight into the impact other possible factors that could have been includes as model regressors. The generally unclear signal picked up by the Gaussian processes suggests within the data available there are no highly informative variables available for inclusion.

5.5.4 Recommendations

The review of the output of this chapter suggests positive and negative recommendations:

- That clear qualitative summaries of results can meet the sort requirements relating to the recommendation from chapter 4 for summary statistics to clearly articulate chapter results. Clarity and brevity are key.

- Models that have contradictory results should have a clear process for comparing results in order to ensure that a stakeholder can understand how summary conclusions have been reached.

Further work

The results of the further analysis leave the conclusions of this work less clear. There are several legitimate interpretations of this.

1. The situation is that there is insufficient information to conclude that roughness length is an important piece of information for urban wind speeds at the around 2m above the ground of the sensors in the BUCL data-set.

2. Roughness length is a very small part of the system that determines urban wind speeds and that 'black box modelling', such as Gaussian processes or neural networks, may be capable of producing similar predictive results without being provided with the roughness length.

3. A third scenario is that the models are not fitting to the data correctly hence suggesting that the results were simply not meaningful.

Further work to differentiate between these different scenarios might be to use a data-set from a more dense sensor network across a city, this would provide further statistical analysis more cases to work with to examine what evidence there is for roughness length and other factors influencing wind speeds. Using techniques such as the model selection by likelihood exhibited in chapter 3 to compare models and improve model fit whilst computationally expensive enable the final model to be much fitted to the data hence provide greater confidence in the conclusions.

Chapter 6

Impact on air temperature of urban density, albedo and altitude in Birmingham

Cities are being increasingly affected by climate change. Many papers have speculated over the extent to which air temperature over a city is affected by the cityscape itself. Key to these assessments is the albedo of the urban terrain, the proportion the incident solar radiation reflected by the surface. Proposed mitigations include adjusting the urban albedo by installation of reflective roofing or lighter coloured paved surfaces. Solecki et al. (2005) discusses the provision of heat island adaptions in New Jersey, which include albedo adjustments.

This chapter discusses evidence that albedo is an important factor in Birmingham for the temperature in general and whether the degree of urbanisation is a larger factor. The height above sea-level was also taken into account and found to have a contributing impact. I found that the evidence for impact of albedo on roughness length was confounded by the strength of the association between temperature and general urban density.

Keywords— Uncertainty Quantification - Urban Climate - Albedo - Air Temperature - Altitude - Sensitivity Analysis

6.1   Introduction

This chapter looks for the relationship between the city surface and the air temperature in Birmingham. The Birmingham University Climate Laboratory data-set Bassett et al. (2016), provided the air temperature data. I investigate the relationship between albedo and urban density by introducing how to differentiate the rural areas from inner cities. The overheating in cities is seen to be the product of many factors. Energy balance models often consider the proportion of reflected incident solar energy or albedo along with other factors such as latent heat flux and infrared emissions.

The urban heat island has been subject to a significant amount of research. The BUCL data-set was used by its original creators for this purpose (Bassett et al. 2016). Tomlinson et al. (2011) discusses the vulnerability of population health to overheating. The study identifies city centre areas as being at much higher risk of overheating, although was unable to verify the results. These

results are supported by Sanchez-Guevara et al. (2019) suggesting that inner city areas are at greater risk. Sanchez-Guevara et al. (2019) also identified low income and older populations as being more at risk. It identified London as being at low risk from overheating. These have relevance to Birmingham in that the work identifies overheating as a risk to modern cities.

This study aims to add to this work by investigating how urban density, rurality (introduced in subsection 6.2.2) and albedo impact on air temperature. The main question investigated in this chapter is whether it is possible to separate the impact of albedo and urban density, or whether rurality is more influential. Bassett et al. (2016) quantified the heat island effect in several components. The main two components are the advection, heating that follows downwind, and local heating components, the heating that appears unrelated to wind. I will instead break it down into components of height and how urban the area is.

A large part of the investigation in this chapter is aimed at evaluating the relationship between albedo and rurality. In principle they might be distinct. Large industrial warehouse areas can have bright roofing whereas city centre areas might have darker roofing with air conditioning usage. Even these descriptions hint at the potential difficulties in separating the effect of reflected light from other effects in the area such as heat dumped in the atmosphere from the air conditioning units.

The role urban terrain albedo plays in urban overheating is different in different contexts and evaluating its impact is not easy. Solecki et al. (2005) recommends that urban greening and high albedo surfaces can be potential solutions to urban overheating. Solecki et al. (2005) observes that urban greening is particularly difficult to improve in lower income areas due to lack of space.

To provide estimates of the albedo this work makes use of the European Commission Joint Research Centre (ECJRC) on Earth Observations Data and Processing Platform Europeean Commision: JOINT (n.d.). This provides state of the art estimates of different albedo values. This work focuses using the Bi-Hemispherical Broadband albedo (BH-BB albedo) values, which would strongly correlate to the infra-red radiation reflected.

This measure provides a proxy for urban density using Google map images company. (n.d.) by assessing the RGB pixels and calculating average pixel locations and providing a ranking on this basis. This is discussed in more detail in subsection 6.2.2.

The work in this chapter also outlines the significance of the factors at different times of day looked at in this work. The data is subdivided into 30 minute windows on each day.

6.2 Methodology

6.2.1 Introduction

In this chapter the goal is to examine the significance of urban terrain on air temperature in Birmingham. The motivation being the known impact of temperature on human comfort. Particular focus is made to sourcing information that is available across different geographical areas.

A striking feature of images of cities and rural areas is that generally cities are grey and rural space is green. This can be codified in the processing of the RGB image as defined below. This process could be refined with a pre-classified data set, see section 6.2.2.

The modelling methodology here uses simple linear regression to identify the broad trend and look for significance. This is covered in section 6.2.3. The temperature data-set used for this analysis is a subset of the BUCL data-set (Bassett et al. 2016).

6.2.2 'Rurality'

There are many ways of measuring the degree of city density. Common intuition suggests metrics such as population density or more specifically number of people with residences per square km. Such metrics require substantial effort to identify individual residences and the estimated population of each residence. Such information was not available for this study. In this study universally available Google maps images of the local area are available.

6.2.2.1 Definition

In this work the aim is to identify the degree to which urban external air temperatures depend on the albedo value of the surrounding land. This is somewhat complicated by the fact that the albedo is strongly linked to the degree of urbanisation in the local area.

In order to evaluate this extent of dependency between air temperatures and urban surfaces in this work a novel approach using publicly available GoogleMap images to estimate the degree of urban extent. The measure selected evaluates the degree of colouration in the RGB GoogleMap image by treating each pixel as coordinate value and then calculating the angle with the $[1, 1, 1]$ vector, then averaging across all pixels. Values along the $[0, 0, 0]$, $[256, 256, 256]$ axis represent purely black, grey and white images (such as white roofed industrial estates and large amounts of impermeable surfaces) whereas higher values represent images that have strong colours such as green/brown fields. Figure 6.2 provides a visual aid to the explanation below.

In equation form the rurality or angle (this is the vertical angle in figure 6.2a) with the $[1, 1, 1]$ can be written for a image of N RGB pixels, $\mathbf{x}_i$:

$$\cos^{-1}\left(\frac{1}{N}\sum_i\left(\frac{\mathbf{x}_i\cdot[1,1,1]}{|\mathbf{x}_i|\sqrt{3}}\right)\right) \tag{6.1}$$

This formulation works easily for the monochrome cities of the UK. However cities in other parts of the world are significantly different to cities in the UK. For example the red tiled inner city of Barcelona. This metric needs some adjustment to work more generally. Two additional locations have been picked out in order to examine the generality of this metric: Barcelona and Kampala. Whilst hardly a comprehensive view of world cities this does provide an example of the potential strengths and limitations of this approach. Figure 6.1 shows the evaluation of this method on different locations.

Barcelona is a very old city with historic red roofing which causes the metric to be less meaningful as the inner city is very strongly coloured compared to Birmingham which has more greys in the city centre. The vegetation is much less vibrant green than the UK vegetation due to the drier Mediterranean climate which makes the rural areas greyer. Kampala was picked due being a city from a developing country as opposed to Birmingham and Barcelona which are in Western developed countries. The road development is less consistent and therefore a different colouration.

The performance of the initial version of the indicator is good for Birmingham but poor for Barcelona indicating the inner city is more rural than the suburbs, clearly not the case. By
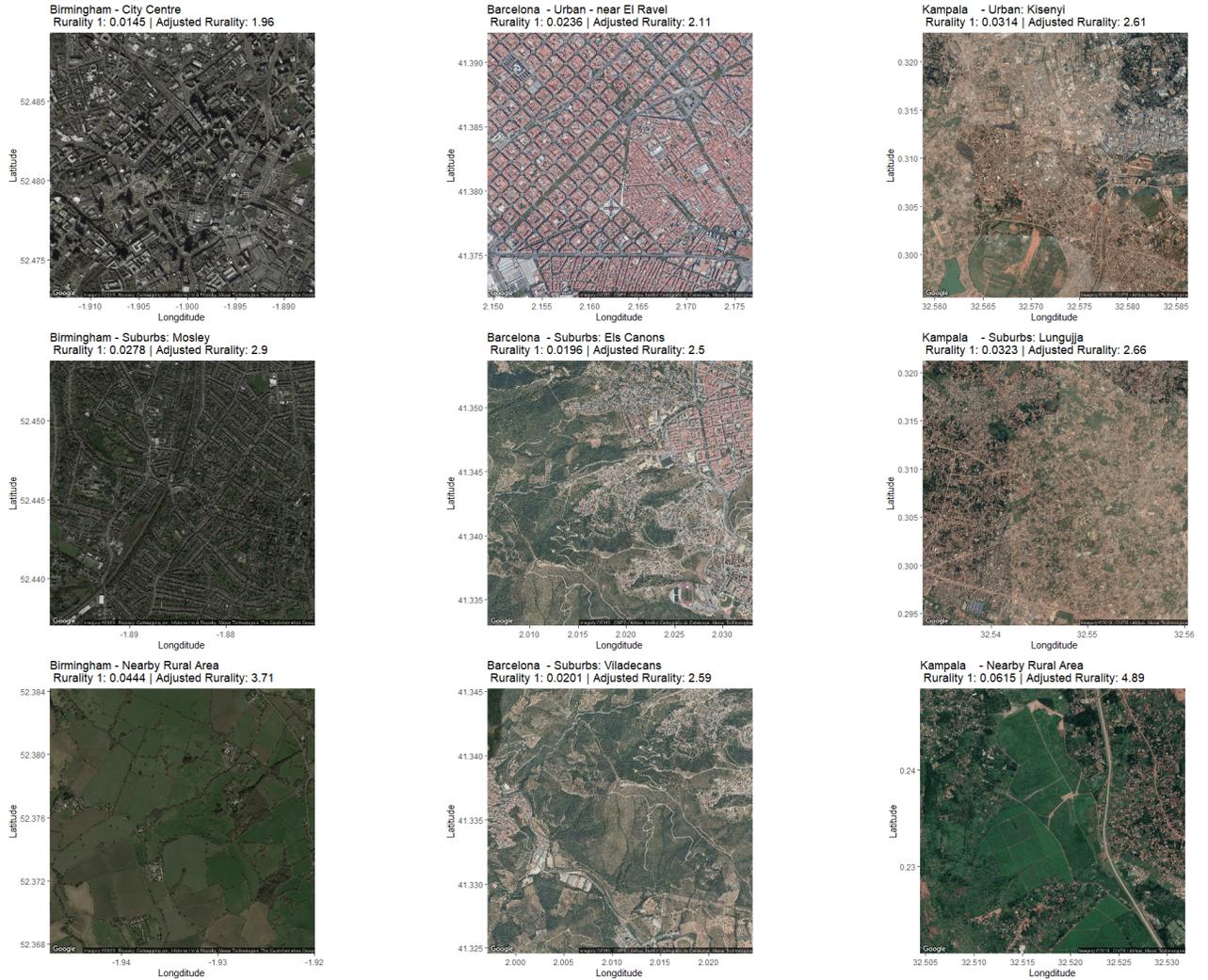
Figure 6.1: Example of Google maps sourced and the evaluated rurality index (high resolution image)

adjusting for the colour of the image (red versus green) the rurality can be made to work for Barcelona. The average colour of the image (or angle around the $[1, 1, 1]$ axis, horizontal angle in figure 6.2a) can be written:

$$\cos^{-1}\left(\frac{1}{N} \sum_i \frac{\mathbf{x}_i \cdot \mathbf{n}_1}{\sqrt{(\mathbf{x}_i \cdot \mathbf{n}_1)^2 + (\mathbf{x}_i \cdot \mathbf{n}_2)^2}} \right) \tag{6.2}$$

where $\mathbf{x}_i$ are again the pixels, $\mathbf{n}_1$, $\mathbf{n}_2$ are vectors normal to $[1, 1, 1]$ and perpendicular to each other. The values are then rescaled and then averaged with a weight of $\frac{2}{3}$ for equation 6.1 and $\frac{1}{3}$ for equation 6.2 to give the adjusted rurality.

