Small-Scale Wind Energy: Methods for Wind Resource Assessment

Shemaiah Matthias Weekes

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

as part of the integrated PhD/MSc in Low Carbon Technologies

The University of Leeds

Energy Research Institute (ERI)

School of Process, Environmental and Materials Engineering (SPEME)

Centre for Doctoral Training in Low Carbon Technologies

January 2014

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 4 and **Chapter 5** include the following published works:

S.M. Weekes and A.S. Tomlin, *Evaluation of a semi-empirical model for predicting the wind energy resource relevant to small-scale wind turbines*. Renewable Energy, 2013. 50: p. 280-288.

S.M. Weekes and A.S. Tomlin, *Comparison of low-cost wind resource assessment tools for small-scale wind energy installations*. Scientific Proceedings of EWEA 2013, Vienna, 4th – 7th February 2013.

S.M. Weekes and A.S. Tomlin, *Low-cost wind resource assessment for small-scale turbine installations using site pre-screening and short-term wind measurements*. IET Renewable Power Generation (in press), 2014.

Chapter 5 includes the following published work:

S.M. Weekes and A.S. Tomlin, *Data efficient measure-correlate-predict approaches to wind resource assessment for small-scale wind energy*. Renewable Energy 2014. 63: p. 162-171.

Chapter 7 includes the following submitted work:

S.M. Weekes and A.S. Tomlin, *Comparison between the bivariate Weibull probability distribution and linear regression for assessment of the long-term wind energy resource using measure-correlate-predict*. Renewable Energy (under review), 2014.

The original work contained within these papers is all the candidate's own work with guidance provided by A.S. Tomlin.

Acknowledgements

Completing this thesis has required great personal effort and determination but there is no doubt that it would not have been possible without the support of many dedicated people who have assisted along the way. First and foremost, I would like to thank my primary supervisor Professor Alison Tomlin for providing direct and timely feedback and always demanding the highest of standards in all aspects of my work. I would also like to sincerely thank my second supervisor Professor William Gale for being available whenever I needed expert advice, direction and anecdotes. I am very grateful to the management committee of the Doctoral Training Centre, particularly Professor Paul Williams and James McKay for creating a stimulating and supportive research environment. Thanks must also go to my collaborators at the UK Met Office, Professor Simon Vosper, Alasdair Skea, Mark Gallani and Dr David Thomson for providing valuable advice as well as extensive access to operational forecast data. In addition, I am thankful to the British Atmospheric Data Centre for providing access to meteorological surface observations and RenewableUK and the Energy Saving Trust for providing access to data from the small-scale wind turbine field trial.

On a personal note I am deeply grateful for the professional and personal support of my many colleagues and friends within the Energy Research Institute at the University of Leeds. In particular I wish to thank Philippa Hardy for her indomitable positive spirit and all those who were there from the beginning of this journey, Joel Millward-Hopkins, Zarashpe Kapadia, Sam Pickard, Gillian Harrison, Hannah James and Tom Lynch. I am also grateful to my fellow PhD students of the Low Carbon Technologies Doctoral Training Centre, of whom there are too many to mention, for providing the most inspirational and supportive environment I have ever worked in. Finally, I would like to thank my parents, without whom this story would never have begun.

Abstract

Small-scale wind energy is a renewable energy technology with exciting prospects in a low carbon energy future. However, in order for the technology to be fully utilized, techniques capable of predicting the wind energy resource quickly, cheaply and accurately are urgently required. Specifically, the direct measurement approaches used in the large-scale wind industry are often not financially or practically viable in the case of small-scale installations.

The subject of this thesis is the development of low-cost, indirect methods for predicting the wind resource using, (i) analytical models based on boundary layer meteorology and (ii) data-driven techniques based on measure-correlate-predict (MCP). The approaches were developed and tested using long-term (11 years) wind data from meteorological stations, short-term (1 year) data from a field trial of small-scale turbines, and output from an operational forecast model.

As a first step, the performance of an existing boundary layer scaling model was evaluated at 38 UK sites and found to result in large site-specific errors. Based on these findings, a revised model was developed and shown to improve prediction accuracy. However, uncertainty analysis and comparison with onsite measurements revealed average errors in the predicted wind power density of over 60% due to uncertainties in the model input parameters. Hence, it was concluded that such an approach is best applied in a scoping context to identify sites worthy of further study.

To investigate the ability of low-cost, data-driven techniques to reduce these uncertainties, MCP approaches were trialled using onsite measurement periods as short as 3 months at a subset of 22 of the above UK sites. In addition to established linear approaches, non-linear Gaussian process regression and bivariate conditional probability approaches were developed. Using a 3 month measurement period, the best performing MCP approaches resulted in average errors in the predicted wind power density of 14%, compared to 26% when using the boundary layer scaling approach at the same sites. The effect of seasonal variability in the prediction errors was investigated in detail and found to be most significant at coastal sites. This variability was found to be reduced by using output from an operational forecast model in place of long-term reference wind data.

This work provides a means for low-cost and rapid wind resource assessment in cases where traditional approaches are not viable.

Contents

List of Tables

List of Figures

Glossary of Terms

Roman symbols:

c	Weibull shape factor
C_p	Power coefficient
d	Zero plane displacement height
d_{eff}	Effective displacement height
d_{local}	Local displacement height
d_{reg}	Regional displacement height
D	Bivariate Weibull association parameter
$f(u)$	Probability distribution function
$F(u)$	Cumulative probability distribution function
$\bar{\mathbf{f}}_*$	Mean function values at test points $\mathbf{x}_*$
$\mathcal{GP}$	Gaussian process
h, h_m	Canopy height, mean height of roughness elements
k	Weibull shape factor
k_{ass}	Assumed Weibull shape factor
k_{obs}	Observed Weibull shape factor
$k(\mathbf{x}, \mathbf{x}')$	Covariance function
$\mathbf{K}(\mathbf{X}, \mathbf{X}')$	Covariance matrix
L	Likelihood function
Lin	Linearity parameter
lnL	Log-likelihood function
$m(\mathbf{x})$	Mean function
P	Instantaneous wind power
P_d	Instantaneous wind power density
$\bar{p}_d$	Mean wind power density

r_u	Linear correlation coefficient between reference and target site hourly averaged wind speeds
r_{WI}	Linear correlation coefficient between reference and target site monthly wind indices
S_i	Sensitivity index
u	Wind speed
u_*	Friction velocity
$\overline{u^3}$	Mean of the cubed wind speeds
$\bar{u}$	Mean wind speed
u_E	Easterly wind vector
u_N	Northerly wind vector
u_{ref}	Observed reference site wind speed
$\bar{u}_{ref}$	Observed reference site mean wind speed
u_{tar}	Observed target site wind speed
$\bar{u}_{tar}$	Observed target site mean wind speed
$\hat{u}_{tar}$	Predicted target site wind speed
WI	Wind index
$\mathbf{x}$	Vector of inputs
$\mathbf{x}_*$	Vector of test inputs
$\mathbf{y}$	Vector of outputs
z_0	Roughness length
$z_{0,local}$	Local roughness length
$z_{0,reg}$	Regional roughness length
$z_{0,eff}$	Effective roughness length
z_{bh}	Blending height
z_{hh}	Turbine hub height
z_{ref}	Reference height

Greek symbols:

α, β	Regression coefficients
Γ	Gamma function
ε	Residual error
$\varepsilon_{\%}$	Residual percentage error
$\varepsilon_{\bar{u}}$	Residual error in predicted mean wind speed
θ	Wind angle, direction from
κ	Von Karman constant
λ_f	Frontal area density
λ_p	Plan area density
ρ	Air density
σ	Standard deviation
σ_f^2	Function variance
σ_s^2	Noise variance
σ_{res}	Standard deviation of residuals
τ	Reynolds shear stress

Abbreviations:

ANN	Artificial neural network
BS	Boundary layer scaling
BW	Bivariate Weibull
cdf	Cumulative probability distribution function
CFD	Computational fluid dynamics
CL	Canopy layer
ECMWF	European Centre for Medium-Range Weather Forecasts
%Error	Absolute percentage error
EST	Energy Saving Trust
FiT	Feed in Tariff
GPR	Gaussian process regression

GUI-HDMR	Graphical user interface - high dimensional model representation
IBL	Internal boundary layer
ISL	Inertial sublayer
LCM	Land cover map
LiDAR	Light detection and ranging
LR	Linear regression
LSM	Least squares method
MAE	Mean absolute error
MBE	Mean bias error
MCP	Measure-correlate-predict
MCS	Microgeneration Certification Scheme
MERRA	Modern Era Retrospective Analysis for Research
MIDAS	Met Office Data Archive System
MM	Method of moments
MML	Method of maximum likelihood
NCEP/NCAR	National Centers for Environmental Prediction/National Center for Atmospheric Research
NCIC	National Climate Information Centre
NOABL	Numerical Objective Analysis of Boundary Layer
NWP	Numerical weather prediction
OR	Orthogonal regression
PBL	Planetary boundary layer
pdf	Probability distribution function
PSD	Power spectral density
PWT	Wind turbine power curve
Rf	Reference sites (meteorological stations)
RMSE	Root mean square error
RRA	Round robin approach

RSL	Roughness sublayer
UBL	Urban boundary layer
UK4	Met Office 4 km UK forecast model
UKV	Met Office 1.5 km UK forecast model
UM	Met Office Unified Model
VMM	Virtual Met Mast
VR	Variance ratio
WAsP	Wind Atlas Analysis and Application Program
WYET	Wind yield estimation tool

1 Introduction

Climate change, sustainable development and renewable energy are just a few of the phrases that have only recently entered common usage as it has become increasingly clear that the Earth's resources are not inexhaustible. Partly in response to this enhanced awareness and partly because it often makes good business sense, governments, organisations and individuals are starting to look for ways to reduce their impact on the environment. One way of beginning to take responsibility for one's environmental impact is through the use of small-scale, decentralised and low carbon energy sources. These are commonly referred to as microgeneration and include small-scale wind energy as well as a whole range of low carbon technologies.

Within the UK, pressures related to the progressive decommissioning of ageing nuclear and coal fired power stations, as well as volatility in primary energy prices are starting to be felt [1]. Added to this, plans for greater diversification of electricity sources [2] the 2020 target of 15% of energy from renewables [3] and the 2050 commitment to reduce CO_2 equivalent emissions by 80% compared to 1990 levels [4], present the energy industry with unprecedented challenges. A whole suite of energy technologies, including new nuclear power, carbon capture and storage and large-scale renewables will likely be needed over the coming decades. In addition to these large-scale contributions, energy efficiency measures and distributed energy sources such as small-scale wind energy must also be fully exploited [5].

Small-scale wind energy has exciting prospects in the UK due to the favourable wind resource, a growing manufacturing base of small-scale turbines and government financial incentives. Despite recent uncertainty in the policy environment, the small- and medium-scale[1] wind turbine industry has experienced rapid growth in the last decade with the UK market size now put at over £105 million [6]. Currently, only China and the USA have installed capacities of small-scale wind turbines that are greater than the UK's, in a global small-scale wind energy market that increased by 27% in 2011 [7]. However, in order for this industry to achieve its full potential, both within the UK and further afield, several issues need to be addressed.

---

[1] Definitions of scale vary worldwide. The upper limits of small- and medium-scale turbines are typically 100 kW and 500 kW rated power respectively [6,7].

One of the most critical of these challenges is the development of suitable techniques for accurate wind resource assessment. An estimate of the likely energy yield is essential for any potential customer who must choose between competing microgeneration technologies and who desires to make an investment choice that maximises financial and environmental benefits. However, predicting the potential energy yield of a small-scale turbine presents a number of specific challenges that are related to costs, timescales and accuracy [7].

1.1 The Challenge of Wind Resource Assessment

Wind resource assessment on any scale is complicated by several factors: (i) the cubic relationship between wind speed and power, (ii) the temporal variability of wind speeds on a wide range of timescales from years to seconds and (iii) the spatial variability in the mean wind speed. Due to the cubic relationship between wind speed and wind power, small changes in wind speed can result in large changes in wind power, necessitating highly accurate wind speed predictions. Additionally, due to the temporal and spatial variability in the wind flows, long-term onsite wind measurements at the location of the proposed turbine site are generally required. These measurements are used to produce frequency distributions of wind speeds averaged over a predefined time period, from which, long-term statistical averages can be obtained. These statistics, possibly along with estimates of other atmospheric variables, are used to define the average characteristics of the wind resource at a particular site. This information, with the addition of a specified wind turbine power curve, can be used to predict the average energy yield for a specific wind turbine and location. Assuming the performance of the turbine has been well characterised, the issue of wind resource assessment ultimately becomes one of defining the statistical averages which describe the wind flow at a specific location. However, in the case of small-scale wind installations, this is non-trivial to achieve.

The most direct way to estimate the potential wind resource is to use onsite anemometry to make long-term measurements. For large-scale wind farms, which involve long-term planning and very large investments, this typically involves collecting 1-3 years of onsite data in order to obtain sufficient statistics to define reliable averages [8]. For small-scale wind installations, however, these timescales are often not practical and the impact of such a measurement campaign on the total investment cost makes the approach unrealistic [7]. In the absence of long-term measurements, a potential customer must resort to indirect methods of estimating the potential energy yield. To achieve widespread uptake of small-scale wind

turbines, these indirect methods should be relatively quick and cheap to implement, applicable to a broad range of sites without detailed, site-specific modelling, and accurate enough to inform investment decisions.

1.1.1 Wind atlas methods

Given a specific location, a wind atlas may be used to quickly and easily estimate the uninterrupted mean wind speed on a predefined grid. For the UK, wind atlases such as NOABL (Numerical Objective Analysis of Boundary Layer) [9] and NCIC (National Climate Information Centre) [10] have been developed using interpolation of historical surface wind speed measurements with a resolution 1 km^2. Despite the superior performance reported for NCIC [10], NOABL has gained widespread popularity due to the fact that it is publically available. However, both NOABL and NCIC are unable to account for local obstacles or spatial variability in surface roughness and topography on length scales smaller than 1 km and this can significantly affect the predicted wind resource. Unsurprisingly, it has been shown that NOABL frequently overestimates the mean wind speed, particularly in areas of high surface roughness [11] and hence cannot be used without the addition of correction factors that attempt to account for local effects [12, 13]. As will be demonstrated later, these correction factors are themselves a source of considerable uncertainty.

1.1.2 NOABL-MCS

Within the UK, small-scale wind turbine installations are eligible for financial support under a feed-in-tariff system. To qualify for this support, a turbine and its installation must conform to certain standards mandated in the Microgeneration Certification Scheme (MCS). This includes an Installation Standard that sets out a standardised method for estimating annual energy production at a site [13]. The method involves applying a correction factor to the NOABL mean wind speed based on an estimate of the local terrain and the presence of nearby obstacles, the so-called NOABL-MCS method. The installer is obliged to communicate this estimate to the customer along with a number of caveats including, "[*This estimate*] *is given as guidance only and should not be considered to be a guarantee*" and "*The energy performance of wind turbine systems is impossible to predict with a high degree of certainty due to the variability in the wind from location to location and from year to year*." [13]. While standardised procedures for estimating annual energy production are very welcome, these statements alone highlight the challenge of making accurate predictions using simple correction factors. Application of the NOABL-MCS method to a number of sites detailed later in this thesis shows that the method often results

in predictions that are too conservative, thus excluding potentially viable sites from development.

1.1.3 Boundary layer scaling

Using a wind atlas as a starting point, the principles of boundary layer meteorology may be applied to provide more precise predictions of the spatially averaged mean wind speed compared to those obtained from simple correction factors. Boundary layer scaling methods attempt to account for the effect of local and regional roughness using simple parameterisations of the surface characteristics [14, 15]. Such methods are attractive since after their development, they can be deployed rapidly at multiple sites with little investment. However, there are large uncertainties involved at each stage of the scaling process, resulting in final wind speed predictions that frequently do not reach the required accuracy [16]. Some improvements may be achieved in the case of building mounted turbines or in very complex terrain through coupling scaling methods to detailed flow models using computational fluid dynamics [14, 17]. While these approaches can result in general siting recommendations, they generally require site-specific modelling which negates some of the advantages of a simple scaling approach. In addition, scaling methods that use wind atlas inputs only produce mean wind speed predictions rather than predictions of the full wind speed distribution and this results in further uncertainties in the predicted wind power.

1.1.4 Data-driven techniques

The starting point for the approaches mentioned above is generally some form of spatially averaged mean wind speed. In contrast, data-driven techniques make use of wind data in the form of a time series, either measured or forecast, at the location of the turbine site. Data-driven approaches are required in the large-scale wind industry in order to provide sufficient confidence to satisfy investors. Typically, in addition to long-term (> 1 year) onsite measurements, correlation to a nearby reference site is sought using an approach known as measure-correlate-predict (MCP) to provide predictions over several decades [18]. In the small-scale wind industry, however, the time and expense of a 12 month measurement campaign make such an approach the exception rather than the norm. Thus, while MCP offers an opportunity for increasing the value of any short-term measurement campaign, it has not been widely studied in the context of the small-scale wind industry where measurement periods are likely to be significantly less than the recommended 12 months.

Related techniques may use forecast wind data in place of onsite measurements. The UK Met Office have recently developed a tool coined the 'Virtual Met Mast™' (VMM) [19]. The VMM applies local corrections to long-term forecast data in order to produce a time series of predicted wind speeds at a specific site without the need for onsite measurements. The VMM is a promising development but at present it has been optimised for heights somewhat above those where small turbines are likely to operate in order to capture the medium- to large-scale wind market. Hence, the current implementation of the VMM does not perform well at heights below 30 m, the very region where many small turbines will operate [20].

1.2 Why Research is Needed

As demonstrated above, the accuracy of wind resource assessment in the small-scale wind industry is constrained by the practical limitations of timescales and costs as well as the technical limitations of currently available tools. These factors have contributed to the lack of a unified and consistent approach to wind resource assessment, and not infrequently, practices that would be intolerable to investors in large-scale wind projects. Inevitably, this has resulted in the installation of turbines at locations that are not viable [11], and most likely, the underdevelopment of viable sites due to insufficient information or lack of investor confidence.

The need for suitable wind resource assessment tools has been highlighted by several field studies. A field trial of 26 building mounted small wind turbines, known as the Warwick Wind Trials [12], investigated turbine performance in a variety of locations. The study highlighted the challenge of accurately predicting the potential wind resource in complex environments using simple methods such as NOABL and the importance of choosing appropriate locations. Following the Warwick Wind Trials, in 2009 the Energy Saving Trust [11] completed a larger field trial of both pole and building mounted small wind turbines in a variety of locations throughout the UK. While the results showed a promising potential for turbines installed at appropriate locations, the study highlighted the underperformance of many installations and the critical importance of identifying appropriate locations before installation [11, 21, 22].

Ideally, the small-scale wind industry should aim to emulate the rigour of the large-scale wind industry in applying consistent and reliable approaches to wind resource assessment in advance of installation. To achieve this, however, low-cost tools are required that are capable of assessing the long-term wind resource at potential turbine locations both rapidly and accurately. Such tools should be generalised so

that they are applicable at multiple site types without the need for highly detailed, site-specific investigations and the associated consultancy costs.

While this may appear to be a formidable challenge, it should be noted that due to the lower investment costs, the development of rigorous and consistent techniques does not necessarily imply the same strict accuracy requirements of industrial wind farms. At the very minimum, however, the resource assessment procedure should be accurate enough to determine whether the site is viable (likely to repay the embedded carbon and financial investment) as well as giving an estimate of the likely uncertainties. In the case of larger investments, a more precise estimate of the expected energy production and financial yield should be sought.