The adjusted rurality orders the tiles of the three citieshttps://www.overleaf.com/project/5d6bf1678a8a6a6 correctly and demonstrates the potential of this technique to contrast rural and urban areas. The rest of this work uses the original rurality measure. As there is no real benefit when working within a single city, the original rurality measure was used to simplify code revisions.

Figure 6.2: Demonstration of the positioning of an image and how the angles in relation to the black-white axis show us about a pixel. The one directing between the pixel to the black-white axis provides information about how 'grey' the pixel is, with larger values representing more coloured pixels. The other angle around the black-white axis is informative about the particular colour the pixel is.

(a) Diagram showing two angles that define the colour of a pixel alongside the distance from the origin.
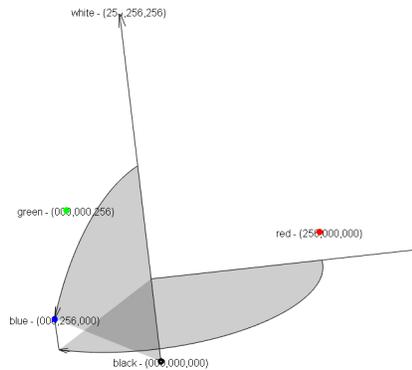
(b) An example image in this case a google maps image of all Birmingham

(c) The plotting of the pixels from 6.2b. The axis are the pixels which in a RGB image vary between 0 and 256 on each axis.
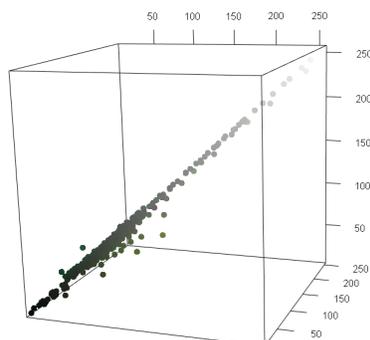
6.2.3 Why use linear modelling for inference

Linear modelling is used across a very broad array of situations. The model relates a response variable y to regression variables x. The strength of a local linear approximation is derived from general calculus where differentiable functions locally are approximately linear. To illustrate this, the first order Taylor expansion is:

$$f(x + h) = f(x) + h \cdot f'(x) + O(h) \tag{6.3}$$

That is to say if the variation, h, in a x is sufficiently small, then the behaviour is linear within this small variation, $O(h)$.

To say y is linear in x would be equivalent to:

$$y = ax + b \tag{6.4}$$

where a and b are the coefficients that do not vary with x. That is not to say a and b do change, in fact in this chapter I examine how they change in relation to times and produce uncertainty estimates for them. In particular a is of interest in that it represents if the value of y increases or decreases in relation to x.

As such linear modelling is suitable if the other factors are held constant or the variation is small enough that the behaviour is locally linear. The other situation that is relevant here is the detection of overall trends. The behaviour of a function may be highly complex, however, if in general there is an increase or decrease of an output y in relation to x, whilst the linear relationship is a poor approximation it remains highly useful to demonstrate the relationship which the framework of linear modelling does.

As such in the complex urban/rural environment around cities linear modelling provides a summary of the effect of regression variables in the case of this chapter albedo, rurality and height (or altitude above sea level).

6.2.4 BUCL and Earth Observation and Processing Platform data-sets

As stated earlier in this chapter the data-set used for the air temperature in different locations in Birmingham was the Birmingham Climate Laboratory data-set The temperature data-set used for this analysis is a subset of the BUCL data-set Bassett et al. (2016). This data-set was filtered for excessively anomalous readings such as a disproportionate number of values exactly 0°C and error values less than -20°C, this potentially could be during freezing/thawing of ice on the temperature gauge. The data was also filtered to the year 2013. This restriction was based on the interval having the vast majority of the sensors active consistently during that interval of time.

The European Commission Joint Research Centre (EC-JRC) Earth Observation and Processing Platform is a EU publicly funded project to generate satellite information on Europe for research purposes. A small part of the processed data was estimated albedo values in Birmingham. Figure 6.3 plots out these values across Birmingham. I chose to use the Bi-Hemispherical Broad Band (BH-BB) albedo values as visual inspect suggested the values were related, and the BH-BB values are averaged across direction and wavelength making them a general rather than specific measure so less likely to have no correlation.
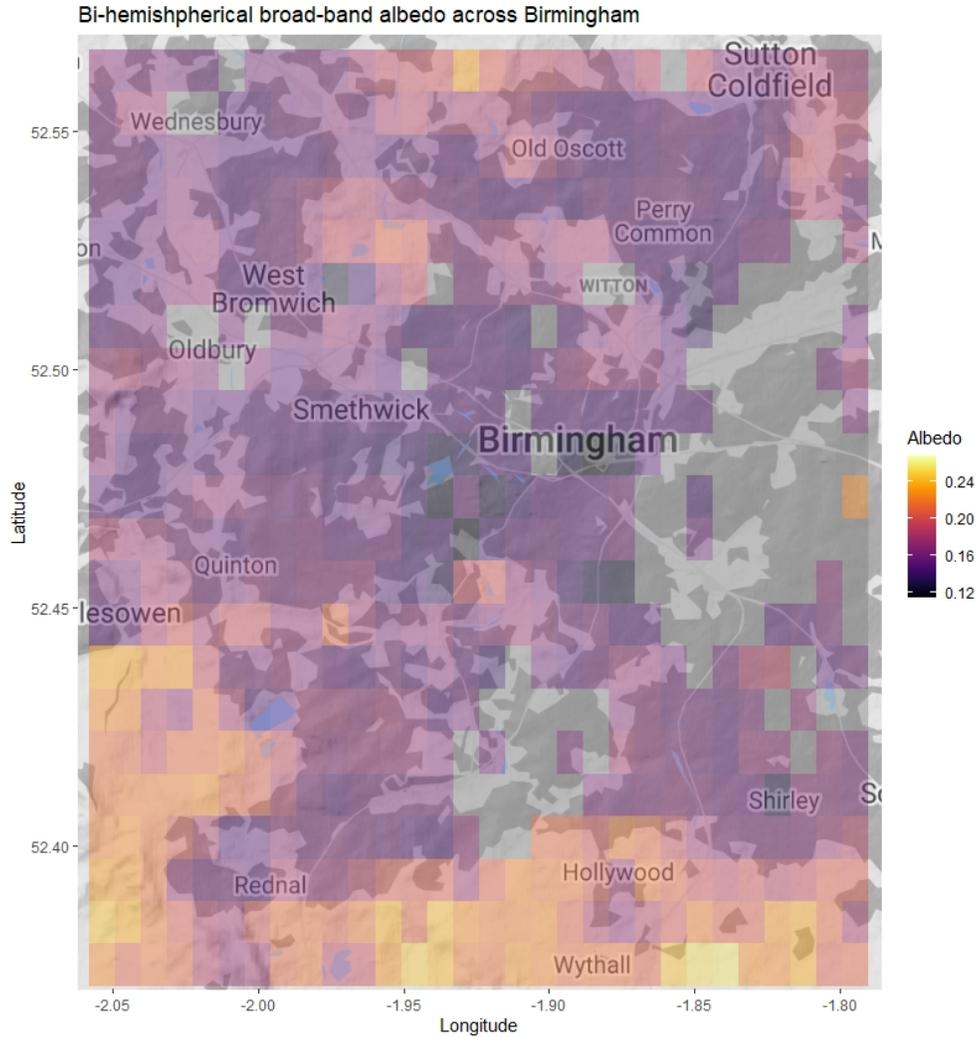
Figure 6.3: Plot of the BH-BB albedo value across Birmingham from Europeean Commision: JOINT RESEARCH CENTRE Earth Observation Data and Processing Platform (n.d.), broadband radiation is typically 0.29–2.5 μm. For comparison Sailor & Fan (2002) lists albedo values mostly between 0.14-0.18 for urban areas.

6.2.5 Data partitioning and models fitted

In order to look at the relationship between different temperature and the regression variables of albedo, rurality and height at different points in time, I subdivided the data into half hour intervals on each day. This means each interval contains values from every minute for each weather station and air temperature station in the BUCL data-set. The different locations of stations can be seen in fig. A.1 in appendix A, further details of the data can be found here.

On each of these intervals I then carry out linear regression. The model is simple; y is the response variable in this case the air temperature; x is the vector of regression variables in stage 1 one, the rurality and albedo, in stage 2 the rurality and height above sea level:

$$y = ax + b + \epsilon \qquad (6.5)$$

where a is the fitted gradient coefficient, b is the intercept and ϵ is a noise term. I will not discuss the

noise term instead discussing the variation in the gradient terms from the different time intervals throughout the day. The intercept, b also has no particularly useful interpretation as none of the rurality, albedo or height values in this case stray close to zero.

I will examine the coefficients themselves to compare how the different models fit. It will also indicate the relative impact of regression variables at different times of day. I will also examine the R^2 values of the regression, and 'relative coefficients'. Relative coefficients were calculated by dividing the coefficient values for the regression variable by the variation in the standard deviation of the regression variable, calculated for each sub data-set. More specifically:

$$\frac{\hat{a}}{\sigma_x} \tag{6.6}$$

where $\hat{a}$ is the fitted value of the regression coefficient and σ_x is the standard deviation for the specific regression variable in question.

6.2.6   Models Investigated

The goal of this chapter is to evaluate the impact of city surface albedo on the air temperature and its significance relative to other factors. These other factors I investigate are rurality and height above sea-level.

To investigate this I subdivided the data as described in subsection 6.2.5. This is subject to the model choices. I do this in two stages:

1 Compare the effect of the albedo with rurality. This is done by using linear regression with each variable individually and then comparing the results. The results can be found in section 6.3.1. The primary results from this comparison are what information about the effect the city-scape has on temperature. These two models were chosen for a direct comparison of impact because of suspected close correlation of impact. 6.4 shows the values of rurality and the BH-BB albedo. The correlation of .41 (2 d.p.) suggests that the spread of the albedo values in relation to the rurality values is not well spaced.

- The two linear models can be written:

$$\begin{aligned} T^t &= \alpha + \beta x^t_{albedo} + \epsilon \\ T^t &= \alpha + \beta x^t_{rurality} + \epsilon \end{aligned} \tag{6.7}$$

where T^t is the temperature at the time interval t, α is the intercept, β is the coefficient of the linear model (results in figure 6.5), x^t_{albedo} and $x^t_{rurality}$ are the value if the albedo and rurality at that time interval, whilst rurality doesn't change there is some limited change between the value for albedo value over the course of the year for the albedo and ϵ is the noise term.

- If the values were better spaced then it would be difficult to attribute similar behaviour to a correlation. As it is you might well expect that the weather station locations in Birmingham with highest albedo would be more rural and darker areas more urban. Even light roofed industrial estates for example have substantial asphalt covered areas for example. The .4 correlation leaves it unclear whether or not two are connected or not.

- Preliminary work suggested that including both variables in the model resulted in the variables have occasional sign flips and exchanges in the magnitude between the variables suggesting a close relationship between variables. Direct comparison between the individual models was therefore adopted to distinguish the better predictor.

2 The second stage focuses on the features of impact of the rurality and height on temperature. Three models are compared.

- The first model regresses temperature on rurality. As such can be used to contrast for any interaction with the combined third model to see if the coefficient varies when the two are used to produce a joint model. The model can be written:

$$T^t = \alpha + \beta x^t_{rurality} + \epsilon \tag{6.8}$$

where T^t is the temperature at time interval t, α is the intercept, β is the coefficient of the linear model and $x^t_{rurality}$ is the values of rurality at time interval t and ϵ is the noise term.

- The second model fulfils the same role for height above sea level. Allowing a comparison with the combined third model. The model can be written:

$$T^t = \alpha + \beta x_{height} + \epsilon \tag{6.9}$$

where T^t, α, β and ϵ are as above in equation 6.8 and x_{height} are the values of the height above sea level.

- The third model is a regression of temperature against both rurality and height. The model can be written:

$$T^t = \alpha + \beta_r x^t_{rurality} + \beta_h x_{height} + \epsilon \tag{6.10}$$

where T^t, α and ϵ are as above in equation 6.8, β_r and β_h are the coefficients for rurality and height above sea-level and $x^t_{rurality}$ and x_{height} are as above in equations 6.8 and 6.9.

Stage one leads to the conclusion that rurality and albedo values contain essentially very similar predictive information about temperature values. This leads to stage two where this is combined to examine the extent to which height affects the temperature anomaly and provides further information for examining the strength of the effect at different times of day.

6.3   Results

To summarise the results, in stage one, it was found that albedo was a substantially worse predictor than rurality in terms of the temperature. In stage two, its height and rurality have an opposite relationship in terms of how they impact on air temperature: during the day increased height maximum impact on air temperature. Rurality was found to have the biggest impact at night: the more urban the area the warmer it will be.

Plot of rurality against albedo
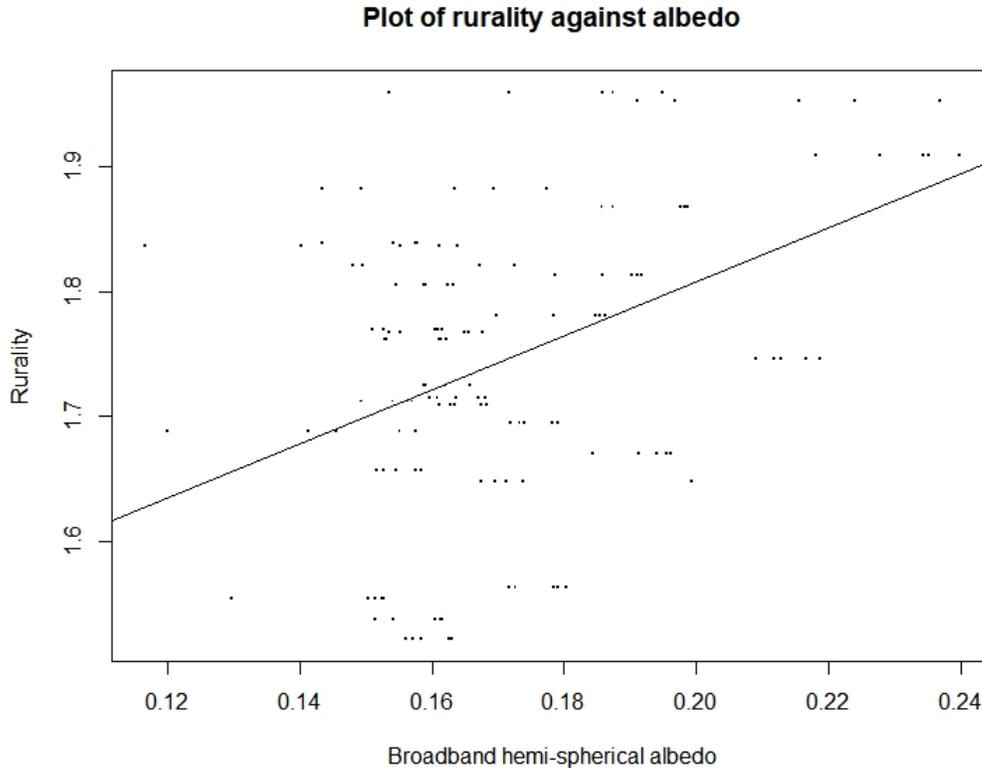
Figure 6.4: Comparison between albedo and rurality values with line of best fit. The correlation between albedo and rurality is .407.

6.3.1   Stage one: Significance of albedo and rurality across different times of day

The results of the stage one analysis are displayed in figures 6.5 and 6.6. These figures display the link between time of day and the effect of different regressors on air temperature. The mean of the different values from sub-data sets from different days is used to represent the values. This gives a metric that whilst influenced by more extreme values is not completely overridden by them. In this case due to the R^2 value being around the magnitude of the order 0.1 and 0.01 extreme values aren't likely to drag the value too far one way or the other.

Figure 6.5 is of two plots of the coefficients of the albedo and rurality. There is a striking similarity between the top and bottom plots. The coefficients are both negative which mean more albedo or more rurality mean more cooling. Whilst the values themselves don't mean much, the shape is interesting. The larger values at night are more suggestive of urban heat island effects which are classically considered to be larger at night than deflecting incident solar radiation which it might be expected to have a larger impact on the temperature during the day. For a discussion on urban heat islands see Taha (1997), and for a more foundational approach and Mirzaei (2015) for further discussion. The similar shape suggests that the same effect might be being observed in both cases.

The issue of whether albedo and rurality are representing the same thing is compounded by the similarity of the plots in fig. 6.6. The R^2 values can be directly compared unlike the coefficient values. R^2 values can be used to assess model fit, tell us how well the data is modelled from

Figure 6.5: Plots comparing the mean gradient coefficient values for the linear model with albedo and rurality respectively

Coefficients by time of day – albedo model only

Time of day (1/2 hour intervals)

Coefficients by time of day – rurality model only

Time of day (1/2 hour intervals)
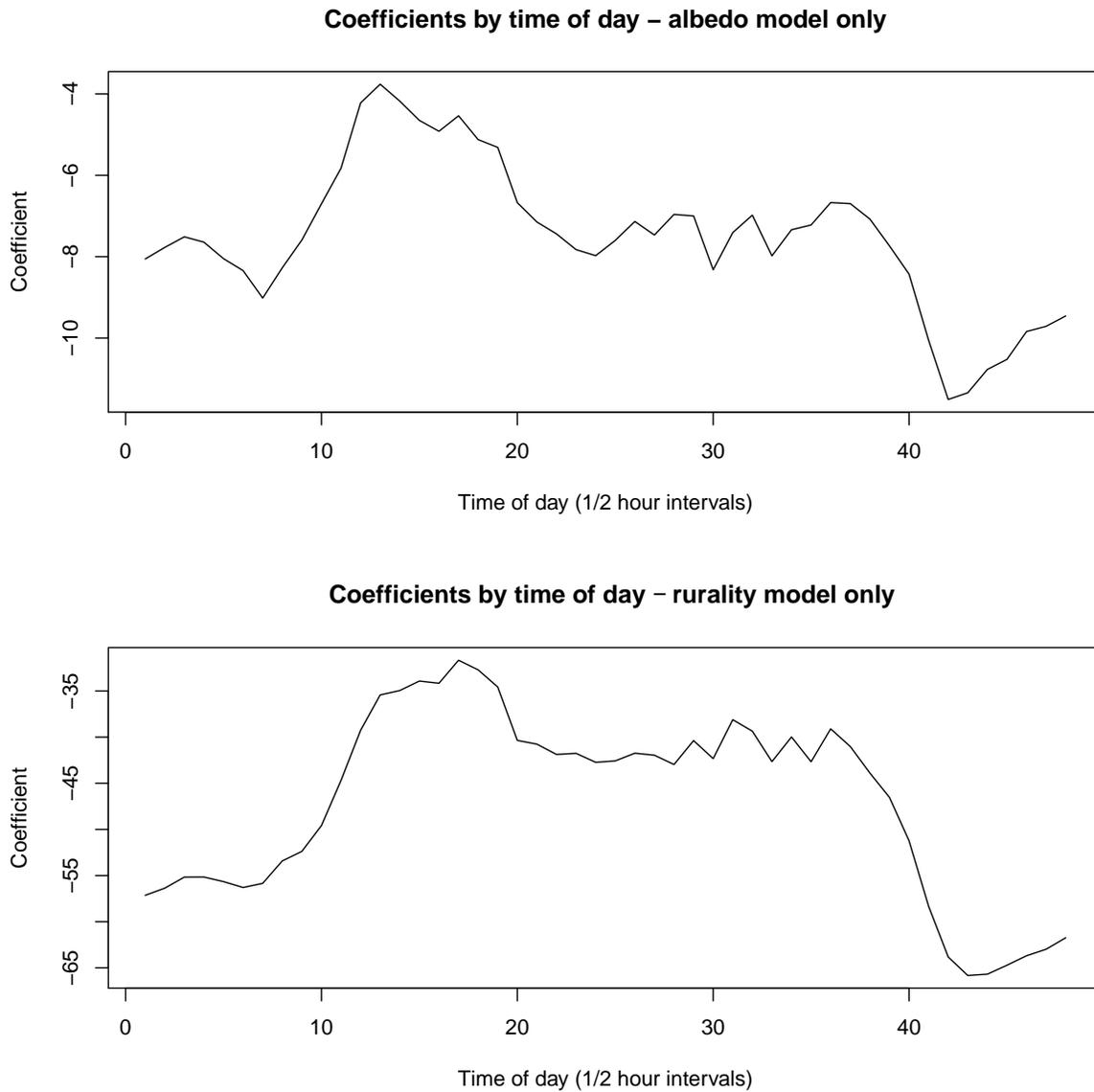
Figure 6.6: Plots of the mean R^2 values comparing the proportion of variance explained by the linear models using albedo and rurality respectively, suggesting rurality is much more informative than albedo. The common shape throughout the day is suggestive of the R^2 contribution from the albedo beign correlated to the rurality, than actual relationship the physical feature, in particular relating the higher R^2 overnight.
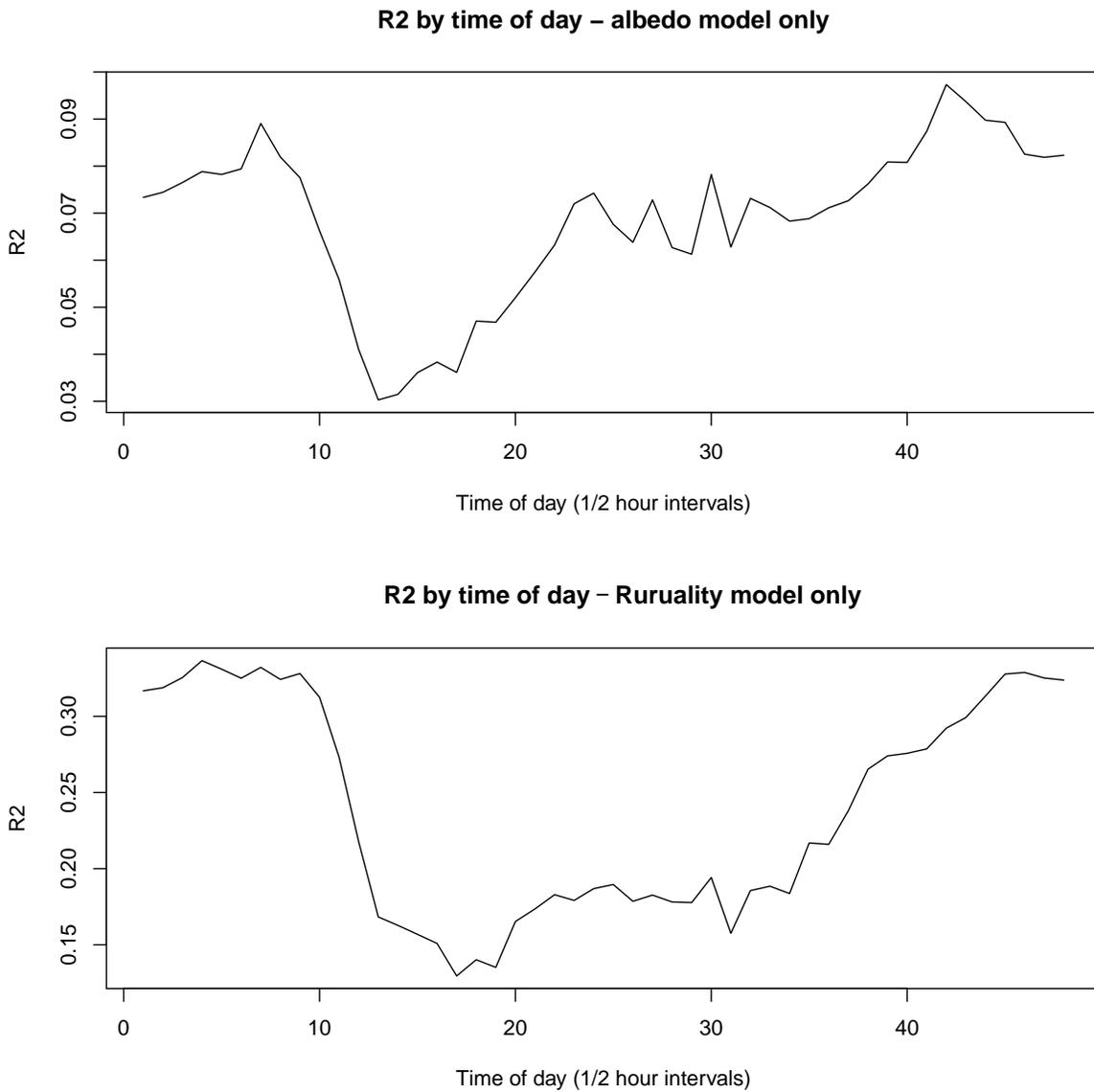
the perspective of variance. The albedo achieves explanations of 3 to 9.5 % of the variance. The rurality explains 14 to 33 % of the variance of the data. Clearly neither are complete in terms of the information required to predict the air temperature.

Another consideration is that the pattern of the time of day that the R^2 value is high. There is a general increase from a minimum in the interval 0600 to 1000 (time intervals 12 to 20) until around 2400 to 0500. The increase takes off faster for the albedo model covering around 50% of the change in variation by 1500 (time interval 30) whereas the rurality improves its prediction in a interval more rapidly between 1500 and 2200 (time intervals 30 and 44). These patterns could be interpreted as being consistent with the features of the effects they represent. Albedo would be expected to see an effect throughout the day as incident radiation is reflected and the impact being accumulated on the reduction in heating of the thermal mass of the system. The rurality is meant to represent the urban heat island, which whilst is a complex group of effects, mostly is human activity producing heat in the evening and the system retaining thermal mass better overnight. The coefficients do not contradict this picture, with the same faster increase throughout the afternoon.