1.3 Aims and Objectives

The overall aim of this thesis is the development of tools to rapidly, accurately and cost-effectively assess the wind resource at potential small wind turbine sites. The term ‘wind resource’ is used here to represent the average wind power density and associated statistical parameters before the application of a specific wind turbine power curve. This overall aim is pursued using a two-pronged approach, (i) development of boundary layer scaling techniques that require only simple parameterisations of the surface and (ii) development of data-driven MCP techniques that provide greater accuracy. It is believed that this two-pronged approach offers utility and flexibility since it allows the resource assessment technique to be tailored to the specific requirements of the project including the available time and funds.

To be of most use, wind resource assessment tools should be widely applicable and not limited to a particular terrain type or local environment. While there is now a growing body of literature related to micro-scale wind energy in the urban environment, and building mounted turbines in particular [23], the current study includes, but is not limited to, this particular application. In many cases, urban wind turbines face very specific challenges and these may require specialised treatment. Thus, the approaches developed in this thesis try to balance specificity with generality.

In developing generalised wind resource assessment approaches, particular attention has been given in this work to the practical constraints encountered in the small-scale wind industry. Specifically, these are short planning and development times, and limited onsite wind measurements. Hence, the work in this thesis is not

limited to a particular turbine size or definition of ‘small-scale’ but is applicable wherever these constraints are encountered.

The approaches are developed and tested using long-term wind data obtained from meteorological stations throughout the UK, predominantly located at 10 m above ground level. Hence, the wind data are representative of the actual meteorological conditions experienced in locations where small-scale turbines are required to operate. This facilitates realistic appraisals of the performance of the approaches and provides useful information regarding their limitations.

The main objectives of this study may be summarised as follows:

- Evaluation of a currently available boundary layer scaling approach to wind resource assessment using long-term measurements at a wide range of UK sites.
- Identification of the limitations in current approaches and the implementation and testing of an improved methodology.
- Investigation of the feasibility of using data-driven MCP approaches using short-term onsite measurements covering less than 12 months.
- Development and testing of new short-term MCP approaches based on statistical and physical considerations.
- Application of forecast data to the data-driven and boundary layer scaling approaches in an attempt to reduce the reliance on wind atlases or onsite measurements.

1.4 Outline of Thesis

Chapter 2 begins with a brief overview of small-scale wind energy including an explanation of the term ‘small-scale’ and a description of current market trends. This is followed by a summary of fundamental concepts related to wind energy in general and wind resource assessment in particular. Next, an overview is given of pertinent boundary layer processes that affect wind flows close to ground level and the complications arising from different terrain types are discussed.

In Chapter 3, specific methods for wind resource assessment using physical and statistical approaches are considered. Firstly, a detailed review is given of the boundary layer scaling methodology employed in Chapter 4. Secondly, a range of data-driven MCP approaches are reviewed in preparation for their application and further development in Chapters 5-8.

In Chapter 4, an existing boundary layer scaling model for predicting the spatially averaged mean wind speed is considered in detail. The performance of the approach is evaluated quantitatively using observed wind data at 38 UK sites in a variety of terrains. Based on this appraisal, an improved methodology, including an estimation of uncertainties, is proposed and evaluated at the same sites. Finally, a sensitivity analysis is applied to identify the relative importance of the model input parameters and to make recommendations for future improvements.

Data-driven approaches are first applied in Chapter 5 where linear MCP techniques are used to predict the 10 year wind resource using multiple 3 month measurement periods at 22 UK sites. The linear models are enhanced through the use of a Gaussian scatter model and this is shown to significantly improve performance. A sliding window technique is developed to obtain average error statistics over multiple training and test periods and these are used along with the error distributions across sites to assess performance. Seasonal effects are considered in detail and the implications of these for the choice of measurement season are highlighted. The core methodology developed in this chapter (short-term onsite measurements and data selection using a sliding window technique) is used in subsequent chapters to provide a consistent framework for comparing the performance of different data-driven approaches.

Chapter 6 extends the work of Chapter 5 through the development of a non-linear, Bayesian approach to MCP using Gaussian process regression. The performance of the approach applied to short onsite measurement periods is compared to linear regression methods and correlation statistics are used to interpret the observed error trends. The approach is further extended to predict time series of wind velocity vectors thus facilitating prediction of the long-term distribution of wind directions at the potential wind turbine sites.

In Chapter 7, a bivariate Weibull MCP approach is developed and training periods of multiple lengths are considered. In part, this approach is inspired by a desire to improve on the Gaussian scatter model that was used in Chapter 5. Building on previous work that applied the approach to artificial wind data, real wind speed observations are supplemented by artificial data drawn from idealised Weibull distributions to demonstrate the additional challenges associated with real-world applications. The approach is used to predict the long-term wind resource using onsite measurement periods of 1-12 months with particular attention given to short measurement periods. The approach is compared with baseline linear methods and the implications for real-world site assessments are highlighted.

In Chapter 8, attention is turned to the possibility of using forecast data as an alternative to long-term wind observations. Output from the UK Met Office Unified Model is applied to wind resource assessment using both MCP and boundary layer scaling approaches. The data are first investigated in terms of their ability to represent long-term trends at the test sites as evaluated by a wind index correlation parameter. Next, the data are used in place of long-term reference observations in an MCP approach and error metrics are compared to those obtained using long-term observed data. Finally, the boundary layer scaling approach developed in Chapter 4 is applied to the forecast time series and the resulting wind resource predictions are compared to those obtained when using conventional wind atlas input.

Chapter 9 outlines the overall conclusions from this work and identifies opportunities for future studies.

2 Background and Fundamentals

This chapter is organised as follows: Section 2.1 gives an overview of small-scale wind energy including a definition of the term ‘small-scale’, the current market trends and the environmental and financial viability of small-scale installations. In Sections 2.2 and 2.3, an outline of the fundamental equations and statistical distributions required for describing the wind resource is given. Finally, Section 2.4 provides an overview of the boundary layer processes relevant to wind resource prediction. Particular attention is given to the effect of a change of roughness on the boundary layer wind flow and the additional complexities arising from coastal and urban terrains are described.

2.1 Overview of Small-Scale Wind Energy

2.1.1 What is small scale?

Within the wind industry, there is no strict definition of the term ‘small-scale’ and the size of small turbines ranges from less than one kilowatt to hundreds of kilowatts. The UK government defines microgeneration as < 50 kW [5] and although the term ‘micro’ is somewhat misleading for wind turbines of 50 kW, this definition is a useful starting point. Definitions and categorisations are also changing in response to market trends. The latest classification from UK trade body, RenewableUK, groups medium and small wind together in a three tier system shown in Table 2.1.

	Power (kW)	Height (m)	Annual energy (homes equivalent)
Micro	0-1.5	10-18	< 0.24
Small	1.5-50	15-35	< 47
Medium	50-500	25-55	< 426

Table 2.1: The RenewableUK classification of micro to medium turbine sizes, adapted from [6]. Homes equivalent energy production is based on average domestic electricity consumption of 4227 kWh/annum [24].

It is noteworthy that the hub heights of small turbines are generally between 10 and 35 m. Since these heights are substantially lower than industrial turbines, the wind resource is likely to be more sensitive to the effects of the local surface. Hence, wind resource prediction methodologies developed for the large-scale wind industry may have limitations when applied to potential small-scale sites.

2.1.2 Current trends

Latest (2013) industry figures [6] show that since a peak in 2007, there has been a steady decline in the number of micro (0-1.5 kW) wind turbines installed in the UK and a subsequent move towards small and medium turbines which have much greater generating potential. This shift has also been reflected in the installed generation capacity with small and medium wind systems now dominating despite the smaller number of units compared to micro turbines. In addition, there has been a sharp decrease in the installation of building mounted turbines, with just 16 installations in 2011 and 2012 combined, compared to a peak of 1054 in 2007. Similar upward trends in the size of small wind turbines have been reported globally [7], although this trend seems to be somewhat stronger within the UK.

As stated in Chapter 1, the work reported in this thesis is not restricted to a particular definition of small-scale. However, the decision to focus on generally applicable resource assessment tools in a range of terrains is somewhat justified given the shift away from micro and building-mounted turbines towards larger, free-standing installations. In addition, the move towards larger turbines implies greater investment costs and associated risks. Hence, this provides a stronger case for the use of at least some onsite measurements and more detailed wind resource assessment procedures.

In general, the amount developers will be willing to invest in wind resource assessment will be related to the total cost of the wind turbine as this is related to the overall risk. These costs are highly variable due to site-specific installation issues and non-linearities between turbine size and price. Based on a sample (rated power 3-20 kW) of currently available MCS certified UK turbines, the installed cost is likely to be in the range of £3000-£5000 per kW, although non-certified and micro turbines may be somewhat cheaper. The World Wind Energy Association [7] quotes an average installed cost of around $6000 (~ £3800) per kW for USA installations and a recent European study [25] put the average at €3900 (~ £3260) per kW. While these are very broad estimates, they provide some perspective in estimating the level of funds likely to be available for assessing the wind resource.

For example, the VMM wind resource assessment product recently launched by the Met Office quotes a current (June 2013) price of £840 for a single site report based on forecast data, rising to £1740 for a more detailed report including corrections from onsite measurements. For a mid-priced 15 kW turbine with an investment cost of £60000, this implies a cost of 1.4 - 2.9% of the total investment, although this percentage will of course be sensitive to the proposed turbine size. Ultimately, the

amount an investor will be willing to pay will depend on many factors, including (i) the degree of risk they are willing to entertain, (ii) their understanding of the accuracy of different methodologies and (iii) the availability of wind resource assessment services through a turbine retailer or external consultancy. However, up to a few percent of the total investment cost is a reasonable first estimate.

2.1.3 Environmental viability

Frequently, a key motivating factor for investing in a small-wind system is to reduce carbon emissions, hence, it is vital to know if small-scale wind can be justified from an environmental perspective. In order to be environmentally viable, a wind turbine must result in sufficient carbon savings during its lifetime to replace the carbon released in its manufacture, transport, operation and disposal. Carbon savings are achieved by displacing electricity produced by more carbon intensive processes. To establish viability, two main inputs are required, (i) a cradle-to-grave audit of carbon emissions and (ii) a predicted energy yield over the lifetime of the turbine. A number of such studies have been undertaken.

To understand the results of these studies it is useful to introduce the concept of the capacity factor, which is the fraction of power delivered by a turbine compared to its rated power. This is defined as:

$$\text{Capacity Factor} = \frac{\text{Average power delivered}}{\text{Rated power}}$$

Equation 2.1

Capacity factors are generally calculated on a yearly basis from the actual energy yield and provide a measure of the performance of a turbine in a specific location. Well located large-scale turbines typically have a capacity factor of 30-40% [26] but this figure can be significantly lower for small-scale turbines [11]. The capacity factor is a useful metric when considering environmental and financial viability since it allows turbines of varying sizes installed in different wind regimes to be compared. As a note of caution, the capacity factor is not on its own a rigorous performance metric since it will also depend on the turbine's rated power which is not always appropriately defined. An artificially high rated power (i.e. power output at an unrealistically high wind speed) will lead to an artificially low capacity factor and vice-versa.

Allen *et al*. [27], carried out a detailed lifecycle assessment of a 600 W turbine along with energy yield estimates based on wind speed measurements at a height of 10 m in a variety of rural and urban areas. For rural areas this can be considered

the height above ground level while for urban areas the height is not specified although the authors state that the resource represents that available to building mounted turbines. Met Office wind speed statistics were used for a total of 26 sites and a correction was included for turbulence based on whether the location was rural or urban. Interestingly, it was found that the component with largest environmental burden (carbon emissions as well as other pollutants) was the building mounting (usually aluminium) or the scaffold pole in the case of pole mounted turbines.

In urban areas, even for the lowest measured wind resource, it was found that the carbon payback did not exceed 15 yrs. However, the impact of heavy metal pollution was only compensated for in urban locations with the highest wind resource (mean wind speed 5.2 ms^{-1}). In rural areas, all pollutants were compensated for within the estimated 15 year turbine lifetime even with the lowest measured wind resource. Overall, the displaced energy payback period (i.e. taking into account the primary energy input for grid electricity) was found to be 3 years for the average urban turbine and 0.6 years for the average rural turbine. These results were based on calculated capacity factors of 3% and 17% for the average urban and rural turbine, respectively.

A similar study was carried out by Celik *et al.* [28] who investigated the life-cycle emissions of a larger 7.5 kW turbine installed as an off-grid system with battery power storage. Using wind speed measurements from urban areas of Turkey, an energy payback time of 1.4 years and a carbon payback time of 0.7 years was calculated based on energy displaced from the average European energy mix. Rankine *et al.* [29] carried out a study of a commercially available 1.5 kW rooftop wind turbine and calculated energy payback times of 13-50 months and carbon payback times of 10-39 months based on capacity factors of 8-31%.

Clearly the environmental viability will depend on a wide range of factors including the materials used in the turbine manufacture, the degree of recycling and most crucially, the wind resource. However, the studies noted above indicate that even in the most challenging cases, such as a building-mounted turbine in an urban area, a net environmental benefit will generally be achieved within the turbine's lifetime. As an added caveat, if the wind speed is frequently lower than the turbine cut-in speed (typically around 3-4 ms^{-1}) the turbine will never achieve a net environmental benefit. While this is unlikely in rural areas, it is a real possibility close to roof height in urban areas, highlighting the need for an accurate wind resource assessment prior to any installation.

2.1.4 Financial viability

The question of financial viability is generally much harder to address than environmental viability. This is because the wind turbine market, the electricity market and UK Government policy are in all in a period of rapid flux which is having major implications for financial assessments.

Given the relative infancy of the small turbine market and the related economies of scale, it is unlikely that the majority of installations would be financially viable without Government support [30, 31]. The major UK support mechanism for small-scale wind energy is the Feed in Tariff (FiT) for small-scale renewables. Following recent price reviews, the current (January 2014) value of the tariff for small-scale wind is a flat rate of 21.65 p/kWh for installations up to 100 kW. This tariff is paid for every kilowatt hour of electricity generated with an additional 4.64 p/kWh paid for electricity exported to the grid [32]. The level of financial support will clearly have a strong influence on what constitutes a viable installation. A European study [25] of five EU countries (not including the UK) calculated minimum viable mean wind speeds for 10 small wind turbines based on a 15 year payback period. The results indicated that in some cases mean wind speeds as low as 3 ms^{-1} were viable but this value was strongly dependent on the turbine cost and the level of support through FiTs.

Within the UK, viability studies have also been reported based on the findings of a large-scale field trial initiated by the Energy Saving Trust (EST) [11]. For example, James *et al*. [21] reported that a 1 kW roof-mounted turbine, with a purchase cost of £1500, would only achieve financial payback within 10 years if it operated with an average capacity factor of 6%. This study included the additional revenue available under the FiT using a previous, higher generation tariff. Of the roof-mounted turbines included in the trial, only those mounted in rural areas were able to achieve this level of performance. Due to reduced wind speeds and poor siting decisions, roof mounted turbines in urban areas operated with capacity factors closer to 3% [11].

A related study by Sissons *et al*. [22] of pole mounted turbines showed much more promising results. A 6 kW turbine was considered at an investment cost of £20000 and an operating cost of £400 per annum. Including FiT revenues, the financial payback time was shown to be 12 years with a capacity factor of 17%, which is shorter than the period for a MW offshore wind turbine. Increasing the capacity factor to 25% reduced the payback time to 7 years, which is equivalent to the period for a MW onshore wind turbine. Clearly the capacity factors that can be achieved

will depend on a number of factors including: the rated power of the turbine, the wind speed and hence the height of the pole mounting and the nature of the surface. Encouragingly however, the EST field trial indicated average capacity factors of 19% for pole-mounted turbines, while at the best sites, vales of greater than 30% were achieved [11]. Given that the estimated lifetime of a wind turbine is typically 20-25 years, these results indicate that a small-scale pole-mounted turbine can not only payback the initial investment but also provide significant future income if it is appropriately sited.

These studies indicate that given the current market conditions, small-scale wind turbines can be financial viable, particularly in the case of pole-mounted turbines which can achieve payback times comparable to commercial wind. A further general point is worthy of note. The energy yield of a turbine is non-linearly related to the radius of the turbine blades and the wind speed. Hence it is possible to achieve large increases in energy yield by (i) installing turbines which are as large as possible, and (ii) mounting turbines as high as possible above local obstructions so as to access higher wind speeds. These two factors favour the installation of pole mounted turbines in rural/exposed areas or roof mounted turbines mounted above large, tall buildings where both the turbine size and wind speed may be maximized. In contrast, relatively small turbines mounted above the roofs of urban domestic properties are likely to be particularly challenging to justify.

2.2 The Wind as an Energy Source

Wind energy is ultimately a renewable solar energy source. Synoptic flows are driven by differential solar heating of the planet which causes air columns to rise or sink depending on their latitude. These flows combined with the Coriolis effect are responsible for the large-scale, regular wind patterns observed in different regions around the Earth [33].

At distances of tens of meters from the Earth's surface, within the range of heights of small-scale wind turbines, frictional forces become particularly important causing large changes in the magnitude, direction and turbulence of the wind flow. It is on these length scales that predictions become complex and dependent on the exact nature of the surface [15]. It is vital to account for these effects when modelling wind flows relevant to small-scale wind turbines in order to make accurate predictions of the available wind energy resource.

2.2.1 Theoretical power in the wind

In simple terms, the kinetic energy in the wind is converted to electrical energy by rotation of the turbine blades which in turn produce a rotating magnetic flux in the turbine generator. The equations describing the energy and power in the wind can be formulated by applying basic physical principles [26]. The kinetic energy in the wind is given by the expression:

$$E_k = \frac{1}{2}mu^2$$

Equation 2.2

where E_k is the kinetic energy in the wind, m is the mass of air and u is the wind speed. This expression can be used to calculate the wind power available to a wind turbine whose blades intercept an area A:

$$P = \frac{1}{2}\rho A u^3$$

Equation 2.3

where ρ is the density of air.

Hence the available power depends on three factors, the density of the air, the area intercepted by the turbine blades and the cube of the wind speed. This cubic dependence is a crucial factor in determining a turbine's performance at a particular site since a relatively small shift in wind speeds can have a large effect on the power generated. As the power in the wind does not depend linearly on the wind speed, it is necessary to know the shape of the wind speed frequency distribution in addition to the mean wind speed, in order to estimate the wind resource accurately.

2.2.2 Real power in the wind

In reality, the power extracted from the wind never reaches the value indicated by Equation 2.3. To fully extract all the available energy the wind velocity would have to be zero immediately downstream of the turbine blades. Instead, the wind speed is simply reduced after passing through the blades due to the extraction of kinetic energy by the turbine. Since the mass flow rate must remain constant both upstream and downstream of the turbine, the area of the air stream expands after passing through the blades.

This effect was described theoretically by Betz and sets a fundamental limit on the amount of energy that can be extracted from the wind regardless of the turbine

design. It is possible to represent this effect by defining a power coefficient C_p such that [34]:

$$C_p = \frac{\text{Generated power}}{0.5\rho A u^3}$$

Equation 2.4

Using momentum theory, it can be shown that the maximum possible value is C_p = 16/27 = 0.593. Hence it is never possible to extract more than 0.593 times the total power available in the wind, a figure known as the Betz limit [34].