The main reason further analysis was not continued if the effect of albedo was due the very low R^2 values, .09 as a maximum of the mean R^2 value in the different time intervals, represents a very poor proportion of the variance. In summary due to the low R^2 values the conclusions in relation to albedo are weak and not reliable, however the higher degree of certainty afforded by the higher R^2 values associated with the rurality suggest there greater confidence for the associated conclusions.

Figure 6.4 and the correlation of .4 between the rurality and the albedo supports the conclusion from the R^2 values that there is no substantial evidence in this data-set that albedo is affecting the air temperature. The weak correlation combined with a similar pattern of predictive power in R^2 values suggests that the albedo merely acts as a poor proxy for the degree of urbanisation or in this case rurality.

The analysis explored here is an example of statistical study known as causal inference which examines the distinction between correlation and causation. In short this is whether two things happen to move in tandem due to being dependant on something in common or whether variation in one actually depends on the other. In this case it seems highly likely that the changes how urban or the change in rurality are far more significant than change in albedo. The rurality is the cause in this case, a in depth background to this field can be found in (n.d.b).

6.3.2 Stage two: Different patterns in impact of height above sea-level and rurality with uncertainty

The analysis in subsection 6.3.1 doesn't discuss the range of different values in the different time intervals merely considering the mean value. This section focuses on 6.7. The analysis in this section uses scaled coefficients: the coefficient of the linear model is divided by the standard deviation of the sub data-set.

The resultant coefficients are therefore comparable between different models and variables. So a scaled coefficient with a value of .5 would have an impact of 1°C on the predicted air temperature where the input value is two standard deviations from the mean. This means the larger mean coefficients at times in fig. 6.7 suggest variation of around .1°C in the air temperature. Whereas the outlying two standard deviation confidence bounds represent a maximum of around 1°C change with a two standard variation the corresponding variable.

Figure 6.7: Combined plot of scaled coefficients for each model. Three models are compared: Black, rurality used for regression alone; Red, height above sea-level for regression; Green and blue, model using two regressors, height above sea-level and rurality respectively. Dashed lines are confidence bounds at 1.96 standard deviations of the coefficient values. These bound the area confidence interval for each element.
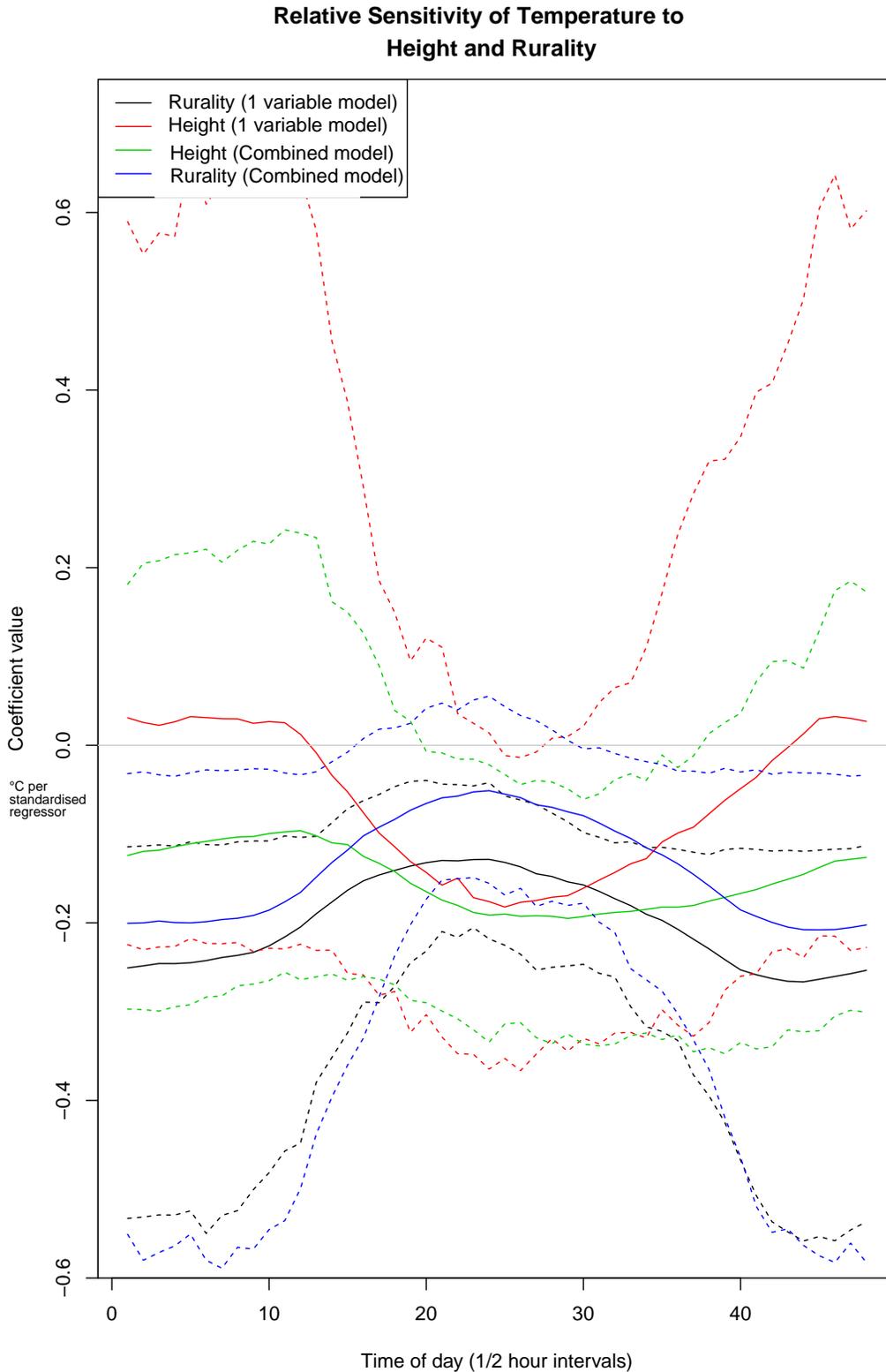
The broad shape of the rurality graphs, in blue and black in fig. 6.7, suggests that the impact of rurality on an area is greater at night than during the day. This supports the analysis in subsection 6.3.1, that the effect is very similar to a heat island effect, with the more rural locations being cooler than more urban areas. The uncertainty, shown by the confidence intervals, is much higher at night which suggests that there is greater variation in behaviour at this time. Possibly when the urban heat island is in effect there are other effects that influence it leading to the variation in impact. During the day the impact is shown to be much lower with coefficient values half those at night-time. This fits with urban heat islands having a largest impact by retaining heat at night.

The height above sea level graphs, in red and green in fig. 6.7, broadly show an opposite relationship from the rurality with the largest effect during the day. The coefficient values are more positive than the rurality values suggesting a smaller influence on the air temperature. The confidence interval around the coefficient for the individual model, in red, for height above sea level is far larger than for the coefficients in the combined model, in green. Interestingly the height above sea level coefficients for the individual model in red stray substantially into the positive area at night. This could imply that at night some higher up areas are warmer, however caution should be considered as the uncertainty is substantial and includes some negative values as well suggesting this could just be noise.

The combined model coefficients, in fig. 6.7 in green and blue, can be noted to contain less varied mean values than with the individual model. The height above sea level in particular has lower variation or a smaller confidence interval. These changes suggest there is some interaction between the terms.

6.4 Conclusions

This chapter evaluates the evidence contained in the BUCL data-set Bassett et al. (2016) contains as to the impact on air temperature of albedo from the ECJRC on Earth Observations Data and processing platform Europeean Commision: JOINT RESEARCH CENTRE Earth Observation Data and Pro (n.d.) and the urban extent as evaluated in the form of rurality in this chapter. The new approach allowed the use of widely available Google Maps company. (n.d.) to evaluate the urban extent systematically without manual inspection.

The ability to utilise additional sources of public data to analyse the effects of urban density on the urban environment has potential to facilitate studies on other areas of policy such as air pollution. There is other open-source data such at OpenStreetMaps used in Munir et al. (2019). The advantage of Google maps is its ubiquity in being available for nearly every city on the planet, though there are some pitfalls as illustrated with the discussion centred on figure 6.1.

Figure 6.7 illustrates that there is a lot of uncertainty over the conclusions around the impact of urban density, however the data supports the common conception that the impact of how urban an area is largest at night. The fact that this pattern however uncertain seems to be borne out across the year suggests this not an illusion in the data but a sustained phenomenon despite the uncertainty over its magnitude. The platform height above sea-level seems to broadly have a slightly weaker impact on the air temperature than urban density with a contrasting relationship to the time of day, with a larger impact during daylight hours.

The main conclusions are as follows:

- Albedo is a weak predictor for air temperature over Birmingham. This does not contradict that it's importance in other climates such as in middle eastern cities such as Bande et al.

(2019), rather that in the climate around Birmingham other factors appear more important. It is possible that some of the relationship between air temperature and albedo are masked by the fact over Birmingham the albedo and how urban the area is, is strongly related.

- The degree of urban extent, measured in this paper as rurality, does have an impact on air temperature over Birmingham. The impact is largest at night peaking around 2200 hours and diminishes during the day. This supports the theoretical work on the established mechanisms of the urban heat island effect acting to retain heat during night time periods.

It would be possible to re-evaluate the methodology to carry out similar analysis in different cities as discussed at the end of subsection 6.2.2.1 and in figure 6.1.

6.5   Model Fitness

6.5.1   User Requirements

The users for this chapter will be looking at how the air temperature is affected by the urban fabric. Their user requirements are similar to the in section 5.5. These requirements are:

- They need to understand how air temperature is related emprically by albedo, urban density and and altitude.

- They need to know which related effects have been considered and modelled.

- They need to understand the methodology behind the modelling in this chapter.

6.5.2   Chapter Output

The output in this chapter can be summarised as follows:

- Mapping of albedo values across Birmingham.

- Methodology for calculating how rural an area is (rurality).

- Visual demonstration of the rurality in different regions.

- Sensitivity analysis of temperature to altitude/height, rurality and albedo.

6.5.3   Evaluation of outputs against user requirements

The mapping of albedo values across Birmingham provides good user understanding of how albedo varies across Birmingham in figure 6.3. The lack of mapping of the rurality across Birmingham was due to cost constraints of making GoogleMap API calls. Figure 6.1 does however demonstrate that within regions the rurality does perform as expected.

This chapter has a similar aspect to chapter 5 in terms of providing inference on the sensitivity of climate to urban fabric. Following on from 5.5 this chapter has different methodology for drawing inference. This chapter makes use of single methodology which is easier to communicate. This chapter addresses casual inference more directly, this field of statistics focuses on the distinction between correlation and causation.

The figures 6.6 & 6.5 highlight the sensitivity of air temperature to rurality and albedo. In the initial stage of analysis the conclusion that albedo was correlated to temperature, rather than being a significant part of the causation for changes in temperature, is useful for making decisions about how to plan Birmingham. In the case of albedo, the degree of rurality or greening is more important for keeping the city and resident warm/cool. Causal inference is a field that should be applied in more detail and rigour in future as it will help support decision making analysis which is demonstrated in this chapter.

The pattern of sensitivity across the daily cycle, in figure 6.7, supports wider analysis of urban heat island effects. The research provides validation to the conclusion that urban heat islands are worth further investigation in terms how they affect Birmingham such how heat island could affect heating bills and cooling costs.

Reviewing the recommendations in section 5.5, it is apparent this chapter whilst not having numerically clear summaries has a much more clearly explained methodology and exploration of data. Data exploration is supported with graphical mapping across Birmingham. Application of the rurality application in different contexts raises the question of the robustness of this work applied in different contexts. The exploration of related variables to air temperature communicates a clearer narrative that contrasts with the conflicted results in the previous chapter, making more apparent to user groups how to use these results.

6.5.4   Recommendations

In this chapter the following recommendations arise when considering model validation:

- To make use of mapping out results in the locations when graphically presenting results, as this communicates properties of the variables.

- To ensure that research has a clear narrative explaining the exploration of data why each step was taken.

- To draw conclusions about causation to distinguish between causation and correlation.

Chapter 7

Conclusions and further research

7.1 Contributions of thesis

This thesis has provided analyses of several different fields in order to bring together architectural ideas and civil engineering ideas around city design and building modelling with modern machine learning and inference techniques. This is based on a future of modelling and decision making that bring together data from multiple sources and drawing inference where there are many influencing factors.

The interdisciplinary nature of this work has led to two main threads forming in parallel in this thesis, rather than a single concentrated thread. They answer questions in different contexts, scales and perspectives. The discussion of model fitting provides context for the models beyond providing novel research. This chapter will review the wider research contributions and then provide recommendations from different chapters dealing with the outcomes of the model fitness reviews at the end of each chapter.

Chapters 3 and 4 look at the smaller scale of the impact of the climate from the perspective of a single building. In particular the management decisions are those made by a private organisation or individual predicting impact. They analyse optimising an asset using techniques rather than drawing generalised conclusions.

Chapters 5 and 6 focused on more holistic consideration of how the micro-climate over a city depends on the city landscape. Chapter 5 draws the conclusion that wind speed at low altitudes is not measurably affected by the roughness length of the underlying urban form. Chapter 6 suggests that whilst the existence of local urban heat island is very definitely detectable, the impact of different albedo values is far less detectable and may be the results of correlation between albedo and the degree of urban density.

7.2 Building energy modelling

Chapters 3 and 4 present a method for bringing together data from multiple sources to develop a method to forecast climate change impacts on building systems. Corporate responsibility has increasingly driven zero carbon plans as laid out in this report, Microsoft & PWC (2020). This puts pressure on all corporate assets for innovation and improvement in efficiency, being able to assess the efficiency of assets as detailed in these chapters has good potential for facilitating planning to

meet these targets. Robert & Kummert (2012) is an example discussing a method for how this can be achieved using down scaled weather files that could then be used by the methods developed in chapters 3 and 4 to create future climate projections.

- The flexible nature of the Gaussian process (GP) is good for handling the general behaviour of a chiller behaviour to external weather, where the external weather actually justifies positive chiller use. Where the weather is sufficiently cold, the system can draw cool air from outside and having zero energy use in this system can be managed by having a classification to distinguish between different modes. This methodology could be extended to handle systems with more complex modes.

- The uncertainty quantification facilitates potential improvement of risk management within building system management in terms of the likely demand relative to current system capacity. This potentially could have more impact on buildings in the context of overheating and future climate change though this was beyond the scope of chapters 3 and 4.

7.3   Uncertainty and the city

In complex urban environments, it can be difficult to separate out clear inference as to the impact of city design features such as the effect of roughness length on wind speed and urban albedo on air temperature. Papers such as Wicht & Wicht (2018) advocate modelling of cities to determine air flow corridors to optimise air flow. In the case of air flow the benefits are largely hypothetical, which is not to say the benefits do not exist, rather that supporting evidence has not been found in Birmingham.

This thesis provides two examples of inference under uncertainty. I draw the following conclusions:

- Chapter 5 concludes that there is some evidence that in Birmingham higher roughness lengths correspond to lower wind speeds. However the confidence and the magnitude of this effect is small.

  - The method of evaluating the roughness lengths by going pixel by pixel over the area of study is in line with standard methodology, if slightly adapted. Trying to account for topology is also an established practice, with debatable success in this context.

  - There are limitations to this study. The more complex study results disagreed with the more simplistic studies. Further work is required to resolve why this discrepancy exists. This demonstrates how more complex modelling doesn't necessarily result in clearer results.

- Chapter 6 concludes that the impact of different urban albedos was swamped by the general impact of urban density and height above sea level. The use of Google Map processing presented in this chapter provides an alternative method for evaluating urban density.

  - Using publicly available information for this sort of analytics facilitates custom metrics that can be tailored to the task at hand. This particular urban metric would need to be further tailed to handle different countries as figure 6.1 shows. The adjusted metric shows that minor adjustments can make the metric more robust, combined with further

measures designed to accommodate for differences in road construction (Kampala) and construction materials (Barcelona), it could be made to work between countries. Potential applications could be used for automatic land scanning for simulation without manual classification.

– Papers such as Gašparović et al. (2019) use satellite imagery directly. Such imagery has not been processed for public consumption in the way GoogleMaps have which has advantages in that how the image has been processed is more transparent, at the cost of being less standardised.

In summary this thesis highlights the challenges of drawing out the precise importance of design factors in urban design from weather data-sets. It also raises the importance of bringing together data from multiple sources in order to draw inference.

7.4 Uncertainty modelling and inference in practice

This thesis is aimed at handling data when there is substantial uncertainty. The raw number of data points may be very large, as with the BUCL data-set, with many 100,000s of rows. However obtaining a paucity of information about how variables such as air temperature vary over space is challenging. This is not because the experiment was in any way poorly designed. The sensors were spread out to cover a variety of locations and not clustered in unhelpful formations. Rather the variables change faster than the spacing of these sensors, despite being fairly numerous. The costs of having complete coverage would be substantial.

It is these limits of a data-set that rub up against the costs of obtaining data that create uncertainty. There are other sources of uncertainty that don't create trade-offs in the same way. The way heating energy used in the Information Commons was measured for example where the values of the metering were rounded after a certain date, leaves the uncertainty in the actual reading value. Estimating original readings could be useful for modelling purposes, especially if the changes in energy use can be smaller than units per day. Alternative reinterpretation of the prediction problem would be to predict the time next meter increment, which might be addressed with state space models, though would require emulators with short execution times.

There's also much to learn about the use of GPs for uncertainty modelling. Ultimately they are a way of modelling a Gaussian noise between observations with a covariance structure. This covariance structure is generally assumed in this thesis to vary smoothly with location, however relative to the spacing of sensors in cities, this may not be the case. It would be expect that variation in weather data, such as time averaged air temperature anomalies, would be smooth if measured over a small enough variation in local conditions and similar topological features. However the assumption seems to hold well enough when dealing with predicting energy use for building systems which suggests that the weather has a 'nice' predictable impact on energy use.

This thesis has shown how many different data-sets from cities can be brought together to calculate summaries of local areas such as urban density or roughness length or building data-sets that can be used to model a building's response to external climate. Uncertainty quantification allows confidence in the results to be better judged and communicated. The variance associated with GP predictions make this much easier. This was useful in the context of building energy models, however, casts doubt of the certainty of the certainty of linear models in chapter 5 is an important consideration.

7.5   Further research

There are many elements in this thesis that I believe have potential for interesting research in terms of the subject matter:

- Development of Gaussian process modelling of time-series data such as building energy system as in chapter 3 and 4. I believe more sophisticated models that have faster run times and accuracy are possible. In particular I would like to see development of more flexible methods for automatically finding different modes of behaviour in data such as treed GP or mixture of experts methods.

- The results of the additive modelling in chapter 5 could potentially be due to poor predictive power of the statistical model. It would interesting to see if a more powerful predictive model came to the same conclusion.

- The results in chapter 5 were interesting in relation to Gaussian processes and randomised replacement values giving the apparently better prediction power than real roughness lengths. This is speculated to be because the random spacing has enough structure to give better fitting covariance values for the normal distribution that the real values. It would be interesting to see if this bears investigation by making the covariance between sensor locations randomly instead.

- The methods in chapter 5 and 6 both have been applied to just Birmingham. It would be interesting to see what results similar methods applied to more cities may be very different circumstances such as other cultures and continents. In particular it would be interesting applying this methodology of using open source LIDAR and Google Map data to Sheffield.

In relation to model fitting summaries in this thesis, the recommendations from the model validation can be summarised as follows:

- To provide clear a summary of the conclusions. in this thesis several options have been explored: simple numeric summaries - chapters 3 & 4 & qualitative conclusions - chapter 5.

- It is important when articulating research to create a narrative around the investigation. Chapter 5 was particularly challenging to understand as the different methods were employed in parallel. In chapter 6, the narrative was much clearer with initial assessment of causation of the albedo variable.

- Communicating uncertainty is important within chapters. The decomposition of variance within chapter 4 is a prime example of making difficult to interpret 'black box' models, more understandable. By identifying that the model is not the majority source of uncertainty it justifies that the model meets the needs of users.

These recommendations have not been validated themselves, however they provide a basis for further investigation and cross-discipline engagement with user groups.

In summary there are several potential developments that can be taken forward from this thesis in term of modelling methodology. Dual behaviour Gaussian processes in particular could be developed further for inference and predictive models. Models also need to have clear structures and have simple effective summaries in order to ensure stakeholder buy-in. In chapters 3 and 4, the

model structure was easily understood and in chapter 3 the likelihood was clearly understandable. Chapters 5 & 6 have less clearly communicable results. ~~It would also be interesting to see the methodology applied across more varied contexts.~~

Bibliography

(n.d.a).

(n.d.b).

Adolphe, L. (2001), 'A simplified model of urban morphology: Application to an analysis of the environmental performance of cities', Environment and Planning B: Planning and Design 28(2), 183–200.
URL: *http://journals.sagepub.com/doi/abs/10.1068/b2631*

Ahmad, M. W., Mourshed, M., Mundow, D., Sisinni, M. & Rezgui, Y. (2016), 'Building energy metering and environmental monitoring - A state-of-the-art review and directions for future research', Energy and Buildings 120, 85–102.

Andrianakis, I., Vernon, I. R., Mccreesh, N., Mckinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M. & White, R. G. (2015), 'Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda', PLoS Comput Biol 11(1).

ASHRAE (2014), ASHRAE Guideline 14-2002.

Bamdad, K., Cholette, M. E., Guan, L. & Bell, J. (2018), 'Building energy optimisation under uncertainty using ACOMV algorithm', Energy and Buildings 167, 322–333.

Bande, L., Afshari, A., Al Masri, D., Jha, M., Norford, L., Tsoupos, A., Marpu, P., Pasha, Y. & Armstrong, P. (2019), 'Validation of UWG and ENVI-met models in an Abu Dhabi District, based on site measurements', Sustainability (Switzerland) 11(16).

Bassett, R., Cai, X., Chapman, L., Heaviside, C., Thornes, J. E., Muller, C. L., Young, D. T. & Warren, E. L. (2016), 'Observations of urban heat island advection from a high-density monitoring network', Quarterly Journal of the Royal Meteorological Society 142(699), 2434–2441.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2836*

Bastos, L. S. & O'Hagan, A. (2009), 'Diagnostics for Gaussian Process Emulators', Technometrics 51(4), 425–438.

BEIS (2020), 'The uk's nationally determined contribution under the paris agreement'.
URL: *https://www.gov.uk/government/publications/the-uks-nationally-determined-contribution-communication-to-the-unfccc*

BEIS (2021), 'Green homes grant: make energy improvements to your home (closed to new applicants)'.
URL: *https://www.gov.uk/guidance/apply-for-the-green-homes-grant-scheme*

Bienvenido-Huertas, D., Oliveira, M., Rubio-Bellido, C. & Marín, D. (2019), 'A comparative analysis of the international regulation of thermal properties in building envelope', Sustainability (Switzerland) 11.

Bishop, C. M. (1994), 'Neural networks and their applications', American Institute of Physics 1803.
URL: *http://rsi.aip.org/rsi/copyright.jsp*

Bonfigli, C., Chorafa, M., Diamond, S., Eliades, C., Mylona, A., Taylor, B. & Virk, D. (2017), Design methodology for the assessment of overheating risk in homes, Technical report.

Boodi, A., Beddiar, K., Amirat, Y. & Benbouzid, M. (2020), 'Simplified building thermal model development and parameters evaluation using a stochastic approach', Energies 13.

Bottema, M. (1997), 'Urban roughness modelling in relation to pollutant dispersion', Atmospheric Environment 31(18), 3059–3075.

Bou-Zeid, E., Anderson, W., Katul, G. G. & Mahrt, L. (2020), 'The persistent challenge of surface heterogeneity in boundary-layer meteorology: a review', Boundary-Layer Meteorology 177(2), 227–245.

Boukouvalas, A., Gosling, J. P. & Maruri-Aguilar, H. (2014), 'An Efficient Screening Method for Computer Experiments', Technometrics 56(4), 422–431.
URL: *http://www.tandfonline.com/doi/abs/10.1080/00401706.2013.866599*

Bueno, B., Norford, L., Hidalgo, J. & Pigeon, G. (2013), 'The urban weather generator', Journal of Building Performance Simulation 6(November), 269–281.
URL: *http://www.tandfonline.com/doi/abs/10.1080/19401493.2012.718797*

Busby, D. (2008), 'Hierarchical adaptive experimental design for Gaussian process emulators', Reliability Engineering & System Safety 94, 1183–1193.

Caiado, C. & Goldstein, M. (2015), 'Bayesian uncertainty analysis for complex physical systems modelled by computer simulators with applications to tipping points', Communications in Nonlinear Science and Numerical Simulation 26(1-3), 123–136.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S1007570415000428*

Carvalho, D., Rocha, A., Gómez-Gesteira, M. & Silva Santos, C. (2014), 'WRF wind simulation and wind energy production estimates forced by different reanalyses: Comparison with observed data for Portugal', Applied Energy 117, 116–126.

Castillo, E., Fernández-Canteli, A., Hadi, A. S. & López-Aenlle, M. (2007), 'A fatigue model with local sensitivity analysis', Fatigue and Fracture of Engineering Materials and Structures 30(2), 149–168.