The actually delivery of electrical power from a wind turbine will typically be considerably less than the Betz limit for several reasons. Firstly, the turbine will require a finite time to respond to changes in wind speed and direction, resulting in losses that will depend on factors such as the turbine inertia and the presence of a control system. Secondly, aerodynamic losses will occur due to non-ideal behaviour of the turbine blades. Thirdly, further losses will be incurred during conversion from kinetic to electrical energy and the process of electricity distribution. Typically a wind turbine will deliver just 30% of the total energy available in the wind as electrical energy [26] although this can vary markedly depending on the design of the turbine, control system and generator. Since the wind power extracted from a site will depend on both the available wind resource and the performance of the installed wind turbine, the Betz wind power density is a useful turbine independent metric for characterising a site. This metric represents the maximum theoretical wind power available per unit swept area and is defined as:

$$P_d = \left(\frac{16}{27}\right)\frac{1}{2}\rho u^3$$

Equation 2.5

2.3 Describing the Wind Resource

Wind speeds vary considerably on a wide range of timescales including annual, seasonal, daily, hourly, and in the case of turbulence, on timescales of seconds. In practice, wind speeds must be reported using some fixed averaging time, below which, variability in the wind flow is not represented. Figure 2.1 shows the classic representation of power spectral density (PSD) as a function of frequency as identified by Van der Hoven [35]. Low frequency oscillations (long averaging times) are associated with large-scale synoptic weather events while high frequency

turbulence (short averaging times) is associated with microscale effects. Wind resource assessment is generally carried out using averaging times between these extremes in what is termed the 'spectral gap', although in general, this may not be as clearly defined as Figure 2.1 would imply [36]. This approach allows large-scale weather events to be captured, while microscale turbulent fluctuations are smoothed out. Microscale fluctuations are of importance to the performance of small-scale wind turbines, however, treatment of these effects is beyond the scope of the current thesis.

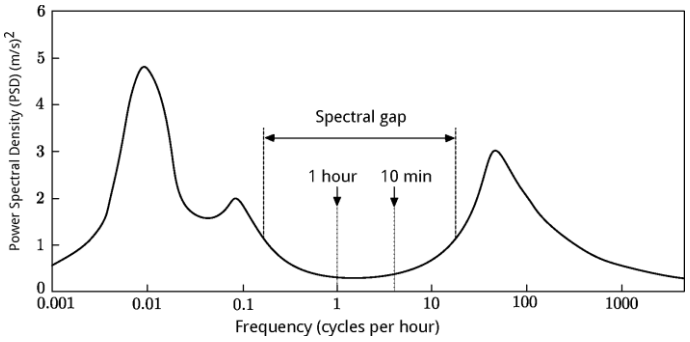

Figure 2.1: A representation of the power spectral density in wind flows as a function of frequency [hours^{-1}] [37].

Due to the inherent variability of wind flows, statistical distributions of wind speeds are required rather than simple mean values. These distributions can be used along with information regarding wind direction (and possibly turbulence) to define the average characteristics of the wind resource at a particular site. Since the power in the wind depends on the cube of the wind speed, the shape of the wind speed distribution provides crucial information regarding the available wind power. The choice of distribution is discussed in the following section.

2.3.1 The Weibull distribution

Active research into statistical distributions to represent the wind resource has been ongoing since the 1940s when targeted wind energy research programmes were carried out in the USA [38]. During this time, a vast number of statistical distributions including univariate, bivariate, multivariate, bimodal and hybrid have been proposed [38]. In 1951, Waloddi Weibull published his seminal paper [39], '*A Statistical Distribution Function of Wide Applicability*', where he described the details of an empirical distribution which came to be known as the Weibull distribution. The distribution was initially applied to failure analysis in the field of materials but in the late 1970s researchers started to apply Weibull's work to wind resource assessment [38]. Due to its empirical success in describing a wide range

of wind regimes, the Weibull distribution has now become the most widely used distribution in wind resource assessment [40, 41].

The standard Weibull distribution belongs to a class of probability distributions characterised by two parameters, a scale factor c and a shape factor k. The scale factor is related to the average wind speed while the shape factor, in effect, describes the spread of the distribution. Due to the cubic relationship between power and wind speed, the shape factor has a strong bearing on the available power. The functions describing the Weibull distribution are [34]:

$$f(u) = k\frac{u^{k-1}}{c^k}\exp\left[-\left(\frac{u}{c}\right)^k\right]$$

Equation 2.6

where $f(u)$ is the probability distribution function (pdf).

$$F(u < u_i) = 1 - \exp\left(-(u_i/c)^k\right)$$

Equation 2.7

where $F(u < u_i)$ is the cumulative distribution function (cdf) which describes the cumulative probability of observing a wind speed $u < u_i$. Note that these distributions are identically zero when the wind speed is zero and hence they are not able to represent periods of zero wind speed (lulls).

The scale factor c is related to the mean wind speed $\bar{u}$ and k by the expression:

$$\bar{u} = c\Gamma\left(1 + \frac{1}{k}\right)$$

Equation 2.8

where Γ is the gamma function.

Figure 2.2 shows examples of Weibull probability distributions for k = 1.5 - 2.3, values that are typical of UK sites [16], and a fixed mean wind speed of 5 ms^{-1}. For a fixed scale factor, low values of k result in wider probability distributions and represent wind flows where there is a large spread of wind speeds about the mean [34]. These distributions are generally associated with higher wind power due to the increased frequency of high wind speeds and the cubic relationship between wind speed and power. The opposite is generally true for high values of k.

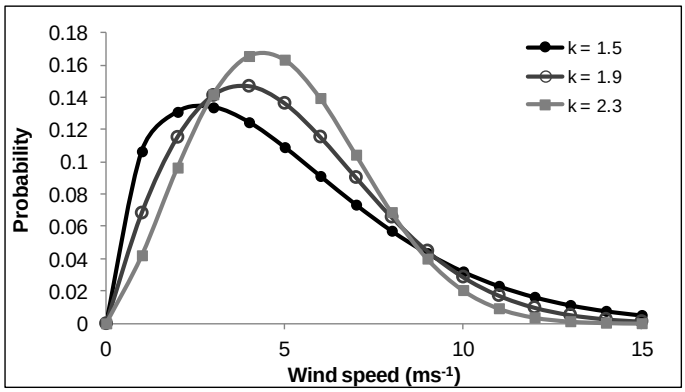

Figure 2.2: A family of Weibull distributions with the range of shape factors typically found at UK sites and a fixed mean wind speed of 5 ms^{-1}.

The defining parameters of the Weibull distribution, namely the scale and shape factors, may be obtained from a best fit to measured wind speed data using a number of well-established methods [42]. Once an appropriate fit has been achieved, the wind speed pdf can be used to obtain the mean Betz power density, $\bar{p}_d$, in the wind using the following expression [43].

$$\bar{p}_d = \left(\frac{16}{27}\right)\frac{1}{2}\rho\int_0^\infty u^3 f(u)du = \left(\frac{16}{27}\right)\frac{1}{2}\rho\overline{u^3}$$

Equation 2.9

Where 16/27 is the Betz limit and $\overline{u^3}$ is the mean of the cubed wind speeds (as opposed to the cube of the mean wind speed). For a Weibull distribution this can be obtained directly from the third non-central moment as [40]:

$$\overline{u^3} = c^3.\Gamma(1 + {}^3/_k)$$

Equation 2.10

The expected energy yield E from a particular wind turbine can be estimated for a time period t with reference to the manufacturer's power curve $PWT(u)$ using:

$$E = t\int_0^\infty PWT(u)f(u)du$$

Equation 2.11

A similar result can be achieved numerically using the actual histogram of measured wind speeds along with the manufacturer's power curve. This method allows one to simply multiply the normalized frequency of a particular wind speed by the power produced at that wind speed. A sum of the results will yield the average power extracted from the wind.

Figure 2.3 shows an example of this process based on the measured power curve for the Kingspan KW6 wind turbine and an ideal Weibull distribution. The shape of the weighted turbine power curve will depend on the characteristics of the turbine and the wind speed distribution at an individual site. The position of the peak in the curve reveals the wind speed at which the most power is extracted from the wind for a specific wind regime and turbine.

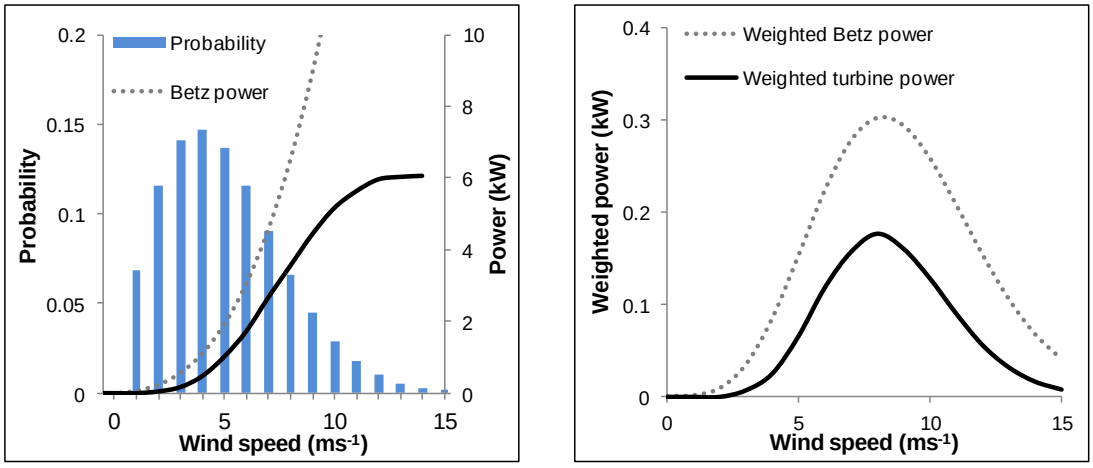

Figure 2.3: Left: Weibull wind speed distribution (k = 1.9, u = 5 ms^{-1}) and a corresponding turbine power curve for the Kingspan KW6 turbine (solid black line). The dotted line shows the theoretical Betz power in the wind. Right: wind power density obtained by multiplying the Weibull distribution by the Betz power or the turbine power curve. The area under the curves represents the total power.

While describing the wind resource statistically in terms of a mathematical distribution inevitably introduces some error, there are several advantages to using this approach compared to a simple histogram of measured wind speeds. Firstly, the use of a mathematical distribution allows greater flexibility in the analysis and manipulation of the data and makes it possible to predict turbine behaviour under a range of wind conditions. Secondly, it allows the wind resource at different sites to be characterized in terms of the mathematical parameters of the distribution making comparisons between sites more meaningful. Thirdly, the application of an appropriate distribution is necessary in order to make power predictions using strategies that predict only a mean wind speed rather than a complete time series of wind speeds.

2.3.1.1 Fitting methods

A range of techniques are available for fitting the Weibull distribution to observed data and a number of studies have been published comparing the relative merits of the different methods. The three most commonly used are the Least Squares

Method (LSM), Method of Moments (MM), and Method of Maximum Likelihood (MML).

The LSM involves writing the Weibull cdf in linear form in order to extract the distribution parameters using linear regression. In practice this requires binning the data according to wind speed resulting in a regression line that gives equal weight to each wind speed bin regardless of the number of data points. Thus, wind speed bins at the extrema of the distribution, which contain few data points, may disproportionately affect the regression fit and the extracted distribution parameters.

The MM takes advantage of the fact that the coefficient of variation for the Weibull distribution can be expressed directly in terms of the distribution parameters. The equation can then be solved iteratively to obtain k and c. However, Carta *et al*. [38] have noted that estimators obtained using MM are not robust (small deviations in the sample can result in large deviations in the estimator) and are not with minimum variance.

MML searches for the distribution parameters that maximize the Likelihood function. Given a measured data set and the constraint of a particular family of distributions (e.g. Weibull) it finds the distribution which has the highest probability given the observed data. The Likelihood function L is the joint density function of n random variables and the unknown distribution parameters. It describes the likelihood of observing the data as a function of the distribution parameters. It is given by [44]:

$$L(u_1, \dots u_n; k, c) = \prod_{i=1}^{n} k \frac{u_i^{k-1}}{c^k} \exp\left[-\left(\frac{u_i}{c}\right)^k\right]$$

Equation 2.12

where the index denotes the i^{th} observation and n is the total number of observations.

Mathematically it is more convenient to work with the Log-Likelihood ($\ln L$) function which transforms the expression into a summation. The maximum likelihood estimates for the shape and scale parameters can be found by maximizing $\ln L$. This is achieved by differentiating $\ln L$ with respect to c and k and setting the partial derivatives to zero.

This process yields the following expression [45] for k that may be solved numerically:

$$k = \left[\frac{\sum_{i=1}^{n} u_i^k \ln(u_i)}{\sum_{i=1}^{n} u_i^k} - \frac{\sum_{i=1}^{n} \ln(u_i)}{n}\right]^{-1}$$

Equation 2.13

Given k it is possible to obtain c from Equation 2.8.

In the context of wind resource assessment, the representativeness of a Weibull distribution, given a series of observed wind speeds, may be assessed directly by considering the wind power. Equation 2.9 and Equation 2.10 can be used to compare the observed and Weibull estimates of $\bar{p}_d$ and hence provide an estimate of the likely error introduced through using a fitted distribution in place of the observed data. Since in the case of wind speed distributions, formal statistical tests are often not appropriate due to violation of the independence assumption, [38, 40], this approach provides a simple quantitative measure of the goodness of fit.

2.3.2 Alternatives to the Weibull distribution

Although the Weibull has become the most widely accepted wind speed distribution in the wind energy industry, it is not always fully justifiable from a theoretical perspective [46]. Hence, a number of studies have been published in recent years comparing the success of different wind speed distributions in representing the wind resource. Unfortunately, many of these studies only consider wind speed data from a single site or a small number of sites and hence the conclusions reached may not be generalisable.

For example, Celik *et al*. [47] compared five probability distributions applied to wind speed data collected at a roof-top site in Edinburgh. They concluded that a bi-modal Weibull distribution was more successful in describing the measured data compared to a standard Weibull. However, as the authors note, general conclusions cannot be drawn from investigating a single site with a specific topographic nature. A more comprehensive study by Chang *et al*. [48] investigated the performance of six probability distributions, including Weibull mixtures, using wind speed data from three wind farm sites in Taiwan. The authors concluded that the standard Weibull distribution performed well except when the observed wind speeds exhibited bimodality.

A particularly useful and comprehensive study was carried out by Carta *et al*. [38] who investigated 12 different probability distributions using data selected from four sites in the Canary Islands. The sites were specifically selected to represent different wind regimes. The study concluded that while the Weibull distribution is not always valid, it has a number of desirable properties in that: (i) it is widely

applicable, (ii) it only requires two distribution parameters, (iii) it can be expressed in a mathematically closed form and (iv) it offers the possibility of estimating confidence intervals (although in practice, autocorrelation between wind observations may violate the assumptions such estimates are based on).

It is of note that bimodality does not necessarily exclude the possibility of fitting Weibull distribution functions to data. For example, in cases where the roughness and/or orography vary with wind direction, the overall wind distribution may contain contributions from several distinct Weibulls that depend on the wind direction. This phenomena is accounted for in commercial site assessment software such as the Wind Atlas Analysis and Application Program (WAsP) [49] where an 'emergent' Weibull distribution is calculated as a weighted sum of the Weibull fits to a number of angular sectors. Mathematically, such distributions can be described by the Weibull mixture distribution [50, 51].

These studies indicate that in the absence of information to the contrary, the Weibull distribution provides a sensible starting point in describing the wind resource. If the data exhibits bimodality, it is likely that a different distribution will be required and a number of alternatives are available, including derivatives such as Weibull mixture distributions.

2.4 Wind Flows in the Boundary Layer

In order to develop physical and statistical approaches to wind resource assessment, an understanding of the processes affecting wind flows close to the Earth's surface is required. These flows are described by boundary layer meteorology and a short overview of the most pertinent boundary layer processes is given in the following section.

2.4.1 Planetary boundary layer

A number of important atmospheric processes occur in the lower level of the troposphere known as the planetary boundary layer (PBL). At the top of this layer (approximately 100 – 2000 m above the Earth's surface, depending on atmospheric conditions), the wind speed is governed by the geostrophic wind which is driven by synoptic scale weather events [52]. Within the PBL, the motion of the air is significantly affected by buoyancy forces instigated by solar heating of the Earth's surface as well as shear forces due to roughness elements. This results in turbulent eddies which are set in motion by free or forced convection. Since wind speeds increase with height from the surface, the effect of turbulent mixing is a net transfer

of momentum from higher to lower levels [53]. The degree of mixing, and hence momentum transfer, depends on the thermal stability conditions within the boundary layer. In the case of unstable conditions instigated by strong solar heating of the surface, turbulent mixing will be enhanced, resulting in greater vertical momentum transfer and a reduced vertical wind speed gradient. The converse will be true in the case of a stably stratified boundary layer where turbulent mixing will be damped.

For wind energy applications, calculations are sometimes simplified by assuming a well-mixed, neutral boundary layer where thermal effects are not significant. This is somewhat justified since neutral conditions are generally associated with higher wind speeds and these are of most importance to wind energy applications. However, the effect of non-neutral stability conditions can introduce significant complexities in certain cases, such as at a land/sea interface for example [54].

As one moves down through the PBL, the effect of surface roughness becomes increasingly important and the wind speed will differ significantly from the geostrophic in both magnitude and direction. Roughness elements such as crops, tress and buildings exert drag forces on the wind flow, resulting in turbulent stresses which spread upwards creating a shear in the vertical wind profile. The magnitude of this shear is dependent on the characteristics (height, density, shape, etc.) of the roughness elements. Due to their typical mounting heights (Table 2.1), small-scale wind turbines will be particularly sensitive to the processes occurring close to the surface in the lowest part of the PBL. In the following sections, these processes are considered in detail.

2.4.2 Inertial sublayer

In a boundary layer model of the troposphere, a number of sublayers are defined within the PBL where specific modelling principles are applied. The most significant of these in the context of small-scale wind energy is the inertial sublayer (ISL) also referred to as the surface layer. The ISL encompasses the lower region of the PBL up to a height much greater than the average height of the roughness elements and much lower than the top of the PBL [15]. Two key characteristics of the ISL are that the shear stress is approximately constant with height and that the vertical wind profile may be described using a logarithmic expression. The justification for using a logarithmic wind profile may be developed in several ways including consideration of the vertical flux of horizontal momentum [53] as well simple dimensional arguments [15]. An intuitive, although somewhat simplified, approach is to consider the transfer of momentum as parcels of air move between different layers [55, 56].

For an air parcel with mean wind speed $\bar{u}$ moving up through the ISL over a vertical distance l, the fluctuation in velocity u' can be written as:

$$u' = \bar{u}(z - l) - \bar{u}(z) = -l\frac{d\bar{u}}{dz}$$

Equation 2.14

where z represents the height above ground level.

In the spirit of Prandtl [57], who related the vertical (w') and streamwise (u') fluctuating velocity components to the vertical flow gradient within the boundary layer, we can write:

$$\overline{u'w'} = -l^2\left[\frac{d\bar{u}}{dz}\right]^2$$

Equation 2.15

where l can be interpreted as the mixing length, loosely defined as the distance traversed by turbulent eddies before they lose their identity [57].

Since the flow is bounded by the surface, l is related to the height above ground level z, with the constant of proportionality known as the von Karman constant κ. Expressing the Reynolds shear stress τ in terms of the average of the fluctuating velocity components and combining with Equation 2.15 gives:

$$\tau = -\rho\overline{u'w'} = \rho\kappa^2 z^2\left[\frac{d\bar{u}}{dz}\right]^2$$

Equation 2.16

Here it is convenient to define a velocity at some reference height where the shear stress is proportional to the square of the velocity [53]. This is known as the friction velocity u_* defined by:

$$u_* = \sqrt{(\tau/\rho)}$$

Equation 2.17

Combining Equation 2.16 and Equation 2.17 and integrating yields a logarithmic vertical wind profile u_z of the form:

$$u_z = \frac{u_*}{\kappa}[\ln(z) + A]$$

Equation 2.18

The integration constant A can be defined such that the wind speed is nominally zero at some height, where this height is equal to the roughness length z_0, giving the expression:

$$u_z = \frac{u_*}{\kappa}\ln\left(\frac{z}{z_0}\right)$$

Equation 2.19

For flows over rough surfaces, a correction to the height above ground level must be used to account for the blocking effect of the obstacles. This is achieved through a displacement height d which acts to shift the profile to a height $z - d$. The resulting log law is:

$$u_z = \frac{u_*}{\kappa}\ln\left(\frac{z-d}{z_0}\right)$$

Equation 2.20

The logarithmic expressions given by Equation 2.19 and Equation 2.20 are only strictly valid in neutral stability conditions and in instances where this assumption does not hold, these expressions can be modified by a stability parameter.

2.4.3 Internal boundary layer at a roughness change

The form of Equation 2.19 implies that the vertical wind speed profile in the ISL is affected by the roughness of the surface. Different land cover may have significantly different roughness characteristics and these cause important modifications to the wind flow close to the surface. For example, the high roughness associated with urban areas can result in a significant reduction in wind speed as momentum is lost to the surface through the drag force exerted by buildings. Conversely, flows over open rural areas may proceed relatively uninterrupted for large distances. These differences result in large variations in wind speeds depending on the surface characteristics.

At the interface between regions of differing roughness, an internal boundary layer (IBL) develops as the downstream wind gradually adjusts to the new surface [58]. Various definitions of the IBL may be used but in simple terms this can be considered as the layer within which the flow is significantly affected by the

presence of the new surface. The development of the IBL downstream from an abrupt roughness change is shown schematically in Figure 2.4

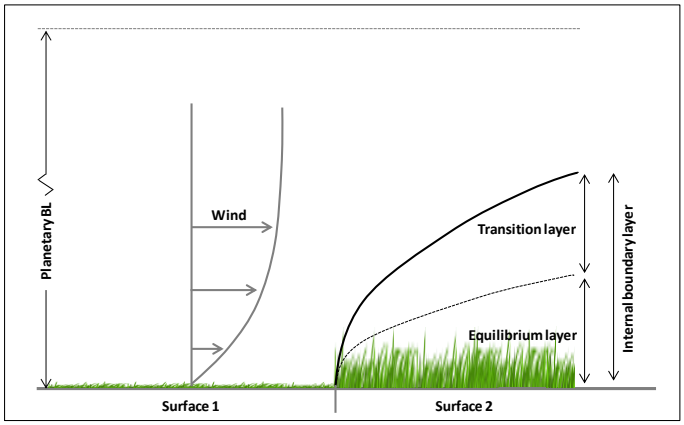

Figure 2.4: Growth of an internal boundary layer at a smooth to rough transition. Adapted from [59].

Above the IBL, the vertical wind speed profile is affected only by the upstream surface, while within the IBL, the profile gradually adjusts to the properties of the local surface. Downstream from the roughness change, the developing IBL can be further divided into an equilibrium layer, where the flow is fully adjusted to the local surface, as well as a transition region where the flow is affected by both the local and upstream surfaces. Peterson [60] showed that the transition region may occupy a significant part of the developing IBL, with only the lower region of this layer fully in equilibrium with the local surface. Clearly, the wind speed at some distance downstream from an abrupt roughness change will depend on the vertical position with respect to these developing layers.

A number of semi-empirical expressions have been developed to describe IBL growth as a function of the distance from the roughness change [58, 61]. These formulations vary in complexity from simple power laws based on the downstream roughness to more complex expressions which account for both upstream and downstream roughness as well as stability effects. Not surprisingly given their different input parameters, these expression lead to a relatively large range of predicted IBL depths [15]. An approximate rule-of-thumb for smooth to rough surfaces is an IBL height to fetch ratio of 1/10, and since only the lowest 10% of this layer will be fully in equilibrium with the downstream surface, this implies that the equilibrium layer shown in Figure 2.4 has a height to fetch ratio of around 1/100 [59, 60]. Rough to smooth transitions are likely to be associated with smaller height to fetch ratios as the flow adjusts more slowly to the smooth surface [62].

In theory, if the heights of the IBL and equilibrium layer are known, a logarithmic vertical wind speed profile (Equation 2.19) may be applied above the IBL and within the equilibrium layer using appropriate roughness parameters for the upstream and downstream surfaces respectively. Within the transition region, however, the wind flow is not fully in equilibrium with either surface and prediction of the wind speed becomes more complex.

2.4.4 Multiple internal boundary layers and blending methods

Frequently, sites where the wind flow is to be estimated are located in regions where the topography is more complex than the single roughness transition described in Section 2.4.3. Even in regions of simple orography, variations in land cover may result in an upstream fetch that consists of multiple surfaces of differing roughness. If the sizes of these patches are such that the IBL is not fully developed before encountering the next roughness change, the developing layers will interact thus increasing the complexity [15]. Figure 2.5 shows a simple representation of the development of regional IBLs over a patchy surface in heterogeneous terrain.

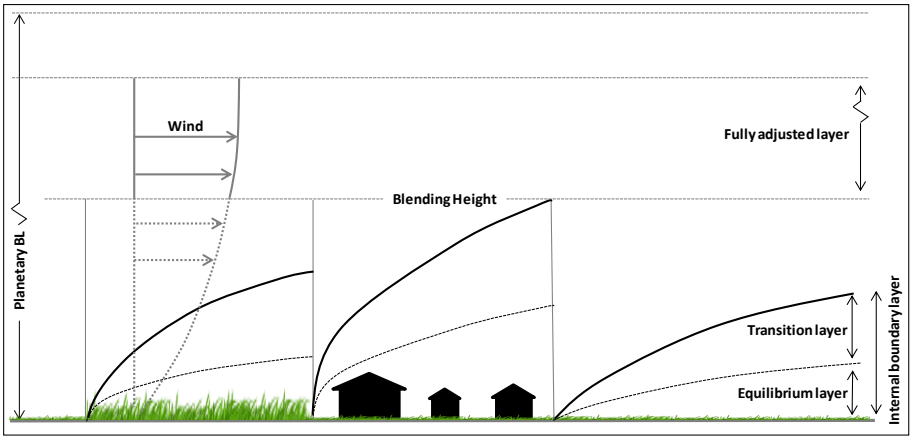

Figure 2.5 Development of internal boundary layers over a region of patchy terrain, adapted from Goode and Belcher [63]. The wind profile is shown on the left for the fully adjusted layer; dotted lines represent the extrapolated profile below the blending height.

Based on the work of Wieringa [64] as well as Mason [65], a blending height can be defined, above which, the individual contributions from multiple roughness patches may be combined into a single average. Wind flows above this height are considered to be fully adjusted to some effective roughness that is representative of the overall surface. The formal definition of the blending height is debated [63], although an intuitive definition is the height at which the flow is horizontally homogeneous and in equilibrium with the combined surface patches, [65]. Note that the characteristics of the ISL given in Section 2.4.2 may only be fully applicable

above the blending height, since below this, and depending on the size of the roughness patches, the ISL related to each roughness patch may only be partially developed or may not exist at all [66].

Estimating a suitable value for the blending height can be problematic. Applying a strict criterion that the flow below the blending height should be in equilibrium with the local surface implies that the blending height is close to the top of the equilibrium layer. Based on the boundary layer growth theory discussed in Section 2.4.3, this implies a blending height of the order of *length*/100 where '*length'* is the horizontal scale of the roughness patches. Mason suggests a value of *length*/200 allowing for cases of slower boundary layer growth but emphasizes that this is an approximate scale rather than a precise value. For real surfaces, a characteristic length scale on which to base such estimates may not exist. In addition, the above discussion assumes that the blending height marks an abrupt transition between the influence of the local and regional surfaces. This implies a discontinuity in the Reynolds stress and ignores the transition layer shown in Figure 2.5. In reality, the stress adjusts from that induced by the local surface to that induced by the regional surface over some depth termed the blending layer [63]. Despite these complexities, the concept of a single blending height is a useful approximation and in many cases, wind speed estimates may not be significantly affected by its exact definition [15].

From the above discussion, it is clear that when applying the logarithmic wind profile of Equation 2.19, account must be taken of which layer is being modelled and suitable parameterisations of the surface roughness must be applied. In addition, below the blending height the developing layers may be somewhat complex due to the influence of multiple surfaces, thus complicating a simple parameterisation of the surface.

2.4.4.1 Blending methods

Since in heterogeneous terrain the wind flow above the blending height is affected by multiple surface patches, a method is required to obtain a parameterisation of the spatially averaged roughness of these patches. This is generally achieved through the calculation of an effective roughness length z_{0eff} applicable above the blending height. Formally, z_{0eff} is defined as the roughness of an equivalent homogenous surface that would give rise to the same average stress as the heterogeneous surface [65]. An expression for z_{0eff} can be derived using the concept of source areas [67] or a blending method [65]. The blending method

provides an intuitive approach that can be applied directly to gridded land cover data as described below.

Considering alternating patches of equal size and differing surface roughness, and with reference to Equation 2.17, the average surface stress for two such patches can be written in terms of the friction velocity [15]:

$$u_*^2 = 0.5(u_{*1}^2 + u_{*2}^2)$$

Equation 2.21

This concept can be further extended by recognising that at the blending height (Figure 2.5), the wind profile is assumed to be in equilibrium with the overall surface and the individual roughness patches. Hence, combining Equation 2.19 and Equation 2.21, z_{0eff} can be expressed in terms of the roughness lengths of the two surfaces and the blending height z_{bh} [15]:

$$\left[\ln\left(\frac{z_{bh}}{z_{0eff}}\right)\right]^{-2} = 0.5\left\{\left[\ln\left(\frac{z_{bh}}{z_{0,1}}\right)\right]^{-2} + \left[\ln\left(\frac{z_{bh}}{z_{0,2}}\right)\right]^{-2}\right\}$$

Equation 2.22

where $z_{0,1}$ and $z_{0,2}$ are the roughness lengths of the two surfaces.

The heterogeneity of real surfaces will often be more complex than the simple alternating patches postulated above [68]. However, as a first approximation, the above arguments can be generalised for surfaces with multiple patches of varying sizes using [15]:

$$\left[\ln\left(\frac{z_{bh}}{z_{0eff}}\right)\right]^{-2} = \sum_i F_i\left\{\left[\ln\left(\frac{z_{bh}}{z_{0,i}}\right)\right]^{-2}\right\}$$

Equation 2.23

where F_i represents the fraction of surface i with roughness length $z_{0,i}$.

Providing an estimate of z_{bh} is available, Equation 2.23 is a convenient expression for calculating effective roughness lengths using automated approaches based on digitised land cover data. In many cases, the height scale in Equation 2.23 must be modified to include an effective displacement height using $z_{bh} - d_{eff}$. The value of d_{eff} is subject to some uncertainty since there is no rigorous method for estimating an effective displacement height for a heterogeneous surface [15].

In developing a methodology for downscaling the spatially averaged mean wind speed, the concept of a blending height and effective roughness parameters leads

naturally to the idea of a regional and local downscaling. The vertical wind profile above the blending height is influenced by multiple surface patches, which, at a sufficient height, are indistinguishable from a single homogenous surface characterised by an effective roughness. This surface can be considered as representative of a regional area, the dimensions of which will depend on the vertical and horizontal scales under consideration. Below the blending height, at least within the equilibrium layer, the vertical wind profile is determined by the local roughness of the specific patch. The blending height in this scheme marks the approximate transition between the local and regional scaling and at this height both wind profiles are assumed to be in equilibrium.

2.4.4.2 Regional and local downscaling

Given a reference wind speed at sufficient height to be independent of the surface roughness, the arguments above provide a framework for downscaling using a logarithmic wind profile and parameterisations of the regional and local roughness. Firstly, the wind speed may be downscaled to the blending height using a regional parameterisation of the surface. Secondly, a local downscaling can be implemented using the local surface roughness with the requirement that the profiles should match at the blending height. Note that such an approach is a simplification since below the blending height, depending on the distance from the roughness change, the equilibrium layer may not extend throughout the entire depth of the developing IBL (Figure 2.5).

In the simplest case of a regionally homogenous, rural area, this procedure simply reduces to the application of a single logarithmic wind profile using an appropriate rural roughness length. If the regional area includes multiple patches of different roughness, the regional downscaling must also account for the effective roughness above the blending height. More challenging cases arise when the local area is at a coastal location or within the built environment. These cases can introduce significant complexities related to parameterisation of the surface roughness and modelling of the IBL. The following sections address these issues in more detail.

2.4.5 Coastal boundary layers

The formation of an IBL at the land/sea interface presents several additional complexities compared to a simple roughness change on land and these may present significant challenges when predicting the wind resource at coastal sites. The land/sea interface is generally associated with an abrupt change in surface roughness leading to the development of an IBL as described in Section 2.4.3. In addition, the interface features changes in temperature and humidity which can lead

to differing stability conditions onshore and offshore. In the case of onshore sites, of primary interest is the development of the onshore IBL as winds from across the sea reach land and encounter a step change in surface roughness and stability. This situation has been studied in some detail [54, 69-71] and some general principles of interest to coastal wind speed predictions can be noted.

If the stability conditions onshore and offshore are the same, the change in wind speed can be modelled based on only the change in roughness. Since the onshore roughness length is generally at least an order of magnitude greater than offshore, this leads to a reduction in wind speed as the wind passes over land [71]. However, even in this simple case, wind speeds will be affected at certain times of day by thermally driven sea breezes due to differential heating and cooling of the land and sea [54]. These effects will have strong diurnal and seasonal components and the direction of such breezes will depend on the orientation of the coastline. In addition, local orography, such as the presence of cliffs, may cause further modifications to the wind flow. These factors illustrate that, even in the simplest case, the coastal climate may be highly site-specific and hence challenging to predict using generalised modelling approaches.

As a first step to developing a model of wind flows in coastal regions, an expression for the IBL growth with fetch length is required. Barthelmie [69] reviewed a number of semi-empirical expressions for coastal IBL growth and demonstrated that even in neutral conditions these lead to widely varying estimates of the IBL height. For illustration, estimated heights for a sea to land transition ($z_{0,land}$ = 0.1 m and $z_{0,sea}$ = 0.0002 m) in neutral stability are sketched in Figure 2.6 using six different IBL growth expressions as reviewed by Barthelmie [69]. At a fetch of 1 km, the IBL height is estimated to be approximately 30 – 100 m while at a 5 km fetch this increases to 120 – 350 m.

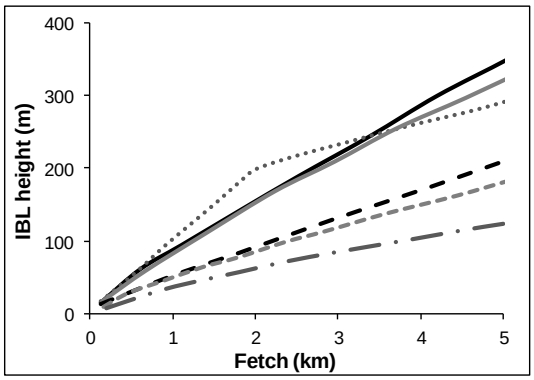

Figure 2.6: Estimated IBL heights for a sea to land transition ($z_{0,land}$ = 0.1 m and $z_{0,sea}$ = 0.0002 m) in neutral stability using six IBL growth expressions reviewed by Barthelmie. Figure adapted from reference [69].

The relatively large uncertainty in the IBL height is further compounded by the influence of non-neutral stability. In general, unstable conditions will result in more efficient vertical mixing and a deeper IBL while the converse will be true for stable conditions. Bergstrom *et al*. [54] investigated coastal stability effects in some detail using coastal observations and simple models and concluded that stability effects had a significant impact on surface wind speeds and IBL height. Their results also indicated that the downstream stability was of more importance than the upstream in determining the wind profile. An interesting observation made by Bergstrom *et al*. was that wind speeds can in unstable conditions increase as the wind moves from sea to land despite the increase in surface roughness. This effect is thought to be due to efficient transfer of high momentum air from higher to lower levels as the wind encounters an unstable atmosphere over land [54]. This effect demonstrates that thermal stability can be as important as the roughness change when considering coastal climates.

In theory, both the logarithmic wind profile (Equation 2.19) and the expressions for boundary layer growth can be modified to account for different stability conditions. However, any such modification is complicated by the fact that stability is likely to depend on fetch, season and time of day. For example, Barthelmie [70] reported that the presence of land in the fetch was linked to a higher frequency of stable conditions while a sea fetch was linked to more frequent near-neutral or unstable conditions at an offshore site in Denmark. Since the fetch in a particular angular sector will be linked to the orientation of the coastline, stability conditions may be highly site-specific.

Seasonal weather patterns are a further source of variable stability conditions between the land and sea. Due to the sea's greater thermal capacity and more efficient heat transfer mechanisms compared to land, near surface air temperatures

offshore will lag the more abrupt changes onshore [53]. In late autumn for example, the air above the sea surface will be relatively warm compared to the air above the land. Thus, the offshore region may be associated with a higher frequency of neutral or unstable conditions compared to onshore. The opposite will be true in spring, particularly during the daytime, where the land surface may experience rapid heating while the sea remains relatively cool [70, 72]. A further complication is the presence of diurnal temperature changes. Since the diurnal temperature cycle is much more pronounced over land compared to the sea, this can result in the development of diurnal stability differences over land and sea [73]. Overnight a stable boundary layer may develop over land while the warm air over the sea may reduce stability offshore. The converse will be true during the day as solar heating instigates instability over land. These effects can lead to significant complexities in the diurnal wind patterns close to the shore and are likely to be greatest in spring and summer when solar heating of the surface is most pronounced [74]. Figure 2.7 shows a schematic representation of some of the major processes discussed above that complicate the modelling of coastal wind speeds.

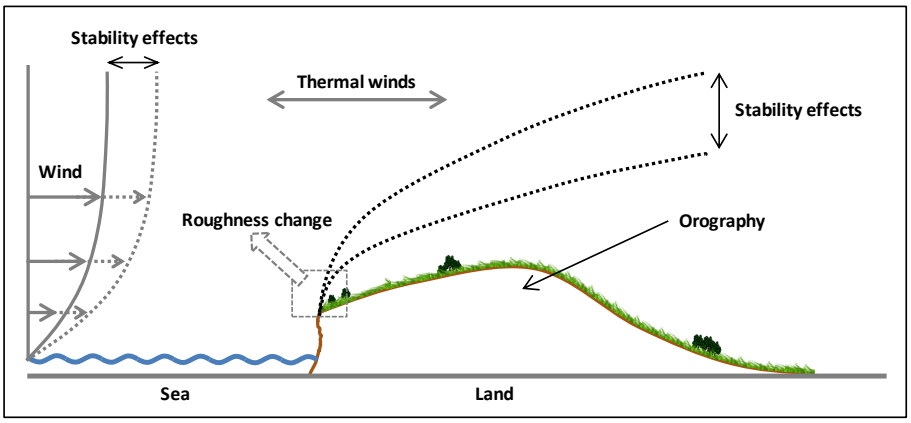

Figure 2.7: Schematic representation of the processes occurring at the land sea interface that complicate wind speed predictions.