Change, C. o. C. (2017), 2017 Report to Parliament – Meeting Carbon Budgets: Closing the policy gap, Technical report.
URL: *https://www.theccc.org.uk/wp-content/uploads/2017/06/2017-Report-to-Parliament-Meeting-Carbon-Budgets-Closing-the-policy-gap.pdf*

Chipman, H. A., George, E. I. & McCulloch, R. E. (1998), 'Bayesian CART Model Search: Rejoinder', Journal of the American Statistical Association 93(443), 957.

Chow, D. H., Levermore, G., Jones, P., Lister, D., Laycock, P. J. & Page, J. (2002), 'Extreme and near-extreme climate change data in relation to building and plant design', Building Services Engineering Research and Technology 23(4), 233–242.

Christenson, M., Manz, H. & Gyalistras, D. (2006), 'Climate warming impact on degree-days and building energy demand in Switzerland', Energy Conversion and Management 47(6), 671–686.

Civil Institute for Building Service Engineers (n.d.).
    URL: *https://www.cibse.org/*

Committee on Climate Change (UK) (2015), The Fifth Carbon Budget: The next step towards a low-carbon economy, Technical report.
    URL: *https://documents.theccc.org.uk/wp-content/uploads/2015/11/Committee-on-Climate-Change-Fifth-Carbon-Budget-Report.pdf*

company., G. (n.d.), 'Google Maps'.
    URL: *www.googlemaps.com*

de Rubeis, T., Falasca, S., Curci, G., Paoletti, D. & Ambrosini, D. (2020), 'Sensitivity of heating performance of an energy self-sufficient building to climate zone, climate change and hvac system solutions', Sustainable Cities and Society 61.

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X. & Bosilovich, M. G. (2012), 'Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations', Journal of Climate 25(6), 1916–1944.

Delgarm, N., Sajadi, B., Azarbad, K. & Delgarm, S. (2018), 'Sensitivity analysis of building energy performance: A simulation-based approach using OFAT and variance-based sensitivity analysis methods', Journal of Building Engineering 15, 181–193.

Dodoo, A. & Gustavsson, L. (2016), 'Energy use and overheating risk of Swedish multi-storey residential buildings under different climate scenarios', Energy 97, 534–548.
    URL: *http://dx.doi.org/10.1016/j.energy.2015.12.086*

Dougherty, E. R. & Lotufo, R. A. (2003), Morphological Processing of Gray-Scale Images, in A. R. Weeks Jr., ed., 'Hands-on Morphological Image Processing', SPIE PRESS, chapter 6, pp. 129–162.
    URL: *https://www.spiedigitallibrary.org/ebooks/TT/Hands-on-Morphological-Image-Processing/Chapter6/Morphological-Processing-of-Gray-Scale-Images/10.1117/3.501104.ch6?SSO=1*

Durosaiye, I. O., Hadjri, K. & Liyanage, C. L. (2019), 'A critique of post-occupancy evaluation in the UK', Journal of Housing and the Built Environment 34(1), 345–352.
    URL: *https://doi.org/10.1007/s10901-019-09646-2*

Eames, M. E. (2016), 'An Update of the UK's Design Summer Years: Probilistic Design Summer Years for Enhanced Overheating Risk Analysis in Building Design', Journal of Building services engineering research and technology (Accepted) .

Eames, M., Kershaw, T. & Coley, D. a. (2011), 'On the creation of future probabilistic design weather years from UKCP09', Building Services Engineering Research and Technology 32, 127–142.
URL: *http://dx.doi.org/10.1177/0143624410379934*

Egmont-petersen, M., Ridder, D. D. & Handels, H. (2002), 'Image processing with neural networks — a review', Pattern Recognition 35, 2279–2301.

Energy, U. D. O. (2010), 'EnergyPlus Engineering Reference: The Reference to EnergyPlus Calculations', US Department of Energy p. 1051.
URL: *http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:EnergyPlus+Engineering+Refere*

Europeean Commision: JOINT RESEARCH CENTRE Earth Observation Data and Processing Platform (n.d.).
URL: *http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/GHSL/*

Fcibse, M. E. I. M. (2010), 'Climate change and future energy consumption in UK housing stock', Building Services Engineering Research and Technology 31(1), 75–90.

Few, J., Pullinger, M., Mckenna, E., Elam, S., Webborn, E. & Oreszczyn, T. (2021), 'Smart energy research lab: Energy use in gb domestic buildings 2021 authors'.
URL: *www.serl.ac.uk*

Foteinaki, K., Li, R., Heller, A. & Rode, C. (2018), 'Heating system energy flexibility of low-energy residential buildings', Energy and Buildings 180, 95–108.

Gál, T. & Unger, J. (2009), 'Detection of ventilation paths using high-resolution roughness parameter mapping in a large urban area', Building and Environment 44(1), 198–206.

Gašparović, M., Zrinjski, M. & Gudelj, M. (2019), 'Automatic cost-effective method for land cover classification (ALCC)', Computers, Environment and Urban Systems 76(December 2018), 1–10.

Gelman, A., Hwang, J. & Vehtari, A. (2014), 'Understanding predictive information criteria for Bayesian models', Statistics and Computing 24(6), 997–1016.
URL: *https://link.springer.com/content/pdf/10.1007%2Fs11222-013-9416-2.pdf*

Ghosh, S., Gavaghan, D. J. & Mirams, G. R. (2018), 'Gaussian process emulation for discontinuous response surfaces with applications for cardiac electrophysiology models', pp. 1–49.
URL: *http://arxiv.org/abs/1805.10020*

Gibson, R., Grant, C., Forristall, G. Z., Smyth, R., Owrid, P., Hagen, O. & Leggett, I. (2009), 'Bias and uncertainty in the estimation of extreme wave heights and crests', Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE 2, 363–373.

Gillich, A., Saber, E. M. & Mohareb, E. (2019), 'Limits and uncertainty for energy efficiency in the UK housing stock', Energy Policy 133(December 2018), 110889.
URL: *https://doi.org/10.1016/j.enpol.2019.110889*

Gneiting, T. & Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', Journal of the American Statistical Association 102(477), 359–378.

GPy (2012), 'GPy: A Gaussian process framework in python'.
   URL: *http://github.com/SheffieldML/GPy*

Gramacy, R. B. & Lee, H. K. H. (2008), 'Bayesian treed Gaussian process models with an application to computer modeling', 103(483).
   URL: *http://arxiv.org/abs/0710.4536*

Guan, Y., Chen, H. & Zhou, X. (2015), 'Study of urban ventilation corridor planning method based on a case study of Guiyang , China', ICUC9 - 9th International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment Study (1).
   URL: *http://www.meteo.fr/icuc9/LongAbstracts/poster_2-21-1321088_a.pdf*

Gusson, C. S. & Duarte, D. H. (2016), 'Effects of Built Density and Urban Morphology on Urban Microclimate - Calibration of the Model ENVI-met V4 for the Subtropical Sao Paulo, Brazil', Procedia Engineering 169, 2–10.
   URL: *http://dx.doi.org/10.1016/j.proeng.2016.10.001*

Haldi, F. & Robinson, D. (2010), 'On the unification of thermal perception and adaptive actions', 45, 2440–2457.

Heo, Y., Choudhary, R. & Augenbroe, G. (2012), 'Calibration of building energy models for retrofit analysis under uncertainty', Energy and Buildings 47, 550–560.
   URL: *http://linkinghub.elsevier.com/retrieve/pii/S037877881100644X*

Heo, Y. & Zavala, V. M. (2012), 'Gaussian process modeling for measurement and verification of building energy savings', Energy and Buildings 53, 7–18.

Hirsch, J. J. & Associates (2010), 'DOE-2 (version2.2-47h2)'.
   URL: *www.doe2.com*

Höfle, B. & Hollaus, M. (2010), 'Urban vegetation detection using high density full-waveform airborne lidar data-combination of object-based image and point cloud analysis', International archives of photogrammetry, . . . XXXVIII, 281–286.
   URL: *http://www.isprs.org/proceedings/XXXVIII/part7/b/pdf/281_XXXVIII-part7B.pdf*

Högström, U. (1996), 'Review of some basic characteristics of the atmospheric surface layer', Boundary-Layer Meteorology 78(January), 215–246.

Hong, T., Buhl, F. & Haves, P. (2008), 'EnergyPlus Run Time Analysis', p. 73.

Humphreys, M. & Nicol, F. (2015), Enviromental Design: CIBSE Guide A, CIBSE.
   URL: *http://www.cibse.org/knowledge/cibse-guide/cibse-guide-a-environmental-design-new-2015*

IPCC (2014), : Climate Change 2014: Synthesis Report. Contr; II and III to the Fifth Assessment Report of the I; R.K. Pachauri and L.A. Meyer (eds.)]. IPCC; Geneva; Switzerland; 151 pp., Technical report.

IPCC (2021), 'Ipcc sixth assessment report'.

IPCC (2022), 'Climate change 2022 - mitigation of climate change - summary for policymakers (spm)'.

Jeanjean, A. P., Monks, P. S. & Leigh, R. J. (2016), 'Modelling the effectiveness of urban trees and grass on PM2.5 reduction via dispersion and deposition at a city scale', Atmospheric Environment 147, 1–10.

Jentsch, M. F., Eames, M. E. & Levermore, G. J. (2015), 'Generating near-extreme Summer Reference Years for building performance simulation', Building Services Engineering Research and Technology pp. 1–27.
URL: *http://bse.sagepub.com/cgi/doi/10.1177/0143624415587476*

Jentsch, M. F., Levermore, G. J., Parkinson, J. B. & Eames, M. E. (2013), 'Limitations of the CIBSE design summer year approach for delivering representative near-extreme summer weather conditions', Building Services Engineering Research and Technology pp. 0143624413478436–.
URL: *http://bse.sagepub.com/cgi/content/long/0143624413478436v1*

Jiang, C., Soh, Y. C., Li, H., Masood, M. K., Wei, Z., Zhou, X. & Zhai, D. (2017), 'CFD results calibration from sparse sensor observations with a case study for indoor thermal map', Building and Environment 117, 166–177.
URL: *http://dx.doi.org/10.1016/j.buildenv.2017.02.007*

Kanda, M., Inagaki, A., Miyamoto, T., Gryschka, M. & Raasch, S. (2013), 'A New Aerodynamic Parametrization for Real Urban Surfaces', Boundary-Layer Meteorology 148(2), 357–377.

Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. (2004), 'kernlab – An S4 Package for Kernel Methods in R', Journal of Statistical Software 11(9), 1–20.
URL: *http://www.jstatsoft.org/v11/i09/*

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z. & Djurovic-Petrovic, M. (2010), 'A review of bottom-up building stock models for energy consumption in the residential sector', Building and Environment 45, 1683–1697.

Keatinge, W. R., Coleshaw, S. R. & Holmes, J. (1989), 'Changes in seasonal mortalities with improvement in home heating in England and Wales from 1964 to 1984', International Journal of Biometeorology 33(2), 71–76.

Kent, C. W., Grimmond, C. S., Gatey, D. & Barlow, J. F. (2018a), 'Assessing methods to extrapolate the vertical wind-speed profile from surface observations in a city centre during strong winds', Journal of Wind Engineering and Industrial Aerodynamics 173(September 2017), 100–111.
URL: *https://doi.org/10.1016/j.jweia.2017.09.007*

Kent, C. W., Grimmond, C. S., Gatey, D. & Barlow, J. F. (2018b), 'Assessing methods to extrapolate the vertical wind-speed profile from surface observations in a city centre during strong winds', Journal of Wind Engineering and Industrial Aerodynamics 173, 100–111.

Krige, D. G. (1951), '"A statistical approach to some basic mine valuation problems on the Witwatersrand', Metal. and Mining Soc. of South Africa 52(6), 119–139.

Lee, D. O. (1979), 'The influence of atmospheric stability and the urban heat island on urban-rural wind speed differences', Atmospheric Environment (1967) 13(8), 1175–1180.

Leeson, G. W. (2018), 'The Growth, Ageing and Urbanisation of our World', Journal of Population Ageing 11(2), 107–115.

Lewis, M. (2010), 'The university of sheffield library information commons: A case study', Journal of Library Administration 50(2), 161–178.

Li, P., Froese, T. M. & Brager, G. (2018), 'Post-occupancy evaluation: State-of-the-art analysis and state-of-the-practice review', Building and Environment 133, 187–202.

Li, X., Ma, Y., Wang, Y., Liu, N. & Hong, Y. (2017), 'Temporal and spatial analyses of particulate matter (PM10 and PM2.5) and its relationship with meteorological parameters over an urban city in northeast China', Atmospheric Research 198(September 2016), 185–193.
URL: *http://dx.doi.org/10.1016/j.atmosres.2017.08.023*

Liu, J., Niu, J., Du, Y., Mak, C. M. & Zhang, Y. (2019), 'LES for pedestrian level wind around an idealized building array—Assessment of sensitivity to influencing parameters', Sustainable Cities and Society 44, 406–415.

Ltd., I. E. S. (n.d.), 'IES Virtual Environment'.

Lukač, N., Štumberger, G. & Žalik, B. (2017), 'Wind resource assessment using airborne LiDAR data and smoothed particle hydrodynamics', Environmental Modelling and Software 95, 1–12.