In summary, several key observations can be made regarding wind flows at coastal sites. Firstly, deviations from neutral stability will affect both the vertical wind profile and the growth of the onshore IBL. Hence, the situation is somewhat more complex than a simple roughness change in conditions of neutral stability. Stability effects are likely to vary with season and time of day and modelling such effects is challenging. These processes are complicated by thermally driven winds which follow diurnal cycles and depend on the orientation of the coast. Finally, the presence of sea cliffs and other complex orography may further impact on the wind flows close to the land/sea boundary. The significance of many of these processes will depend on the presence of sea or land in the fetch. Hence, even for two sites

located in relatively close proximity, the wind climate may be somewhat different depending on the exact distance from the coast as well as the coastal orientation. It is also very likely that any such differences will exhibit diurnal and seasonal variability.

These factors present a number of challenges for the implementation of simple boundary layer scaling approaches. Specifically, assumptions of neutral stability may not hold, the roughness of the fetch is likely to be strongly dependent on the wind direction, IBL heights may be difficult to estimate and complex orography may not be accounted for. This is likely to lead to large uncertainties in wind speed predictions at coastal sites, as discussed in Chapter 4. The implementation of wind resource prediction approaches based on statistical correlations to nearby long-term reference sites may be similarly challenging. Due to the possibility of highly localised wind climates, the wind speeds at sites may become decoupled across relatively short distances. As described in Chapter 5, obtaining suitable reference wind observations can be particularly problematic for coastal sites and alternative data sources may be required. One such alternative is discussed in detail in Chapter 8.

2.4.6 Urban boundary layers

While this thesis is not specifically concerned with urban installations, any resource assessment methodology should be able to make at least approximate predictions within the built environment. As with coastal regions, the application of boundary layer modelling principles to urban areas presents several challenges. Firstly, simple parameterisations of the surface roughness become complex due to the variability in the size and spacing of the buildings [75]. Secondly, surface heterogeneity leads to variability in the local wind flows, and hence, the boundary layer structure below the blending height must be considered in more detail. Thirdly, the presence of individual building wakes can cause local perturbations to the flow on the scale of individual buildings.

A description of the vertical wind speed profile within the built environment can be developed using the general principles of boundary layer flows described previously. However, in urban and sub-urban areas, the surface shear associated with buildings causes significant modifications to the wind speed. These effects can be most easily understood by considering the urban boundary layer (UBL), along with its associated sublayers, as a special case of an IBL.

In the transition from areas of low to high roughness (e.g. rural to urban), the UBL starts to develop as the wind flow adjusts to the new rough surface. The height of

this layer will increase with downwind distance from the edge of the roughness change, eventually reaching a height of several hundred metres [52]. Due to the heterogeneity of the surface, IBLs will develop within the UBL as described in Section 2.4.4. Following the description of Grimmond and Oke [76] and other authors [15, 52], it is convenient to divide the UBL into sublayers as depicted in Figure 2.8. Each of these layers is described in more detail below.

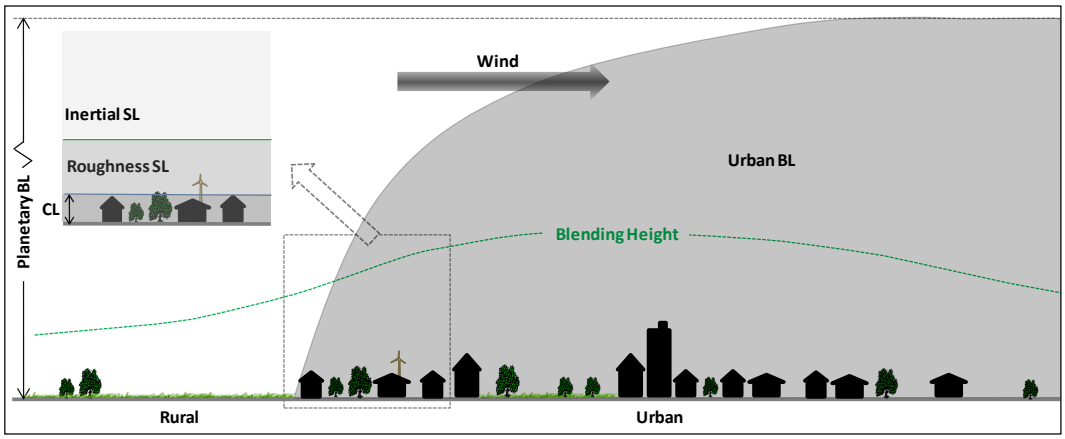

Figure 2.8: Formation of the urban boundary layer (UBL) at a roughness change. Inset: The internal sublayers (SL) within the UBL. Adapted from reference [15].

The blending height

For flows above rough, heterogeneous surfaces, multiple interacting wakes develop above individual roughness elements and these cause horizontal inhomogeneity in the flow. This situation has some parallels with the interacting multiple IBLs over a heterogeneous surface presented in Figure 2.5 (Section 2.4.4). At some height above the surface, the individual wakes are no longer distinguishable and the wind flow can be considered to be horizontally homogenous and influenced by the average roughness of the overall surface. For such surfaces, it is natural to define a height scale at which horizontal homogeneity occurs. Although this concept arises from different theoretical considerations to those discussed in Section 2.4.4 for blending heights above multiple patches of differing surface roughness, the two situations are somewhat analogous. Since the blending height discussed in Section 2.4.4 is a height scale rather than a fixed value, the term ‘blending height’ is also widely used to define the vertical transition to horizontally homogenous wind flows over urban surfaces [52, 75].

The Inertial Sublayer

Above the blending height, the standard logarithmic scaling, which is characteristic of the ISL, can be applied using effective roughness parameters. The flow in this

layer is subject to the characteristics of the regional surface which may include contributions from both urban and non-urban land cover. The ability to use Equation 2.19 to predict the wind speed in the ISL, and by extension in the roughness sublayer (RSL) [17, 66], relies on an estimate of the effective aerodynamic parameters z_{0eff} and d_{eff} based on the underlying surface.

The Roughness Sublayer

The wind profile within the RSL is of interest to small-scale wind energy since the majority of turbines installed above rough surfaces will be located within this layer. The top of the RSL (marked by the blending height) is generally assumed to extend to approximately 2-5 times the average height of the roughness elements, although there is a degree of uncertainty in this value due to the complexity of the built environment [77]. Several experimental methods exist for estimating this in practice, including the height of maximum shear stress [78] or, perhaps more intuitively, the height at which the measured flow becomes horizontally uniform [66].

Below the blending height, the wind flow will be affected by individual roughness elements, and may exhibit a high degree of spatial variability, making wind speed predictions particularly challenging. Without recourse to detailed fluid flow modelling or onsite measurements, it is only possible to define a spatial average for the wind speed in this layer. Based on wind tunnel studies, Cheng and Castro [66] have shown that a single logarithmic profile can be extended from the ISL down through the RSL in order to describe the spatially averaged wind speed, despite the fact that the assumption of constant shear stress does not strictly hold in this layer. However, of particular importance here is the choice of the aerodynamic parameters z_0 and d used to describe the surface.

A review by Garratt [58] found that while the frequently used values of 0.1 and 0.7 times the mean obstacle height, for z_0 and d respectively, are applicable for a number of natural surfaces, in complex environments, these parameters will be a function of the size, shape and coverage of the roughness elements and are non-trivial to determine. It has been noticed by several authors [17, 66, 79] that if these parameters are extracted from fitting a logarithmic profile (Equation 2.20) to the ISL alone, this can result in a significant error in the predicted wind speed in the RSL.

The Canopy Layer

A final region, which extends from the surface to the top of the roughness elements, may be defined as the canopy layer (CL). Within this layer the wind flows are strongly affected by the local building and street geometry resulting in complex

effects such as channelling and recirculation. Due to these effects, defining an average wind speed profile within the urban canopy can be even more problematic than within the RSL. While the wind speeds in the CL are likely to be too low to make turbine installations viable, roof mounted turbines may well be located close to the boundary between the CL and RSL and hence an understanding of the vertical wind profile in the CL is relevant to wind resource estimation. Theoretical investigations of plant canopies as well as wind tunnel studies over arrays of cubes suggest an exponential wind speed profile can be used of the form [80]:

$$u(z) = u(h)exp\left[-a\frac{h-z}{h}\right]$$

Equation 2.24

where h is the canopy height and a is an attenuation parameter obtained from length scales associated with the canopy. Due to the exponential relationship, the predicted wind speed near the top of the canopy changes rapidly with height. Hence, for heights close to the top of the canopy, the predicted wind speed will be particularly sensitive to the choice of canopy height. This raises the question of how the canopy height should be defined for the built environment, which almost always features heterogeneous building heights resulting in an ambiguous definition of the canopy top.

2.4.6.1 Roughness parameters in the built environment

Within the built environment, the aerodynamic parameters of z_0 and d, required to formulate the vertical wind speed profile, can be obtained using two approaches: (i) a morphometric method using the three-dimensional structure of the underlying surface [75] or (ii) a micrometeorological method using observed wind profiles to extract the parameters of the logarithmic wind speed profile [81]. Both approaches present considerable challenges in real-world, complex, environments.

In a morphometric approach, the starting point is to define a number of geometric parameters that describe the size and arrangement of the roughness elements. Typically, the parameters used are the mean height of the elements (h_m) and the frontal ($\lambda_f = A_f/A_t$) and plan ($\lambda_p = A_p/A_t$) area densities as illustrated in Figure 2.9.

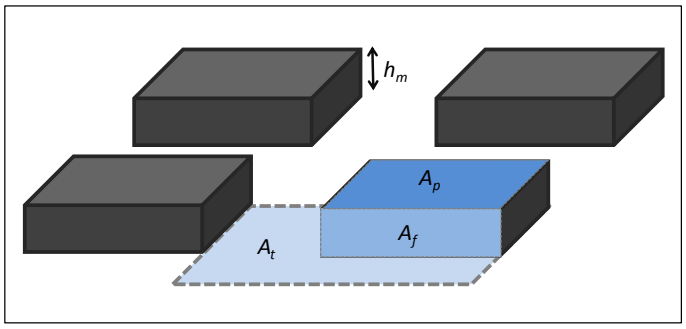

Figure 2.9: Illustration of surface parameters which are useful in determining roughness length and displacement height in a morphometric approach. Adapted from reference [75].

Two of the more successful morphometric methodologies are those outlined by MacDonald *et al*. [82] and Raupach [83]. MacDonald *et al*. [82] developed analytical expressions for z_0 and d based on the drag of the individual obstacles and assuming a logarithmic wind profile above the canopy. It is assumed that only the frontal area of the obstacles above the displacement height influences the drag balance and the derived expression includes a parameter which alters the drag coefficient depending on the geometry of the obstacles. The expressions developed by Raupach [83] are derived by partitioning the total drag between that imposed by the obstacles and that imposed by the underlying surface. This treatment produces similar trends to MacDonald, although according to a review by Grimmond and Oke [75], the method of Raupach [83] provides slightly more accurate results. A feature of both approaches is that while d increases monotonically with increasing plan area density, z_0 peaks before falling at high densities. This behaviour can be understood in terms of different flow regimes that dominate at different area densities [84].

Although there has been some success in applying morphometric models to real surfaces there are some clear issues when using this approach. Firstly, the models are generally based on uniform arrays of roughness elements and idealised wind flows that are simply not representative of real conditions in the built environment. Secondly, the surface parameters shown in Figure 2.9 may not be trivial to define for real built environments with variable area densities and building heights. Thirdly, the majority of models do not take into account building height heterogeneity which can lead to markedly different flow characteristics. Recently, Millward-Hopkins *et al.* [85, 86] have built on the work by MacDonald [82] and Raupach [83] in developing a methodology to predict the surface parameters above surfaces with heterogeneous building heights. The methodology has shown promising results when compared to numerical and wind tunnel studies of urban-like surfaces as well

as wind speed observations in UK cities [87]. However, this methodology requires detailed information regarding the heights and footprints of all buildings within a city and hence, as yet, cannot be implemented using simple parameterisations of the surface.

The alternative to a morphometric approach is to experimentally determine aerodynamic parameters from wind speed measurements above urban surfaces, either by means of field studies or using scaled down models in wind tunnel studies. The general approach [81] is to obtain wind speed measurements with high temporal resolution at one or more heights above a rough surface in order to directly estimate the parameters of z_0 and d. Grimmond and Oke [75] conducted a comprehensive review of such micrometeorological studies, including field trials and scale-model wind tunnel investigations. Disappointingly, it was not possible to recognize clear trends from these studies relating the aerodynamic parameters to surface descriptors or allowing the authors to recommend one approach over another. Given the large uncertainties inherent in currently available morphometric and micrometeorological approaches, Grimmond and Oke [75] recommend the use of tables with typical values of z_0 and d for commonly encountered urban surfaces. These can be used in conjunction with aerial photographs to estimate values for a range of built environments.

From the above discussion, it is clear that despite a large number of studies and differing approaches, it is extremely challenging to accurately estimate values of z_0 and d for the built environment. However, accurate values of these parameters are of great importance in making meaningful estimates of the mean wind speed in the RSL using analytical methods.

2.4.6.2 Local perturbations to the flow

In addition to knowledge of the mean flow above the surface, the local effects due to individual obstacles are also of importance to small-scale wind turbines. Detailed treatment of these small-scale flow perturbations go beyond the scope of the current work which is concerned with predicting the spatially averaged mean wind speed.

However, it is worth noting that the wind profile above an individual building will depend on the exact geometry of the building, the measurement point with respect to the roof and whether or not it is embedded in an array [14, 77]. The array effects are related to the behaviour of the aerodynamic parameters for different flow regimes. Interestingly, recent work by Millward-Hopkins *et al.* [17], investigating several regular arrays of urban-like surfaces, suggested that despite the significant

attention given to investigating flows around individual buildings, the errors associated with estimating the values of z_0 and d may be more significant to wind resource assessment than the errors associated with the spatial variation around individual roughness elements.

2.4.7 Complex orography

The presence of complex orography such as hills, valleys and steps causes further modification to the wind flow that may not be accounted for in a simple boundary layer model. Orographic effects can be divided into those due to (i) differential heating of sloping land and associated localised, diurnal wind flows and, (ii) direct modification of synoptic flows through speed-up or sheltering effects [53].

Differential heating is an important source of localised wind flows in valleys, particularly during the summer months [53]. During the day, the slopes and valley floor heat rapidly causing an increase in the temperature of the air close to the surface. Since the air high above the centre of the valley remains relative cool, this sinks to the valley floor and sets in motion circulating thermal cells with weak up-slope (anabatic) winds. Eventually, a temperature gradient develops between the warming and well mixed air within the valley and the cooler air over nearby planer regions. As the cooler air flows into the valley it creates a localised wind through the valley. At night, due to radiative cooling, the process reverses leading to down-slope (katabatic) winds and a reversal of the valley wind.

Direct modification of wind flows due to aerodynamic effects gives rise to the well-known phenomena of speed-up and flow retardation at the crest and leeward side of a hill respectively, as well as a range of other more complex flow distortions. These processes will tend to dominate thermal effects at moderate to high wind speeds [88] and are not simple to generalise due to the wide range of surface structures that occur in nature [53]. The flow modification in the simplest case of a uniform, isolated hill is shown schematically in Figure 2.10.

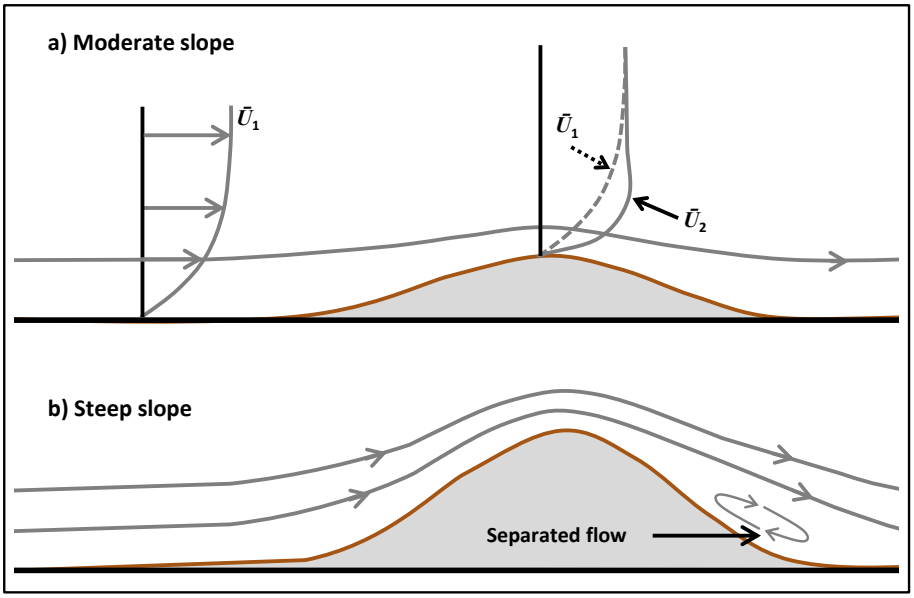

Figure 2.10: Flow modification over a hill of (a) moderate and (b) steep slope. $\bar{U}_1$ and $\bar{U}_2$ represent the mean wind speeds upstream and at the crest of the hill respectively. Adapted from reference [88].

For cases of moderate slope, (< 17º or a height to distance ratio of 0.3), the wind flow over an isolated hill generally remains attached, [53, 88]. For a fixed height above ground level the wind speed is greater at the crest of the hill than at the same height some distance upstream. This speed-up results from constriction of the stream lines in the vertical direction [53] and a subsequent increase in the vertical wind gradient. At the windward and leeward bases of the hill, a reduction in wind speeds is typically observed. Such effects can be described using linear flow models where the velocity is divided into an undisturbed upstream part and a perturbation resulting from the hill [89]. This approach allows simple rules-of-thumb to be developed regarding speed-up ratios (maximum wind speed at crest/upstream wind speed) based on the height and width of the hill [88]. The maximum speed-up ratios are of the order of 1.6-1.8 although these also depend on stability conditions [53]. In the case of steep slopes (> 17º) the flow may become separated at the windward and leeward base of the hill due to the sudden discontinuity in the surface. In these regions, which require description by non-linear models, turbulent eddies develop that reduce the wind speed and even reverse the flow direction. Real-world wind flows over rugged terrain are often significantly more complex than the simple situations shown in Figure 2.10. Hills are rarely uniform or isolated and are frequently in areas of rugged terrain featuring multiple slopes and valleys, escarpments, ridges and varied land cover. Such features increase the flow complexity as well as the likelihood of local perturbations to the flow that are challenging to model using simple methods.