MacDonald, I. & Strachan, P. (2001), 'Practical application of uncertainty analysis', Energy and Buildings 33(3), 219–227.

Mahmoudabadi, H., Olsen, M. J. & Todorovic, S. (2016), 'Efficient terrestrial laser scan segmentation exploiting data structure', ISPRS Journal of Photogrammetry and Remote Sensing 119, 135–150.
URL: *http://dx.doi.org/10.1016/j.isprsjprs.2016.05.015*

Makido, Y., Dhakal, S. & Yamagata, Y. (2012), 'Relationship between urban form and CO 2 emissions: Evidence from fifty Japanese cities', Urban Climate 2, 55–67.
URL: *http://dx.doi.org/10.1016/j.uclim.2012.10.006*

Marston, A. (2018), Post Occupancy Evaluation-How Building Energy Models Bridge the Gap Between Design and Operation: A Case Study, in 'CIBSE Techincal Symposium'.

Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor & (eds.), T. W. (2018), IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global, Technical report.

Menberg, K., Heo, Y. & Choudhary, R. (2019), 'Influence of error terms in Bayesian calibration of energy system models', Journal of Building Performance Simulation 12(1), 82–96.

Met Office: Hadley Centre (2019), UK Climate Projections : Headline Findings (Version 2).
URL: *www.metoffice.gov.uk*

Microsoft & PWC (2020), The Building Blocks for Net Zero Transformation, Technical Report
    September.
    URL: *https://www.pwc.co.uk/sustainability-climate-change/assets/pdf/building-blocks-net-*
    zero-companies-transformation.pdf

Mirzaei, P. A. (2015), 'Recent challenges in modeling of urban heat island', Sustainable Cities and
    Society 19, 200–206.

Mo, Z. & Liu, C. H. (2018), 'A wind tunnel study of ventilation mechanism over hypothetical urban
    roughness: The role of intermittent motion scales', Building and Environment 135(October
    2017), 94–103.
    URL: *https://doi.org/10.1016/j.buildenv.2018.02.031*

Moonen, P. & Allegrini, J. (2015), 'Employing statistical model emulation as a surrogate for CFD',
    72, 77–91.

Moussa, A. (2010), Automatic Classification and 3D Modeling of Lidar Data, in 'IAPRS', Vol.
    XXXVIII, pp. 155–159.
    URL: *http://www.isprs.org/proceedings/XXXVIII/part3/b/pdf/155_XXXVIII-part3B.pdf*

Muller, C. L., Chapman, L., Grimmond, C. S., Young, D. T. & Cai, X. M. (2013), 'Toward a
    standardized metadata protocol for urban meteorological networks', Bulletin of the American
    Meteorological Society 94(8), 1161–1185.

Munir, S., Mayfield, M., Coca, D., Jubb, S. A. & Osammor, O. (2019), 'Analysing the performance
    of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—a case
    study in Sheffield', Environmental Monitoring and Assessment 191(2).

Murphy, J., Sexton, D., Jenkins, G., Boorman, P., Booth, B., Brown, K., Clark, R., Collins, M.,
    Harris, G., Kendon, L. & Meteorology Office Hadley Centre (2010), UK Climate Projections
    science report: Climate change projections, Technical report.

Neto, A. H. & Fiorelli, F. A. S. (2008), 'Comparison between detailed model simulation and
    artificial neural network for forecasting building energy consumption', Energy and Buildings
    40, 2169–2176.

Nguyen, A.-T., Reiter, S. & Rigo, P. (2014), 'A review on simulation-based optimization methods
    applied to building performance analysis', Applied Energy 113, 1043–1058.
    URL: *http://www.sciencedirect.com/science/article/pii/S0306261913007058*

Oakley, J. E. & O'Hagen, A. (2004), 'Probabilistic Sensitivity analysis of complex models: a
    Bayesian approach', Journal of the Royal Statistical Society. Series B 66, 751–769.

O'Hagan, A. (2004), 'Bayesian analysis of computer code outputs', Reliability Engineering &
    System Safety 91(10–11), 1290—-1300.
    URL: *http://link.springer.com/chapter/10.1007/978-1-4471-0657-9_11*

O'Hagan, A. & Oakley, J. E. (2004), 'Probability is perfect, but we can't elicit it perfectly',
    Reliability Engineering and System Safety 85(1-3), 239–248.

P. Rastogi, M. A. (2015), 'EMBEDDING STOCHASTICITY IN BUILDING SIMULATION THROUGH SYNTHETIC WEATHER FILES Interdisciplinary Laboratory of Performance Integrated Design (LIPID), Ecole Polytechnique F', Proceedings of BS 2015 IBPSA .

Park, S. & Wang, J. (2017), On-line Gaussian Process Mixture Model for Robot Model Learning, in 'IEEE International Conference on Robotics and Automation (ICRA)', pp. 6449–6454.
   URL: *http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7989762*

Pastore, L., Corrao, R. & Heiselberg, P. K. (2017), 'The effects of vegetation on indoor thermal comfort: the application of a multi-scale simulation methodology on a residential neighborhood renovation case study', Energy and Buildings 146, 1–11.
   URL: *http://www.sciencedirect.com.ludwig.lub.lu.se/science/article/pii/S0378778817312550*

Pisello, A., Pignatta, G., Castaldo, V. & Cotana, F. (2015), 'The Impact of Local Microclimate Boundary Conditions on Building Energy Performance', Sustainability 7(7), 9207–9230.
   URL: *http://www.mdpi.com/2071-1050/7/7/9207/*

Qu, Y., Liang, S., Liu, Q., He, T., Liu, S. & Li, X. (2015), 'Mapping surface broadband albedo from satellite observations: A review of literatures on algorithms and products', Remote Sensing 7(1), 990–1020.

Rajasekar, U. & Weng, Q. (2009), 'Urban heat island monitoring and analysis using a non-parametric model: A case study of Indianapolis', ISPRS Journal of Photogrammetry and Remote Sensing 64, 86–96.

Rallapalli, H. S. (2010), A Comparison of EnergyPlus and eQUEST Whole Building Energy Simulation Results for a Medium Sized Office Building, PhD thesis.

Ramallo-González, A. P., Eames, M. E., Natarajan, S., Fosas-de Pando, D. & Coley, D. A. (2020), 'An analytical heat wave definition based on the impact on buildings and occupants', Energy and Buildings 216.

Rasmussen, C. E. & Ghahramani, Z. (2002), 'Infinite Mixtures of Gaussian Process Experts', Advances in Neural Information Processing Systems 2, 881–888.
   URL: *https://papers.nips.cc/paper/2055-infinite-mixtures-of-gaussian-process-experts*

Rasmussen, C. E., Williams, C. K. I. & NetLibrary Inc. (2006), Gaussian processes for machine learning, the MIT Press.
   URL: *http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=156015*

Robert, A. & Kummert, M. (2012), 'Designing net-zero energy buildings for the future climate, not for the past', Building and Environment 55, 150–158.
   URL: *http://dx.doi.org/10.1016/j.buildenv.2011.12.014*

Robinson, D. (2011), Computer modelling for sustainable urban design : physical principles, methods and applications, Earthscan.

Rose, S. & Apt, J. (2016), 'Quantifying sources of uncertainty in reanalysis derived wind speed', Renewable Energy 94, 157–165.
   URL: *http://dx.doi.org/10.1016/j.renene.2016.03.028*

Roth, M. (2000), 'Review of atmospheric turbulence over cities', Quarterly Journal of the Royal Meteorological Society 126(564), 941–990.

Roustant, O., Ginsbourger, D. & Deville, Y. (2012), 'DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization', Journal of Statistical Software 51(1), 1–55.
URL: *http://www.jstatsoft.org/v51/i01/*

Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P. (1989), 'Design and analysis of computer experiments', Statistical Science 4(4), 409–435.

Sailor, D. J. & Fan, H. (2002), 'Modeling the diurnal variability of effective albedo for cities', Atmospheric Environment 36(4), 713–725.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. & Tarantola, S. (2010), 'Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index', Computer Physics Communications 181(2), 259–270.
URL: *http://dx.doi.org/10.1016/j.cpc.2009.09.018*

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. (n.d.), Global Sensitivity Analysis The Primer, John Wiley & Sons, Ltd.

Sanchez-Guevara, C., Núñez Peiró, M., Taylor, J., Mavrogianni, A. & Neila González, J. (2019), 'Assessing population vulnerability towards summer energy poverty: Case studies of Madrid and London', Energy and Buildings 190, 132–143.
URL: *https://doi.org/10.1016/j.enbuild.2019.02.024*

Sharp, E., Dodds, P., Barrett, M. & Spataru, C. (2015), 'Evaluating the accuracy of CFSR re-analysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information', Renewable Energy 77, 527–538.

Shinzato, P., Simon, H., Silva Duarte, D. H. & Bruse, M. (2019), 'Calibration process and parametrization of tropical plants using ENVI-met V4–Sao Paulo case study', Architectural Science Review 62(2), 112–125.
URL: *https://doi.org/10.1080/00038628.2018.1563522*

Simon, H. (2016), Modeling urban microclimate, PhD thesis.

Singh, R., Lazarus, I. J. & Kishore, V. V. (2016), 'Uncertainty and sensitivity analyses of energy and visual performances of office building with external venetian blind shading in hot-dry climate', Applied Energy 184, 155–170.

Snelson, E. & Ghahramani, Z. (2006), 'Sparse {Gaussian} processes using pseudo-inputs', Advances in Neural Information Processing Systems 18, 1257.

Solecki, W. D., Rosenzweig, C., Parshall, L., Pope, G., Clark, M., Cox, J. & Wiencke, M. (2005), 'Mitigation of the heat island effect in urban New Jersey', Environmental Hazards 6(1), 39–49.

Staddon, P. L., Montgomery, H. E. & Depledge, M. H. (2014), 'Climate warming will not decrease winter mortality', Nature Climate Change 4(3), 190–194.
URL: *http://www.nature.com/doifinder/10.1038/nclimate2121*

Stokes, G. (1851), 'On the Effect of the Internal Friction of Fluids on the Motion of Pendulums', Transactions of the Cambridge Philsophical Society 9, 8–106.

Strong, M. & Oakley, J. E. (2013), 'An efficient method for computing single-parameter partial expected value of perfect information', Medical Decision Making 33(6), 755–766.

Summerfield, A. J., Lowe, R. J. & Oreszczyn, T. (2010), 'Two models for benchmarking UK domestic delivered energy', Building Research and Information 38(1), 12–24.

Swarup S. Medasani, (Thousand Oaks, C., Yuri Owechko, (Newbury Park, C. & Thommen Korah, (Marina Del Rey, C. (2013), 'STREET CURBAND MEDIAN DETECTION USING LIDAR DATA'.
URL: *https://patentimages.storage.googleapis.com/db/1e/2a/6452eb2a2cc3ab/US8537338.pdf*

Taha, H. (1997), 'Urban climates and heat islands: albedo, evapotranspiration, and anthropogenic heat', Energy & Buildings (25), 99–103.
URL: *http://ci.nii.ac.jp/naid/110007068242/*

Taleghani, M., Marshall, A., Fitton, R. & Swan, W. (2019), 'Renaturing a microclimate: The impact of greening a neighbourhood on indoor thermal comfort during a heatwave in Manchester, UK', Solar Energy 182(January), 245–255.
URL: *https://doi.org/10.1016/j.solener.2019.02.062*

Tian, W. & De Wilde, P. (2011), 'Uncertainty and sensitivity analysis of building performance using probabilistic climate projections: A UK case study', Automation in Construction 20(8), 1096–1109.
URL: *http://dx.doi.org/10.1016/j.autcon.2011.04.011*

Tomlinson, C. J., Chapman, L., Thornes, J. E. & Baker, C. J. (2011), 'Including the urban heat island in spatial heat health risk assessment strategies : a case study for Birmingham , UK', International Journal of Health Geographics 10(1), 42.
URL: *https://doi.org/10.1186/1476-072X-10-42*

Tresp, V. (2000), 'Mixtures of Gaussian processes', Advances in neural information processing systems pp. 654–660.

U.K._Environment_Agency (n.d.), 'Environment Agency Geomatics Survey Data - LIDAR Composite DSM'.
URL: *https://environment.maps.arcgis.com/apps/MapJournal/index.html?appid=c6cef6cc642a48838d38e?*

Ulam, N. M. S. (1949), 'The Monet Carlo Method', Journal of the American Statistical Association 44(247), 335–341.

UNCC (2016), 'The paris agreement'.
URL: *https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement*

Vernon, I., Goldstein, M. & Bower, R. G. (2010), 'Galaxy formation: a Bayesian uncertainty analysis', Bayesian Analysis 5(4), 619–669.
URL: *http://projecteuclid.org/euclid.ba/1340110846*

Volodina, V. & Williamson, D. (2020), 'Diagnostics-Driven Nonstationary Emulators Using Kernel Mixtures', SIAM/ASA Journal on Uncertainty Quantification 8(1), 1–26.