A number of approaches are available for accounting for the effect of orography on wind flows. The linear flow models mentioned above form the basis of the widely used WAsP software developed by the Technical University of Denmark [49]. WAsP applies a linear flow model to local wind observations using terrain and roughness information to produce a description of the wind climate across an extended region. To overcome the limitations of such models in steep terrain, the latest release of WAsP includes a computational fluid dynamics (CFD) module that can resolve the flow in more complex orography. CFD approaches use numerical methods to solve the Navier-Stokes equations that govern the motion of a fluid, subject to specified boundary conditions. These approaches can be applied to detailed terrain models using high resolution meshes and are capable of resolving more complex flows compared to linear models, albeit at additional computational cost [15]. A further class of approaches are the so-called 'mass-consistent' models which use wind observations at multiple sites and the principle of conservation of mass [90]. The applicability of such models is somewhat dependent on the availability of sufficient wind observations to allow the effects of the orography to be properly accounted for [15]. Numerical weather prediction models (NWP) are also capable of representing detailed orography but due to their computational requirements, these are generally run at relatively large sizes. Corrections to NWP for local orography can be achieved using downscaling techniques such as those based on linear flow models [91] or statistical approaches [92].

The presence of complex orography can complicate wind resource predictions using scaling approaches based on boundary layer meteorology, and correlation approaches based on wind observations at correlated site pairs. In the case of the former, the resolution of the input climatology will limit the size of the orographic features that can be resolved without the application of detailed flow modelling. In the case of the later, complex orography may result in highly localised flows that are poorly correlated to nearby reference sites.

2.4.8 Summary of boundary layer processes

Wind flows close to the earth's surface have been the subject of much study, and in many simple situations, they are relatively well understood. Subject to certain restrictions, the vertical wind profile can be described by a logarithmic expression that takes account of the characteristics of the surface. At a roughness change, an internal boundary layer develops as the wind flows gradually adjust to the new surface, and hence, the logarithmic profile will be affected by both the upstream and downstream surfaces. Frequently, multiple surface patches contribute to the overall

surface roughness and this necessitates the use of blending methods to parameterise the effect of the overall surface above the blending height. Coastal and urban sites present significant challenges to simple boundary layer modelling approaches due to the thermal and mechanical properties of these surfaces. Complex orography introduces further challenges as thermal and aerodynamic phenomena may significantly influence the wind flow in rugged terrain.

As will be discussed in the following chapter, the principles of boundary layer meteorology may be used to develop wind resource assessment approaches based on the regional and local scaling of a reference climatology. However, these applications require the use of a number of simplifications and assumptions. The principles outlined above can also be useful in illuminating the physical basis of results obtained from statistically based wind resource assessment approaches using correlated site pairs.

3 Analytical and Data-Driven Approaches to Wind Resource Assessment

In this chapter, attention is turned to specific techniques for assessing the wind resource for small-scale turbine installations. As outlined in previous chapters, wind flows are subject to both spatial and temporal variations that occur on a range of timescales. While such variability can in theory be represented using long-term onsite measurements or detailed computational models, these approaches are often not practical for widespread deployment at potential small-scale wind installations. Hence, in the current chapter, two approaches to wind resource assessment capable of predicting the long-term wind resource are discussed in detail. The first is a direct application of the principles of boundary layer meteorology in a simple scaling approach, while the second is a data-driven approach based on short-term onsite measurements.

This chapter is organised as follows: Section 3.1 describes an analytical methodology, based on the principles of boundary layer meteorology, for predicting the spatially averaged mean wind speed and wind power density. The practical implementation of the approach is described including the assumptions and limitations of the model. Section 3.2 reviews a number of data-driven measure-correlate-predict approaches for assessing the long-term wind resource using short-term measurements. Existing approaches are reviewed with particular attention to their applicability to onsite measurement periods of much less than 12 months and potential new approaches are also proposed.

3.1 A Boundary Layer Scaling Methodology: The Met Office Approach

Given information regarding the regional climatology, as well as details of the local and regional surface characteristics, it is possible to develop an analytical methodology to predict the spatially averaged mean wind speed at a specific point near the Earth's surface. Although this does not provide a complete description of the wind flow, it can then be used as a starting point for predicting the wind resource at a particular location or an aggregated estimate over an extended region. The methodology is based on the principles of boundary layer meteorology

outlined in Section 2.4 along with certain simplifications that allow the approach to be deployed rapidly, without detailed, site-specific information.

A related approach was used by Heath *et al.* [14] to investigate the potential energy output from a hypothetical micro-wind turbine installation in urban London. In addition to predicting the spatially averaged mean wind speed, the study also considered the detailed effects of building scale wind flows using CFD and made recommendations regarding turbine siting. In 2008, a similar approach was generalised under a joint initiative by the Met Office and the Carbon Trust in their assessment of the UK potential for small-scale wind energy [15], albeit without considering detailed microscale flow effects. The study was used to produce gridded estimates (1 km^2) of the generating potential for small-scale wind energy, based on mean wind speeds, and this was aggregated into a UK-wide estimate under various scenarios. The same principles were subsequently applied in the development of the Carbon Trust's online Wind Yield Estimation Tool (WYET) [93]. The tool was intended as an improvement to the simple NOABL-MCS method for wind resource prediction but unfortunately, after an initial period of operation, the tool was removed in 2012 reportedly due to funding cuts.

The boundary layer scaling approach involves the calculation of a reference climatology valid near the top of the IBL which can then be downscaled using several steps in order to estimate the mean wind speed at the turbine hub height. An overview of this methodology is shown in Figure 3.1 and a further discussion of each stage is given below.

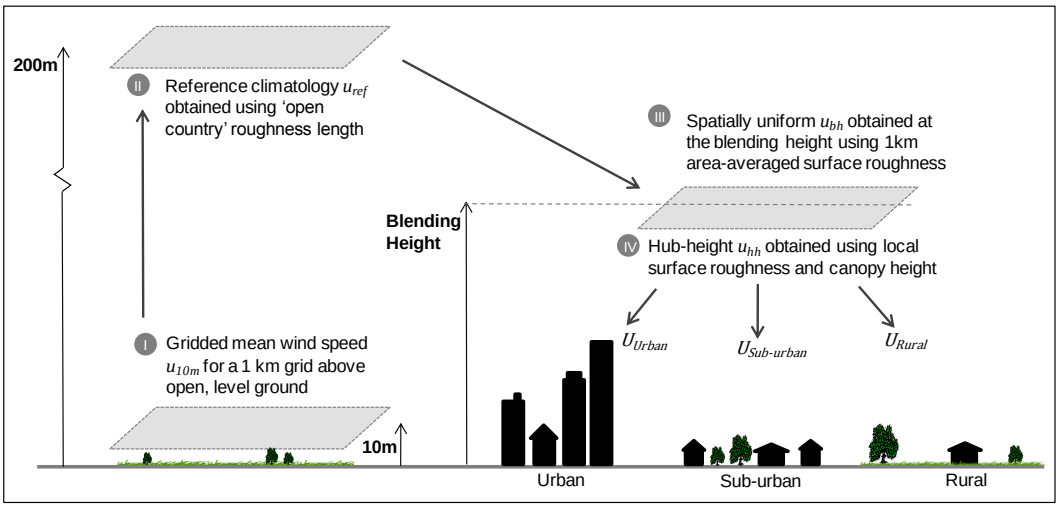

Figure 3.1: Outline of the methodology employed by the Met Office to predict the spatially averaged mean wind speed. Starting with an input mean wind speed (I) and culminating in a predicted mean wind speed at turbine hub height (IV).

3.1.1 Large-scale reference climatology

As a first input, climatology data produced by the Met Office's National Climate Information Centre (NCIC) is used. The database consists of mean wind speeds at a height of 10 m above ground level, estimated over 1 km grid squares using geostatistical interpolation. This technique takes into account elevation, the proportion of sea within a 5 km radius and terrain shape in calculating the gridded averages from surface measurements [94]. The NCIC database includes surface measurements covering 30 years and 220 stations compared to only 10 years and 56 stations for the widely used NOABL database [10].

In order to convert this to a large-scale reference climatology, the gridded mean wind speeds must be transformed upwards to a reference height where the flow can be considered independent of the local surface. In the Met Office methodology, a reference height of 200 m and a roughness length of z_0 = 0.14 m (representative of 'open country') are used. Note that while a smaller roughness length (~0.03 m) would be expected for short grass, extended regions of short grass without interruption from hedges and bushes are relatively unlikely in UK rural areas [15]. Hence, in practice, z_0 = 0.14 m is considered more representative of open country.

The reference height is, in effect, an estimate of the height of the boundary layer over a rough surface. Strictly, this height will not be fixed and should be determined as a function of fetch and surface roughness using an expression for boundary layer growth [58, 61, 88]. However, such an approach is challenging to implement in an automated tool since it requires potentially subjective estimates of significant roughness changes. Fortunately, the predicted wind speed close to the surface appears to be relatively insensitive to the exact choice of this height provided a sensible value is used [15]. In addition, the wind speeds at 10 m above the surface adjust relatively quickly to major roughness changes with the most significant adjustments occurring within 1 km [14]. However, the effect of using a fixed height for the internal boundary may be more problematic close to a roughness discontinuity such as at a coastal site or close to the edge of a city, as discussed in Section 2.4.

The vertical transformation to the reference height is achieved by scaling the vertical logarithmic profile described in Section 2.4.2:

$$u_{ref} = u_{10m} \frac{\ln(z_{ref}/z_0)}{\ln(10/z_0)}$$

Equation 3.1

where z_{ref} is the reference height (200 m), u_{ref} is the reference wind speed at this height and u_{10m} is the wind speed at a height of 10 m obtained from the NCIC database.

3.1.2 Wind speed at the blending height

Given a large-scale reference climatology, the next step is to transform this to a mean wind speed representative of the regional area. This is an average over a 1 km grid square at the blending height, taking account of the effective surface roughness on a regional scale as described in Section 2.4.4. To achieve this, a second logarithmic profile is applied of the form [15]:

$$u_{bh} = u_{ref} \frac{\ln\left[(z_{bh} - d_{eff})/z_{0eff}\right]}{\ln\left[(z_{ref} - d_{eff})/z_{0eff}\right]}$$

Equation 3.2

where u_{bh} is the wind speed at the blending height z_{bh}, and z_{0eff} and d_{eff} are the effective roughness length and effective displacement height respectively.

At this point, we are faced with the difficulty of choosing appropriate values for z_{0eff} and d_{eff} to represent the combined effect of multiple patches with differing roughness. This can be achieved by considering the land cover for the grid-square of interest in terms of specific surface categories. Categorised UK land cover data are available from the Centre for Ecology and Hydrology (CEH) [95]. In general, land cover categories can be linked to the aerodynamic parameters of z_0 and d using standard tables. However, relating surface cover to roughness parameters is non-trivial in the case of the built environment. In the current methodology, surface parameters of h_m, λ_f and λ_p are obtained for the sub-urban and urban categories based on values associated with typical UK cities [75]. These are subsequently used along with the method of Raupach [83] in estimating urban and sub-urban aerodynamic parameters.

After obtaining appropriate estimates of z_0 and d, the fraction of each surface type within the grid square is used to calculate a grid-box average roughness length according to the blending method described in Section 2.4.4. The effective displacement height is simply taken to be the maximum of those identified within the grid box. The blending height is taken to be the larger of 10 m or twice the maximum canopy height within the grid square, in line with observations over rough surfaces described in Section 2.4.6. Note that the definition of this height is subject

to some uncertainty as described previously. Finally, these parameters are applied to Equation 3.2 to estimate the blending height mean wind speed u_{bh}.

3.1.3 Wind speed at hub height

The final downscaling step transforms the wind speed at the blending height to the spatially averaged mean wind speed at the turbine hub height, taking into account the local surface properties. In the original Met Office methodology [15], three surface classes were considered, 'rural', 'sub-urban' and 'urban'. However, these classes were extended to include four classes of urban/sub-urban surfaces from 'low height and density' to 'very high height and density', as well as a 'woodland' class when implemented in the WYET. In all cases, a simple extrapolation of the logarithmic profile is used with local aerodynamic parameters defined by the surface class.

For the rural or open countryside case, the final downscaling is achieved by simply applying the following expression: [15]:

$$u_{hh} = u_{bh}\frac{\ln(z_{hh}/z_0)}{\ln(z_{bh}/z_0)}$$

Equation 3.3

where z_{hh} and u_{hh} are the turbine hub height and associated mean wind speed respectively.

For the remaining cases, the logarithmic profile is extrapolated below the blending height to the turbine hub using the local roughness length (z_{0local}) and displacement height (d_{local}) as suggested by Cheng and Castro [66]. Hence the expression becomes [15]:

$$u_{hh} = u_{bh}\frac{\ln[(z_{hh} - d_{local})/z_{0local}]}{\ln[(z_{bh} - d_{local})/z_{0local}]}$$

Equation 3.4

where z_{0local} and d_{local} depend on the local surface.

Note that this simple extrapolation below the blending height ignores the full complexity of the flow in this region as described in Section 2.4.

For sites within the built environment where the hub height is below the canopy top, an exponential relationship is used which takes into account the frontal area density as recommended by Macdonald [80]:

$$u_{hh} = u_h exp\left[-9.6\lambda_f \frac{h - z_{hh}}{h}\right]$$

Equation 3.5

where u_h is the wind speed at the canopy top, as calculated by Equation 3.4, h is the canopy height and $9.6\lambda_f$ represents an empirical estimate of the attenuation coefficient [80].

Figure 3.2 summarizes each stage of the downscaling process and the approximate shape of the vertical wind speed profiles in each sublayer.

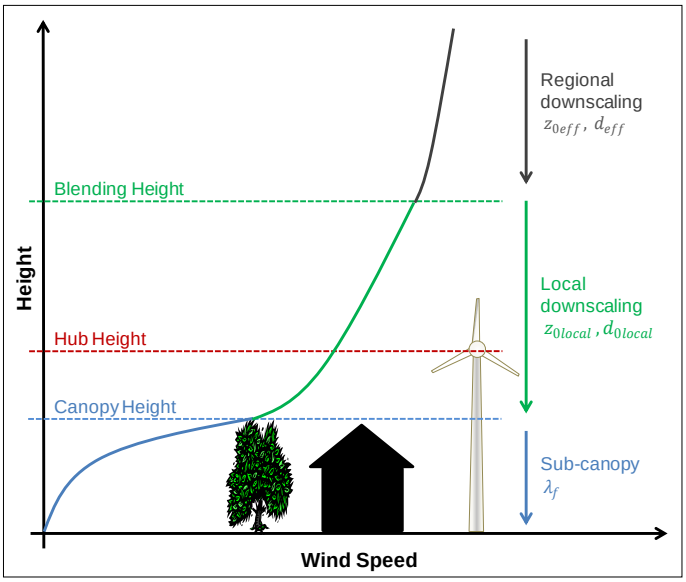

Figure 3.2: Summary of the regional and local downscaling of the mean wind speed. The curves show the approximate shape of the vertical wind profile in each layer.

Given the predicted mean wind speed at the turbine hub height, a prediction of the wind resource can be made by assigning a specific wind speed frequency distribution. The Met Office approach assumes a Weibull distribution with a fixed shape factor of 1.8 based on an observed range of 1.5-2.1, and a mean wind speed defined by u_{hh}. This distribution may be combined with a suitable turbine power curve to calculate the available energy.

3.1.4 Limitations

The accuracy of this prediction methodology is limited by simplifications and sources of uncertainty that are present at each stage. When combined, these can lead to large errors in the estimation of the wind resource at a specific site. These uncertainties may be divided into those affecting predictions in all terrains and those that are terrain specific, as listed below:

Uncertainties affecting all terrains

(i) The reference climatology suffers from uncertainties due to the interpolation process and unresolved orography on scales below 1 km (Section 2.4.7).

(ii) The reference climatology contains no directional information and this prevents the directional variations in the upwind surface roughness from being properly accounted for.

(iii) Regional aerodynamic parameters are based on land cover in a local region of 1 km^2. However, considerations of boundary growth imply that the mean wind speed will be affected by the upwind roughness over a much larger fetch (Section 2.4.3).

(iv) Wind flows below a fixed blending height are assumed to be fully in equilibrium with the local surface and described by a logarithmic wind profile. This ignores many complexities related to developing IBLs (Section 2.4.4).

(v) The methodology predicts only a temporally averaged mean wind speed, thus requiring assumptions regarding the form of the wind speed distribution in order to make power predictions. This is ultimately a limitation of the reference climatology which contains no information regarding the distribution of wind speeds.

Uncertainties affecting specific terrains

(i) The use of a fixed IBL height of 200 m does not fully account for edge effects close to roughness boundaries. These will be particularly important for coastal sites (Section 2.4.5) as well as sites close to the rural/urban boundary (Section 2.4.6).

(ii) The logarithmic scaling assumes neutral stability conditions. This is unlikely to be appropriate at coastal sites, which are subject to complex thermal effects (Section 2.4.5).

(iii) There are large uncertainties in the estimation of aerodynamic parameters for real urban areas that are not accounted for using simple impressions of UK cities (Section 2.4.6).

(iv) For urban areas, the definition of the canopy height and mean building height are ambiguous due to surface heterogeneity. This affects the scaling of the roughness parameters as well as the vertical wind speed profile close to the canopy top.

These uncertainties indicate that while the methodology is attractive since it is relatively simple to implement and requires no direct wind speed measurements, it must be applied with caution when making site-specific wind resource assessments. In addition, these uncertainties offer opportunities for improvements to the approach. As discussed in Chapter 4, some of these improvements can be achieved relatively simply, while others require the development of more sophisticated techniques and input data.

3.2 Data-Driven Measure-Correlate-Predict Approaches

As outlined above, simple boundary layer scaling approaches, whilst being rapid and cheap to implement, are unable to account for many of the complexities of real-world boundary layer wind flows. An alternative to such analytical methods are data-driven approaches where the collection of onsite wind data forms part of the resource assessment procedure. In the large-scale wind industry, detailed on-site wind speed measurements are the standard before any investment decision is taken. Typically this involves 1-3 years of onsite monitoring in addition to correlation to a long-term reference site using measure-correlate-predict (MCP) [8]. Although this may not be practical in the case of small wind turbine installations, a shorter on-site monitoring program combined with correlation to a reference site has the potential to provide wind resource estimations of greater rigour compared to simple boundary layer scaling approaches.

Although MCP approaches may be used in an ad-hoc basis in the small-scale wind industry, it is a technique primarily associated with large-scale wind resource assessment. Thus, little work has been done to test and formalise MCP approaches used in the context of small-scale wind energy. This area of application is hence ripe for further investigation.

The MCP strategy is shown schematically in Figure 3.3 and may be summarised by the following three stages:

1. **Measure** wind speeds at the target site, as close to the location and height of the proposed installation as possible, over a short training period.
2. **Correlate** to wind data from a local long-term reference site, such as a meteorological station or airport.
3. **Predict** the long-term resource by extrapolating from the on-site measurements using the correlated reference data.

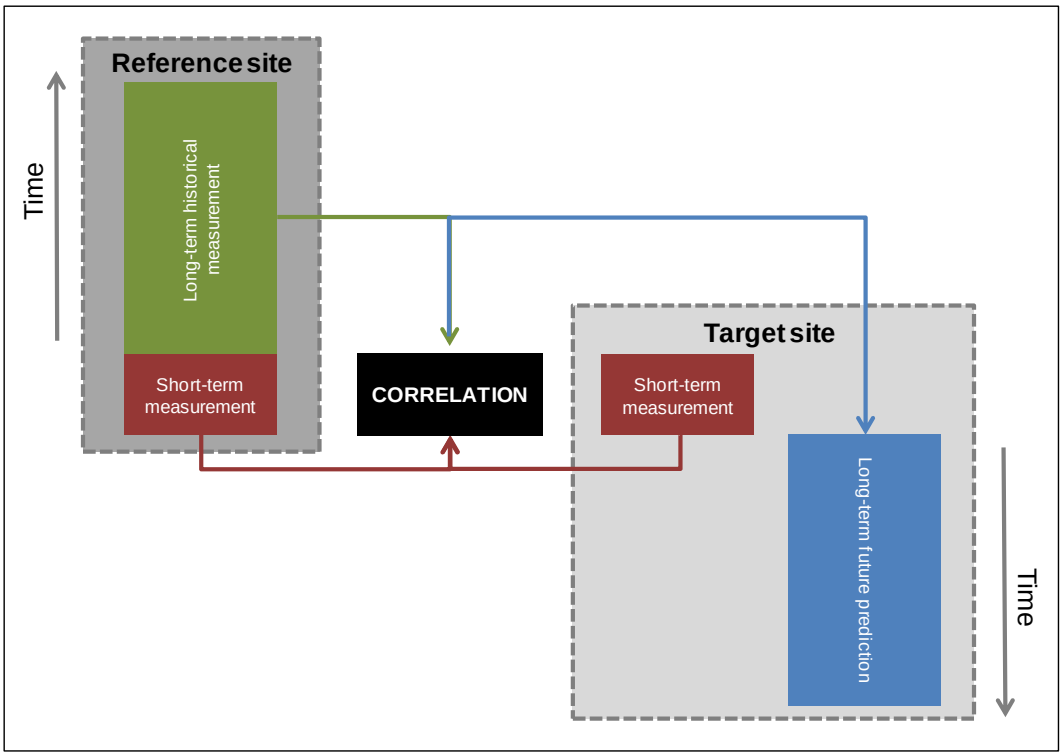

Figure 3.3: Schematic diagram of the measure-correlate-predict process.

Generally, the aim of an MCP approach is to find a correlation between short-term concurrent measurements taken at a reference and target site. The correlation is applied to historical data from the reference site in order to construct a long-term time series of wind speeds (and possibly directions) at the target site. From this time series it is then possible to extract parameters describing the wind speed distribution and the long-term energy potential of the target site. In this thesis, the short-term, concurrent measurement period at the reference/target site pair is denoted as the training period since it is used to establish the relationship between the two sites.

The literature related to MCP is broad including numerous peer reviewed studies, conference proceedings and technical reports. There are almost unlimited methods for describing a relationship between correlated variables and this has led to a large variety of proposed MCP techniques, as illustrated by a recent review by Carta *et al.* [18] which considered over 150 studies. The usefulness of these studies varies enormously. For example, not all are peer reviewed, some draw conclusions based on application to a single reference/target site pair, others use data from a single year to test new approaches and many consider predictions of mean wind speeds whilst failing to investigate the predicted distribution of wind speeds which is required to estimate wind power. Despite this, many detailed studies do exist and even those that are less rigorous may provide useful hints for directing future work.

In reviewing this material, however, it is prudent to give more weight to studies which include predictions of the distribution of wind speeds or the wind power rather than just the mean wind speed. In addition, studies applied to a small number of sites or those that are not tested using wind data collected over multiple years should be treated with caution.

The following sections are intended to provide an overview of the major classes of MCP approaches that have been proposed. Particular focus is given to approaches that have been rigorously tested and that are already widely used in the wind industry. Since there is a lack of studies investigating the application of MCP to small-scale wind installations, Section 3.2.5 considers studies that have applied MCP to short measurement periods, since this is the major challenge in such an application. While the review is not intended to be exhaustive, some variations to established techniques, as well as promising new approaches, are considered with an emphasis on techniques that may be best suited to the small-scale wind industry.

3.2.1 Preliminary considerations

While it is not appropriate here to outline detailed site assessment protocols, it is informative to highlight the key assumptions of the MCP approach.

Application of MCP is based on the hypothesis that the wind climate at the reference and target sites is similar. If the two sites experience very different mesoscale phenomena for example, it is unlikely that the reference site will serve as a good predictor for the target site without the use of special measures [96]. Ensuring similarity at the reference/target site pair is generally achieved by choosing site pairs with the minimum separation possible and avoiding site pairs with obviously different climates (coastal paired with land-locked mountains for example), although such considerations are always subject to data availability. Frequently, a metric such as the linear correlation coefficient (r) between concurrent wind speeds at the two sites is used as a measure of similarity [18]. There is, however, some debate as to the usefulness of such a metric due to the possibility of non-linear relationships or time-of flight delays that may lead to low values of r even for highly correlated sites [97]. In addition, the calculated value of r may be unreliable when using training periods of less than 12 months due to the limited length of target site data.

Since MCP is applied to historical data at the reference site, there is also an implicit assumption that the future wind resource can be predicted by the past. While this

may be a reasonable approximation, studies of historical wind data [98] do show variability on decadal timescales and this could potentially impact MCP estimates. In the case of MCP applied to training periods of less than 12 months, however, these uncertainties are likely to be small compared to those associated with the short correlation period.

A final point worthy of note is that depending on the approach, the MCP predictions may directly predict the target site wind direction, or else, make some assumption that the long-term distribution of wind directions is represented by the short-term distribution at the target site or the long-term distribution at the reference site [99]. In the case of large-scale wind farms, this distribution is important in wind farm design and properly accounting for wake effects. For small scale installations, however, while the distribution of target site wind directions is of potential use in micro-siting decisions, it is of lesser importance. In the work presented in this thesis, prediction of the target site wind directions is not considered a key requirement, although some consideration is given to this issue in Chapter 6.

3.2.2 Regression approaches

3.2.2.1 Climatological adjustment

Perhaps the earliest and simplest use of long-term reference data in an MCP approach is the so called ‘climatological adjustment’ proposed by Putnam [100, 101]. The method involves simply scaling the long-term reference site mean wind speed by the ratio of the short-term means at the target and reference sites. The result is a prediction of the long-term mean wind speed at the target site. A simple extension [102] is to apply the same scaling to the long-term time series of wind speeds at the reference site using the expression:

$$\hat{u}_{tar,i} = \frac{\bar{u}_{tar}^{ST}}{\bar{u}_{ref}^{ST}} u_{ref,i}$$

Equation 3.6

where $\hat{u}_{tar,i}$ and $u_{ref,i}$ are the i^{th} predicted target site and observed reference site wind speeds respectively, and $\bar{u}_{tar}^{ST}$ and $\bar{u}_{ref}^{ST}$ represent the short-term mean wind speeds observed at the reference and target sites during the training period.

This method improves on climatological adjustment in that it can be used to predict the full long-term time series, and hence, the wind speed distribution at the target site. Several variations on simple ratio methods have also been proposed [101],

including those which use an estimate of the variance at the target and reference sites, although there is little to distinguish such methods from linear regression.

3.2.2.2 Linear regression

Regression based MCP approaches are widely used in the wind industry, presumably due to their ease of implementation and empirical success [103-105]. A typical approach [99, 106, 107] involves the application of sector-wise linear regression (LR) to wind speed observations that have been binned into angular sectors according to the reference site wind direction. The sector approach accounts for the fact that the relationship between the reference and target site may vary with wind direction due to angular variations in roughness, orography or local obstacles.

For each sector, the individual wind speeds at the target site are described by the equation:

$$u_{tar} = \alpha + \beta u_{ref} + \varepsilon$$

Equation 3.7

where u_{tar} is the observed wind speed at the target site, u_{ref} is the corresponding observed wind speed at the reference site, α and β are the regression coefficients obtained using a fit to the training data and ε is an error term which represents the residual scatter. Fitting is frequently achieved by minimising the sum of squares of the residuals (least squares method) although other techniques exist including methods based on minimising the perpendicular distances of the observations from the fitted line [18, 102].

The MCP approach involves collecting target site wind data for sufficient time to obtain the regression coefficients, which are assumed to be constant with time. Assuming the data are well represented by a linear fit, the mean prediction at the target site given a reference wind speed u_{ref} is:

$$\hat{u}_{tar} = \alpha + \beta u_{ref}$$

Equation 3.8

Bardsley and Manly [108] derived estimators for the long-term mean wind speed and variance of wind speeds using such an approach more than three decades ago. Interestingly, they also pointed out that while linear regression produces unbiased estimates of mean wind speed, estimates of the width of the wind speed distribution, as described by the variance, are not unbiased. This issue is related to

the contribution to the variance of the residual scatter term ε in Equation 3.7 and is an important observation since this will lead to biased wind power predictions.

In recent years, LR has been investigated by many authors and is frequently used as a baseline against which to assess the performance of alternative MCP approaches [106, 109, 110]. It is noteworthy that although the importance of the residuals ε in estimating wind power is known, [18, 103, 108], it is often not explicitly considered in scientific studies. This can lead to unfair comparisons between alternative MCP approaches and LR approaches where ε is not accounted for, particularly with respect to wind power predictions.

3.2.2.3 Variance ratio method

In addition to the LR approach, the variance ratio method (VR) has been shown to be an MCP approach of wide applicability. The approach was proposed by Rogers *et al.* [99] in response to the observation that in the simple LR case, where no account is taken of the residual scatter ε, the standard deviation of the predicted wind speeds $\sigma(\hat{u})$ will be smaller than the standard deviation of the observed values $\sigma(u)$ by a factor r, where r is the linear correlation coefficient. Here $\hat{u}$ and u are used to indicate that the standard deviations are a function of the predicted or observed wind speeds respectively. The underestimation of the standard deviation will impact on the shape of the predicted wind speed distribution and hence will result in an under prediction of the available wind power at the target site. Note this is an equivalent observation to that made by Bardsley and Manly [108] mentioned previously. To account for this, Rogers *et al.* [99] proposed a variation of simple LR based on setting $\sigma(\hat{u}) = \sigma(u)$. Using least squares theory it can be shown that this condition results in a linear equation of the form:

$$\hat{u}_{tar} = \left[\bar{u}_{tar} - \left[\frac{\sigma_{tar}}{\sigma_{ref}}\right]\bar{u}_{ref}\right] + \left[\frac{\sigma_{tar}}{\sigma_{ref}}\right]u_{ref}$$

Equation 3.9

where σ_{tar} and σ_{ref} represent the standard deviation about the mean wind speeds at the target ($\bar{u}_{tar}$) and reference ($\bar{u}_{ref}$) sites respectively, as calculated from the short-term training data. Note that in this approach no direct attempt is made to model the residual scatter term ε. Instead, Equation 3.9 postulates zero scatter (in effect, r is set to unity) resulting in a gradient which is a factor of $1/r$ larger than that obtained using simple LR thus forcing the required increase in $\sigma(\hat{u})$.

Rogers *et al.* [99] tested the performance of the VR approach in terms of the predicted mean wind speed, wind speed distribution parameters and wind power using eight target sites in a variety of terrains. The results showed notably improved predictions using VR compared to LR particularly for parameters related to the wind speed distribution. The approach has since been utilised in a number of studies [110-112] and along with LR is frequently used as a benchmark against which to compare new MCP techniques.

While the approach appears to produce surprisingly accurate results considering its simplicity, it is perhaps somewhat unfair to directly compare the predictions to LR where no account is taken of the residual error term ε which is known to produce biased estimators. Standard linear regression theory [113] postulates normally distributed residual errors with constant variance and hence a more considered approach should attempt to include a representation of these within the LR technique to provide a more suitable comparison.

3.2.2.4 Mortimer/matrix method

Whilst not strictly a regression approach, the method suggested by Mortimer [114] based on the ratios of binned wind speeds has been increasingly applied to wind resource assessment [99, 103]. In this approach, data are binned with respect to wind speed and direction to produce a matrix of mean ratios between the target and reference site wind speeds. A second matrix is constructed containing the standard deviation of these ratios and this is used to calculate a triangular distribution centred on the mean ratio. The target site wind speeds are then predicted using:

$$\hat{u}_{tar} = (\gamma + e)u_{ref}$$

Equation 3.10

where γ is the ratio of wind speeds and e is a random number drawn from the triangular distribution. Note that unlike LR and VR this approach allows for non-linear relationships between the reference and target site wind speeds. Mortimer applied the method to hourly wind data covering 10 months from an unspecified reference and target site pair in Scotland and observed improved performance at high wind speeds compared to linear regression. More comprehensive tests were carried out by Rogers et al. [99] as well as the commercial organisation Renewable Energy Systems [102], both of which indicated improved performance of the Mortimer method compared to simple sector-wise linear regression approaches. Woods and Watson [115] also used a matrix representation of wind directions in order to correct for wind veer. Their results indicated that such a method could

provide improved predictions of the angular sector populations in complex terrain compared to the standard assumption of equivalent sector populations at the reference and target sites.

Matrix methods have the advantage of allowing the relationship between the reference and target sites to be non-linear and to vary with small changes in wind speed and direction. Inclusion of the e term also allows for scatter in these relationships. However, as pointed out by Derrick [107], achieving sufficient data coverage for each matrix element (i.e. wind speed and direction bin) in order to produce reliable correlations can be challenging. While methods such as interpolation may address this, it is likely to be a particularly challenging issue in the case of very short onsite measurement periods and hence it may not be well suited to small-scale wind resource assessment.

3.2.2.5 Alternative regression approaches

In addition to LR and VR, a number of alternative regression approaches have been proposed. For example, Derrick [107] suggested a power relationship of the form $u_{tar} = \alpha(u_{ref})^{\beta}$ may sometimes be appropriate. Clive [116] provided some theoretical basis for such an approach, as well as analytical expressions for β based on the assumption of Weibully distributed speeds. Other authors have suggested linear regression of orthogonal wind vectors [106]. This approach involves transforming the time series of wind speed measurements into eastern and northern components and performing linear regression on each component. The predicted orthogonal components can then be recombined to estimate the resultant wind vector with the advantage of predicting both wind speed and direction.

Achberger *et al.* [106], building on the work of Hanson *et al.* [117] suggested a method known as vector correlation. Using a vector representation of the target and reference site wind speeds it is possible to derive expressions for the scaling, veer and translation of the wind vector between the reference and target sites. While this method is capable of predicting both wind speed and direction at the target site, the approach imposes a single veer angle between the mean wind vectors at the reference and target sites. In reality, any vector rotation may be dependent on many factors including wind speed, wind direction and synoptic conditions [107]. Achberger *et al.* [106] compared the performance of linear regression of orthogonal wind vectors, vector regression and standard linear regression using wind data from a single target site covering a period of approximately nine months. The techniques did not appear to result in significant prediction differences, although the study was too limited to draw robust conclusions.

Nielsen *et al.* [118] proposed a further technique based on expressing measured wind speed and direction in the form of two orthogonal vectors. A two-dimensional linear regression is then performed using:

$$\begin{bmatrix}\hat{u}_E\\ \hat{u}_N\end{bmatrix} = \begin{bmatrix}\alpha_E\\ \alpha_N\end{bmatrix} + \begin{bmatrix}\beta_{EE} & \beta_{EN}\\ \beta_{NE} & \beta_{NN}\end{bmatrix}\begin{bmatrix}u_{E,ref}\\ u_{N,ref}\end{bmatrix}$$

Equation 3.11

where u_E and u_N represent orthogonal easterly and northerly wind vectors. Despite the fact that the method is capable of accounting for both the wind speed and direction at the target site, in practice it was found that the technique resulted in biased wind speed predictions. This bias has been attributed to an underestimate of the target site variance, as was noted for simple linear regression [99].

A regression approach coined 'Weibull regression' was recently investigated by Perea *et al*. [110]. In this approach, the correlated wind speeds at the reference and target sites are assumed to be sampled from a joint Weibull distribution. Such an approach yields a non-linear regression curve that can be used to predict the target site wind speed given an observation at the reference site. However, tests on artificially created wind data yielded no significant improvement compared to LR.

Overall, the results of these studies appear to offer no clear mandate for using alternative regression methods in place of the more widely studied LR or VR approaches.

3.2.3 Distribution-based approaches

While regression approaches are often convenient to implement, unless the residual scatter is modelled, they are limited in that they predict a single-valued target site wind speed for each input reference site wind speed. However, two-dimensional frequency plots of reference/target wind speeds readily demonstrate that this is a significant simplification. Given a series of identical reference site wind speeds, the target site wind speeds are likely to vary considerably about the mean regression line due to the stochastic nature of wind flows. Hence, the reference/target site relationship is more accurately represented by a two-dimensional joint distribution of wind speeds at the reference and target sites rather than a single regression line.

3.2.3.1 Linear regression with scatter

An assumption of linear regression is that the variability, or residual scatter ε, about the regression line is represented by a zero-mean Gaussian distribution of the form [113]:

$$\varepsilon \sim \mathcal{N}(0, \sigma_{res}^2)$$

Equation 3.12

In the MCP approach σ_{res} represents the sample standard deviation of the residuals about the predicted target site wind speeds $\hat{u}_{tar}$, as calculated from the training data using [113]:

$$\sigma_{res} = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(u_{tar,i} - \hat{u}_{tar,i})^2}$$

Equation 3.13

These properties can be used to reconstruct the variability using a Gaussian scatter model. As an example, Figure 3.4 shows a scatter plot of short-term wind speed observations at a pair of correlated sites, the colour shading indicates the frequency of observations in wind speed bins with approximate width 0.5 ms^{-1}. A linear fit is shown by the solid line and Gaussian distributions conditioned on the average standard deviation of the residuals σ_{res}, are shown schematically at several wind speeds.

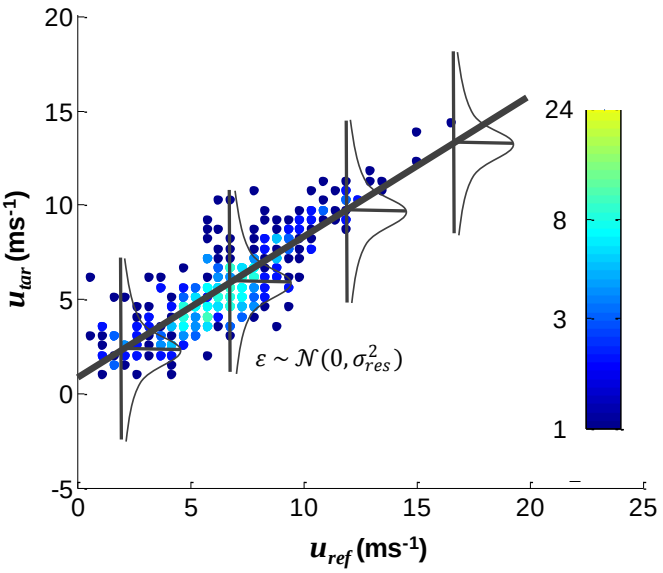

Figure 3.4: Schematic representation of a Gaussian scatter model applied to linear regression. The solid line represents a linear fit to the data, the shading represents frequency within a wind speed bin and the distributions represent zero mean Gaussians conditioned on the residuals.

The variability in u_{tar}, given a fixed u_{ref} is accounted for in the long-term prediction by adding samples drawn from the Gaussian distribution described by Equation 3.12 to the mean prediction (solid line, Figure 3.4) using either a long time series or a Monte-Carlo sampling technique. The approach assumes that the residual variance is not dependent on wind speed and that its magnitude can be estimated from the short-term measurement period. Both these assumptions are likely to be simplifications. In particular, observations at high wind speeds are often sparse, as demonstrated in Figure 3.4. This makes it necessary to extrapolate low wind speed estimates of σ_{res} across the whole wind speed range. In addition, for training periods covering just a single season, the long-term representativeness of the residual scatter estimated from the training period will be sensitive to seasonal variability in the wind climate. This situation could potentially be improved through the use of a seasonal correction factor.