Vonlanthen, M., Allegrini, J. & Carmeliet, J. (2017), 'Multiscale interaction between a cluster of buildings and the ABL developing over a real terrain', Urban Climate 20, 1–19.
URL: *http://dx.doi.org/10.1016/j.uclim.2017.02.009*

Vuckovic, M., Kiesel, K. & Mahdavi, A. (2017), 'Title: Studies in the assessment of vegetation impact in the urban context Studies in the assessment of vegetation impact in the urban context', Energy & Buildings 145, 331–341.
URL: *http://dx.doi.org/doi:10.1016/j.enbuild.2017.04.003*

Wang, B., Cot, L. D., Adolphe, L., Geoffroy, S. & Sun, S. (2017), 'Cross indicator analysis between wind energy potential and urban morphology', Renewable Energy 113, 989–1006.

Wang, L., Mathew, P. & Pang, X. (2012), 'Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building', Energy and Buildings 53, 152–158.

Wang, Z. & Srinivasan, R. S. (2017), 'A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models', Renewable and Sustainable Energy Reviews 75(October 2016), 796–808.
URL: *http://dx.doi.org/10.1016/j.rser.2016.10.079*

Warren, E. L., Young, D. T., Chapman, L., Muller, C. & Grimmond, C. S. B. (2016), 'Data Descriptor : The Birmingham Urban Climate Laboratory — A high density , urban meteorological dataset , from 2012 – 2014', Nature Scientific Data 3, 1–8.

Warren, E. L., Young, D. T., Chapman, L., Muller, C., Grimmond, C. S. & Cai, X. M. (2016), 'The Birmingham Urban Climate Laboratory-A high density, urban meteorological dataset, from 2012-2014', Scientific Data 3.
URL: *doi: 10.1038/sdata.2016.38*

Wate, P., Coors, V., Iglesias, M. & Robinson, D. (2019), 7 - uncertainty assessment of building performance simulation: An insight into suitability of methods and their applications, in U. Eicker, ed., 'Urban Energy Systems for Low-Carbon Cities', Academic Press, pp. 257–287.
URL: *https://www.sciencedirect.com/science/article/pii/B978012811553400007X*

Wate, P., Coors, V., Robinson, D. & Iglesias, M. (2016), 'Qualitative screening method for impact assessment of uncertain building geometry on thermal energy demand predictions', The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 42, 127.

Wate, P., Darren, R., Iglesias, M. & Coors, V. (2020), 'Emulation based uncertainty and sensitivity analysis for complex building performance simulation'.

Wicht, M. & Wicht, A. (2018), 'LiDAR-Based Approach for Urban Ventilation Corridors Mapping', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11(8), 2742–2751.
URL: *https://ieeexplore.ieee.org/document/8268056/*

Wilke, U., Haldi, F., Scartezzini, J. L. & Robinson, D. (2013), 'A bottom-up stochastic model to predict building occupants' time-dependent activities', Building and Environment 60, 254–264.

Wong, N. H., Kwang Tan, A. Y., Chen, Y., Sekar, K., Tan, P. Y., Chan, D., Chiang, K. & Wong, N. C. (2010), 'Thermal evaluation of vertical greenery systems for building walls', Building and Environment 45, 663–672.

Wood, M. J., Eames, M. E. & Challenor, P. G. (2015), 'A comparison between gaussian process emulation and genetic algorithms for optimising energy use of buildings', Proceedings of the Building Simulation Conference 2015 .

Yassaghi, H. & Hoque, S. (2019), 'An overview of climate change and building energy: Performance, responses and uncertainties', Buildings 9(7).

Young, A. D. (1989), 'Boundary Layers', NASA STI/Recon Technical Report A 91.

Zhang, X., Schildbach, G., Sturzenegger, D. & Morari, M. (2013), Scenario-based MPC for energy-efficient building climate control under weather and occupancy uncertainty, in '2013 European Control Conference, ECC 2013', EUCA, pp. 1029–1034.

Zhao, F. (2012), 'Agent-based modeling of commercial building stocks for energy policy and demand response analysis', (May).
URL: *https://smartech.gatech.edu/handle/1853/43704?show=full*

Zhao, H. X. & Magoulès, F. (2012), 'A review on the prediction of building energy consumption', Renewable and Sustainable Energy Reviews 16(6), 3586–3592.
URL: *http://dx.doi.org/10.1016/j.rser.2012.02.049*

Zou, P. X., Wagle, D. & Alam, M. (2019), 'Strategies for minimizing building energy performance gaps between the design intend and the reality', Energy and Buildings 191, 31–41.

Zou, P. X., Xu, X., Sanjayan, J. & Wang, J. (2018), 'Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives', Energy and Buildings 178, 165–181.

Appendix A

Birmingham University Climate Laboratory (BUCL) data

A

The Birmingham University Climate Laboratory project Warren, Young, Chapman, Muller, Grimmond & Cai (2016). The sensor network was originally designed to assist in modelling and examining urban heat islands, such as their intensity, heterogeneity and extent.

From the perspective of working with data-set the data is well formatted and in .csv is easy to read into a variety of platforms without the need for more specialist packages and software. The completeness of the data does lead to some challenges in terms of which periods is best to do inference studies in chapters 5 and 6. The commentary made about the sensor being prone in winter months to producing repeated readings due low power. This seems to correlate with excessive numbers of 0°C values found in the data. Some data cleaning was carried out to remove repeated values and value where the data flag doesn't appear to correctly identify completeness of the data. The approach adopted was to remove values that appeared suspicious rather than try to correct them.

Sensor Deployment

The sensor were mostly placed within the boundaries of Birmingham. A total of 7 stations were placed in rural location to provide contrasting background observations. The sensors were placed at 3 ± 0.5m in slightly higher than normal to facilitate security concerns and the standardisation across sites. Most were located in school grounds which have varied surroundings between light industrial, dense urban and suburban. Metadata was collected in accordance with the UMN metadata protocol Muller et al. (2013).

Date of observation | year	Date of observation | month	Date of observation | day	Time of observation | hours UTC	Time of observation | minutes UTC	Air Temperature | degC	Dew point Temperature | degC	Relative Humidity | percent	Station Pressure | hPa	Mean Sea Level Pressure | hPa	Solar Radiation | W m-2	Rainfall Total | mm	Rainfall Rate | kg m-2	Wind Speed | m s-1	Wind Direction | deg	Maximum Wind Gust | m s-1	Hail | Hits cm2	Data Quality Control Flag |
2014	1	1	0	0	-999.0	-999.00	-999.0	-999.0	-999.00	-999.00	-999	-999	-999.0	-999	-999.0	-999	9
2014	1	1	0	1	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	2	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	3	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	4	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	5	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	6	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	7	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	8	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	9	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	10	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	11	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	12	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	13	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	14	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	15	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	16	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	17	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	18	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	19	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	20	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	21	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	22	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	23	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	24	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2
2014	1	1	0	25	6.1	3.06	80.8	982.7	999.66	0.84	-999	-999	0.7	148	1.8	-999	2

Table A.1: Example of data in the BUCL data-set from a WXT station.

Date of observation | year	Date of observation | month	Date of observation | day	Time of observation | hours UTC	Time of observation | minutes UTC	Air Temperature | degC	Data Quality Control Flag |
2012	10	18	0	1	10.30	0
2012	10	18	0	2	10.29	0
2012	10	18	0	3	10.26	0
2012	10	18	0	4	10.28	0
2012	10	18	0	5	10.28	0
2012	10	18	0	6	10.26	0
2012	10	18	0	7	10.25	0
2012	10	18	0	8	10.17	0
2012	10	18	0	9	10.22	0
2012	10	18	0	10	10.20	0
2012	10	18	0	11	10.19	0
2012	10	18	0	12	10.21	0
2012	10	18	0	13	10.16	0
2012	10	18	0	14	10.15	0
2012	10	18	0	15	10.16	0
2012	10	18	0	16	10.13	0
2012	10	18	0	17	10.09	0
2012	10	18	0	18	10.07	0
2012	10	18	0	19	10.07	0
2012	10	18	0	20	10.06	0
2012	10	18	0	21	10.01	0
2012	10	18	0	22	10.05	0
2012	10	18	0	23	10.01	0
2012	10	18	0	24	9.98	0
2012	10	18	0	25	10.01	0
2012	10	18	0	26	10.04	0

Table A.2: Example of data in BUCL data-set from a ASM station

Figure A.1: Weather station locations taken from Warren, Young, Chapman, Muller, Grimmond & Cai (2016). ASM stations often in schools are Aginova Sentinel Micro (ASM) air temperature sensors. The WXT sites are Vaisala WXT520 (WXT) automatic weather stations, coupled with a Skye Instruments SKS1110 pyranometer, at minute resolution, with an average spacing of 3 per km2, details taken from Warren, Young, Chapman, Muller, Grimmond & Cai (2016)
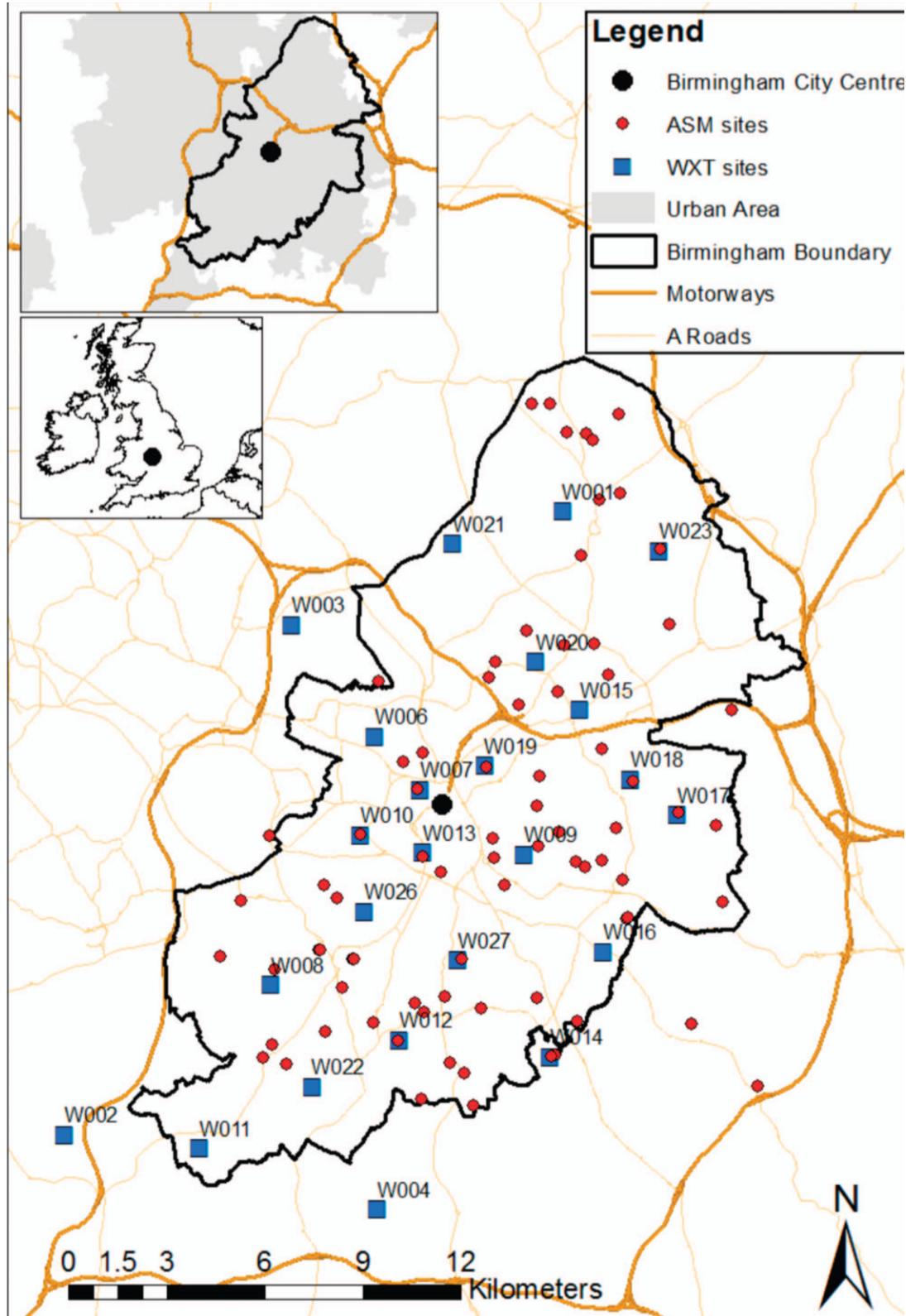
Figure A.2: Activity of WXT stations for BUCL data-sets, it should be noted that within these periods there are issues with data completeness. For more details see the plots in figure 2 on page 5 or 6 in paper Warren, Young, Chapman, Muller, Grimmond & Cai (2016)