The Gaussian scatter model can be thought of as a first approximation to reproducing the two-dimensional probability distribution since the resulting target site predictions will no longer be single-valued for a given reference site wind speed. Instead they will follow a distribution of wind speeds centred on $\hat{u}_{tar}(u_{ref})$. Modelling the scatter in this way is expected to provide a more accurate representation of the target site wind speed distribution and will thus reduce the negative bias in predicted wind power that occurs with simple LR. Despite these advantages, this approach is not routinely implemented when LR is applied to MCP

and there appear to be few, if any, systematic studies reporting the performance of such an approach.

While such a model is in line with the standard assumptions of linear regression, it should be noted that it represents a simplification in the case of jointly distributed wind speeds at correlated reference/target site pairs. Such correlations are expected to result in Weibull rather than Gaussian distributions [110, 116] and this topic is discussed in more detail in the following section as well as in Chapter 7.

3.2.3.2 Empirical multivariate probability distributions

A more rigorous treatment of the issue of scatter in the reference/target site relationship is to model the full conditional probability distribution of reference and target site wind data. Unlike regression techniques, this approach seeks to directly model the underlying distribution of wind speeds (and possibly wind directions) rather than a historical time series.

For example, given a set of two, correlated, random variables, their relationship may be described in terms of a bivariate probability density function (pdf). The height of the pdf surface at a point describes the probability of observing a particular combination of variable pairs. The distribution can be thought of as being composed of a series of one-dimensional, conditional probability distributions, or vertical slices, through the two-dimensional probability surface. Each vertical slice describes the probability of observing particular values of one variable given a fixed value of the second variable. In addition, the conditional probability slices can be integrated across one of the variables to yield the marginal, or complete, distribution of the other variable.

The approach can be thought of as a generalisation of the linear regression Gaussian scatter model since rather than the restriction that a specific reference site wind speed corresponds to a specific target site wind speed, a conditional distribution of target site wind speeds is predicted, conditioned on the reference site observation. However, the approach is more general in this case since the conditional distributions are not constrained to be univariate Gaussian. Since wind power is proportional to the cube of the wind speed, accurately predicting the form of these conditional distributions is likely to be important in achieving accurate wind resource predictions.

Garcia-Rojo [119] applied an approach based on the multivariate probability distribution of the reference and target site wind directions and wind speeds to predict the long-term wind resource at two target sites in Spain. The approach used

the short-term, joint probability mass function obtained from a training period of approximately 5 months. The probability mass function is simply an empirical representation of the discrete, joint probability of observing some combination of target and reference site wind speeds and directions, as calculated from the observed frequency of that combination. The short-term probability function was used along with the long-term distribution at the reference site to predict the mean wind speed and distribution of wind directions at the target site over a period of 4 years. The approach is claimed to offer advantages over regression methods although the results were not compared with other MCP methods. Casella [120] extended the method by defining a multivariate probability distribution between three masts (two reference and one target), for a single target site in Australia. The study, based on one year of data, concluded that the use of two reference masts improved predictions but again, only mean wind speeds and directions were investigated and no comparisons were made with other MCP approaches.

Unfortunately, these studies did not investigate the accuracy of the predicted wind speed distribution or power density despite the fact that the joint probability approach is likely to have most utility in the prediction of these parameters. In addition, the number of test sites and the length of the long-term data used were both very limited. A further point of note is that while approaches based on the use of empirical distribution functions constructed directly from the observed wind data are easy to implement, they are limited in that they give equal weight to all data, including outliers [116]. Hence, methods based on mathematical functions may be more appropriate despite the fact that they require increased complexity and some sacrifice in flexibility. Such methods are discussed in the following section.

3.2.3.3 Bivariate probability functions with fixed mathematical form

In the case of wind speeds observed at a pair of correlated reference and target sites, the conditional probability density may be expressed as [18]:

$$f\left(u_{tar}\middle|u_{ref}=u'_{ref}\right)=\frac{f(u'_{ref},u_{tar})}{f(u'_{ref})}$$

Equation 3.14

where u_{ref} and u_{tar} represent wind speed observations at the reference and target sites respectively and u'_{ref} is a specific value of u_{ref}, $f(u_{ref},u_{tar})$ is the bivariate pdf and $f(u_{ref})$ represents the univariate pdf at the reference site.

The marginal pdf at the target site, $f(u_{tar})$, is obtained by integrating the product of the conditional pdf in Equation 3.14 and the marginal pdf at the reference site, $f(u_{ref})$, over all reference site wind speeds using [18]:

$$f(u_{tar}) = \int f(u_{tar}|u_{ref} = u'_{ref})\, f(u_{\boldsymbol{ref}})\mathrm{d}u_{\boldsymbol{ref}}$$

Equation 3.15

The marginal pdf of wind speeds at the target site $f(u_{tar})$, represents the key descriptive quantity of the target site wind resource.

Given an appropriate choice of bivariate probability function, this approach can be used to directly predict the long-term target site wind speed distribution using a short-term concurrent measurement period at the reference and target sites and long-term observations at the reference site. The use of probability distributions with a fixed mathematical form may have some advantage over the use of empirical distributions since the probability function may be evaluated even in regions of sparse data coverage. However, data sparsity when using short training periods may still present problems related to obtaining an adequate functional fit to the observed data, particularly at high wind speeds as shown in Figure 3.4.

Carta and Velázquez [121] used contingency-type distributions (a class of bivariate distributions constructed from specified marginal distributions) to implement this approach at six sites in the Canary Islands. The approach involves first fitting the short-term training data to a fixed distribution function, before using the function, along with long-term observations at the reference site to predict the long-term wind resource at the target site. The success of the approach was compared with linear and non-linear regression methods as well as the multivariate method of Garcia-Rojo [119] described above. In terms of the predicted wind power, the approach generally outperformed the regression methods but failed to outperform the multivariate Garcia-Rojo method. One limitation of the study was that while the long-term test period covered eight years, a single, fixed year was used for training the algorithms. Hence, variations in the relative performance of the algorithms across different years were not tested. In addition, the sites used were all located in the archipelago of the Canary Islands and the marginal distributions used to construct the bivariate contingency-type distributions were known to be representative of that particular climate. Thus, it is not clear how transferrable the results may be.

A related, theoretical study was carried out by Perea *et al*. [110] using artificially generated wind data. The study observed that widely used linear regression

methods implicitly assume a bivariate Gaussian distribution between reference and target site wind speeds, while experience shows that the Weibull distribution is generally more appropriate, an observation also made by Clive [116]. Based on this observation, a bivariate Weibull probability approach was proposed to model correlated wind speeds at two sites. The distribution requires five parameters that describe the shape and scale of the distributions at the two sites as well as the strength of the correlation. An example of such a distribution is shown in Figure 3.5 where the target site conditional distributions are also highlighted.

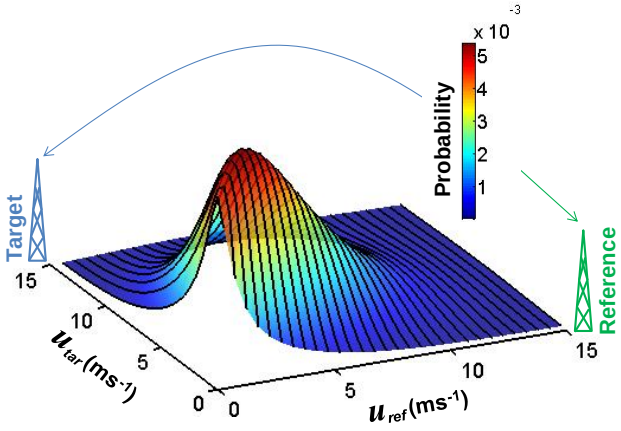

Figure 3.5: A bivariate Weibull probability distribution of wind speeds at a pair of correlated sites. The solid lines trace the target site conditional distributions for increments of 0.5 ms^{-1} in the reference site wind speed.

Using the artificial data, the performance of the bivariate Weibull approach was compared by Perea *et al.* to conventional linear methods, as well as a bivariate Gaussian approach, and in all cases shown to outperform the other methods. The bivariate Weibull approach is attractive in that it provides a strong theoretical basis for modelling correlated wind speeds while avoiding the possible problems associated with empirical probability distributions. However, since the approach has only been tested using artificial data, it is not yet clear how such a method may perform in the real world where wind data are unlikely to conform to idealised Weibull distributions. This issue, as well as the performance of the approach using short-term training periods is worthy of further consideration and is returned to in detail in Chapter 7.

3.2.4 Learning-based approaches

The MCP methods considered thus far, except in the case of the empirical probability distribution approach, have assumed some fixed functional relationship between the reference and target site wind data. Thus, the main task in applying

such methods has been to identify a suitable functional relationship and to estimate the parameters of this function which best map the reference site wind observations onto the target site observations.

An alternative is to apply a learning based method where fewer assumptions are made about the specific relationship between the reference and target sites. Instead, the main task is to use a training data set to establish patterns between wind observations at the two sites. If successful, these patterns can be applied to new observations at one of the sites in order to make predictions at the second site. Since this is exactly the task required in the implementation of the MCP approach, there has been some interest in applying learning based methods to MCP.

3.2.4.1 Artificial neural networks

Artificial neural networks (ANN) are perhaps the most widely studied learning-based MCP approach. ANNs are so-called in that they can loosely be thought of as mimicking the learning processes of biological neurons [122]. In simple terms, the ANN architecture involves a series of neurons organised into layers. Each neuron is interconnected through synapses which represent the application of a mapping or transfer function [122]. A given neuron receives input from a series of interconnections and these inputs are combined in a weighted sum. The output of an individual neuron is a function of this weighted sum.

The role of the ANN is to establish patterns between input and output data, given a training data set of known inputs and outputs. This can be achieved by adjusting the weights applied to each connection for example such that, given the input training data, the ANN is able to reproduce the outputs. A successfully trained ANN is then able to generalise these patterns to new data in order to make predictions. This is in effect an analogue of the MCP process whereby short-term reference and target site wind measurements are used to establish a relationship or pattern between two sites. New data (long-term reference site observations) are then introduced to make long-term target site predictions.

As with many studies related to MCP, it is difficult to reach objective conclusions about the general success of ANNs compared to other MCP techniques due to the limited scope of many studies, both in terms of the number of sites investigated and the length of data analysed. For example, Bechrakis *et al.* [123] used the ANN approach to predict the wind resource over 1 year using training lengths of 1-2 months at 3 sites in Ireland with favourable results. Similarly, Lopez *et al*. [124] investigated an ANN approach using a single target site in Spain and demonstrated that the mean wind speed could be well predicted using data collected during the

first five days of each month of the year. These studies are too limited to allow general conclusions to be drawn, however.

A more comprehensive study was initiated by Velazquez *et al.* [125] at six sites in the Canary Islands where 1 year of training data was used to predict the wind resource over a period of 8 years using one or more reference sites. Their results indicated that ANNs performed better than the linear variance ratio method, with the best predictions obtained using the ANN with two reference stations. A large-scale study [126] carried out in partnership between three industrial partners (Renewable Energy Systems, Risø National Laboratory and Ecoténia) using 41 site pairs compared the performance of ANNs with various linear MCP methods. Overall, the ANN approach tended to result in predictions with the lowest error, while simple ratio methods tended to result in the smallest bias. However, for many of the MCP approaches tested, the authors found few statistically significant differences in prediction accuracy.

Overall, ANNs applied to MCP may well offer a promising route for further study. However, at present, the degree of benefit they may confer is not clear.

3.2.4.2 Gaussian process regression

Gaussian process regression (GPR) is a powerful learning-based technique which has arisen from the field of machine learning and is increasingly being applied to a diverse number of research problems [127-129]. In the field of wind energy, GPR has mainly been applied to short-horizon (several hours) wind power forecasting to establish relationships between observed meteorological variables and wind power production. See for example the studies by Jiang *et al.* [130], Mori and Kurata [131] as well as Kou *et al*. [132]. However, it is believed that GPR has not yet been used in the context of MCP-based wind resource assessment.

The main attractions of GPR as applied to MCP are that it is capable of mapping non-linear relationships between the reference and target site wind speeds and that it employs a probabilistic approach in establishing both the form of this relationship and the uncertainty in the predictions. It is thus of interest to establish whether non-linear relationships between the reference and target sites emerge using this approach and what implications this may have for long-term wind resource predictions. A further area of investigation is whether probabilistic methods are more efficient than conventional regression methods in extracting the form of the reference/target site relationship, as well as the underlying variability, particularly when applied to very short training periods.

The formal definition of a Gaussian process is '*a collection of random variables, any finite number of which, have a joint Gaussian distribution*' [133]. Ultimately this means that GPR is an application of the properties of multivariate Gaussian distributions to regression analysis. GPR involves both a training and prediction phase. In the training phase, the relationship between one or more inputs and the corresponding outputs is established. This relationship is then generalised to make predictions in the prediction phase including an estimate of the uncertainty. In the current application the inputs and outputs are reference and target site wind speeds respectively.

Since it is a non-parametric, Bayesian approach, GPR does not constrain the relationship between the target and reference sites to be of a single functional form. Instead, the relationship is allowed to arise naturally from the training data and is expressed as a probability distribution over functions. Such an approach is less prescriptive than parametric techniques and allows for functional flexibility at different reference/target site pairs. A detailed application of GPR in an MCP context is undertaken in Chapter 6.

3.2.5 Short measurement periods

MCP approaches are predominantly studied in the context of the large-scale wind industry where small uncertainties can have very large financial consequences due to the size of the investment. In this context, a number of studies [99, 107, 134] have suggested that minimum short-term onsite monitoring periods of 8-12 months are required in order to successfully implement the MCP technique. In cases where studies have specifically considered the effect of training data length on the accuracy of the predictions, the issue is approached from a somewhat different perspective to that required for small-scale turbine installations. However, such studies can still provide valuable insights as to the potential challenges of applying MCP using short measurement periods.

An early attempt to apply MCP to training periods of less than one year was reported by Barros and Estevan [135]. Using wind data sampled from 20 stations within a large (10^3 km) region of the USA, the study attempted to predict the annual wind resource at one station through correlation to the remaining stations using principle component analysis. The study reported impressive error statistics including an average error of 5% in mean wind speed and 10% in mean wind power using a 13 week measurement period. However, the validity of the study was criticised by Skibin [96, 136] on a number of counts. Firstly, the use of multiple reference stations over a very large area implies that many of the stations

experience a different climate to the target station and hence will be poor predictors of the long-term resource at the target site. Secondly, the study was based on wind data sampled just four times per day which is insufficient to capture the full range of meteorological conditions occurring at the sites. In addition, Skibin was particularly concerned about the risk of using very short-term data in decision making for large wind farms when a wrong decision could have significant public and political repercussions for what was at the time a fledgling large-scale wind industry.

The study has important parallels with the current challenge facing the small-scale wind industry. In the present case, however, the biggest risk is not that a long-term, rigorous measurement campaign may be sacrificed in favour of a short-term one but rather that no measurements may be taken at all and that resource assessments may be neither consistent nor reliable.

More recently, Oliver and Zarling [137] investigated the impact of measurement length on prediction accuracy using linear MCP approaches applied to 14 site pairs in the USA. The study reported significant variability in the error in the long-term predicted mean wind speed depending on the season used for the onsite measurements as well as the regression approach employed. While the study highlights the importance of seasonal variability in the reference/target site correlations, the results are not conclusive. Firstly, only a single year of data was analysed and overlapping periods where used for both the regression and assessing the predictions. Hence, the results provide limited information regarding the predictive capacity of the approaches when applied to unseen data and additionally, the error statistics do not account for variability between years. Secondly, the study only considered predictions of the mean wind speed rather than the full distribution of wind speeds necessary for estimating wind power.

Hence, there is scope for a more detailed analysis of the effect of short measurement lengths using data from multiple years and considering the predicted wind speed distributions and associated parameters in addition to the mean wind speed.

3.2.5.1 Uncertainties related to measurement length

Predicting the likely uncertainties introduced by the application of MCP approaches is complicated given that in a real-world site assessment only short-term target site measurements are available. Uncertainties must thus be estimated from the short-term training period and extrapolated to the long-term prediction period. The issue is further compounded when the training period is less than 12 months since additional uncertainties are introduced related to seasonal variability.

Taylor *et al.* [134] reported results of a study where simple regression MCP methods were applied to three target sites using training data of various lengths in order to predict the mean wind speed over a 4 year period. Their results showed a strong reduction in error with increased measurement periods up to 12 months, after which, little change occurred. However, seasonal effects were not considered explicitly in the study and the limited number of target sites as well as the relatively short test period makes it unwise to generalise the uncertainty estimates. In addition, no account was taken of the errors in the predicted wind speed distribution or mean power density.

As part of their comparison of the performance of a range of MCP techniques and the development of the VR method (discussed earlier), Rogers *et al*. [99] also considered the effect of the length of the training data on MCP performance. The study was implemented using 8 site pairs and concurrent data covering between 1.7 and 5 years. Using the standard deviation of the predicted mean wind speed across different training subsets as an uncertainty metric, the study recommended training lengths of at least 9 months. These conclusions were based on the data length required for convergence of the predictions, which of course is desirable but perhaps not essential in the case of small-scale installations. In addition, the study was relatively limited in terms of the number of sites and data periods considered.

In a subsequent study, the same authors considered methods of predicting uncertainty in MCP when using variable training periods [109]. Although the application of MCP to short training periods was not the main focus of the study, the results are of relevance to this application. The study firstly demonstrated that standard statistical techniques, such as those based on the variance of the gradient and intercept of a linear fit to the data [107], underestimate the uncertainty in the predicted wind resource since they do not account for serial correlations in the wind speed observations. This situation can be improved using so-called Jacknife estimates based on estimating the variance across multiple subsets of training data. However, while such an approach is better able to represent variance within short training periods, it is still unable to account for variance between training periods (i.e. seasonal terms). Hence the authors suggest that sources of seasonal variability in the modelled relationships require further exploration.

Jung *et al.* [112] presented a Bayesian framework for quantifying the error in the estimated annual energy production at a site, including all steps in the estimation procedure. To account for the uncertainty in the MCP step, a Bayesian regression model was used, resulting in probability distributions for the fitted regression

parameters rather than fixed values. The Bayesian approach is attractive in that it allows the uncertainty in the model parameters, and hence the predicted wind resource, to be related to the length of training data. Unfortunately, however, as with the method described by Derrick [107] this approach does not account for phenomena such as seasonal trends, which according to Rogers [109], are likely to be the dominant source of uncertainty.

The development of analytical techniques for estimating uncertainty in MCP, even for relatively long training periods, remains an open challenge [18]. The problem is particularly severe in the case of training periods of less than 12 months where seasonal trends may be the dominant source of uncertainty. In the absence of reliable analytical techniques, rigorous analysis of MCP performance when applied to short training periods, including an investigation of seasonal effects, may serve as a first step to relating uncertainties to measurement length. At present, however, there is a lack of studies that have performed such analyses using a sufficient number of sites, test data lengths and terrain types from which to draw general conclusions. Such studies are vital if MCP is to be applied to the small-scale wind industry with a full understanding of the inherent uncertainties. This issue is considered in Chapter 4 where a detailed assessment of the performance of linear MCP approaches applied to very short training periods is undertaken.

3.2.5.2 Novel measurement approaches

Given the competing issues of seasonal variability in short measurement periods and the increased cost and time involved in long measurement campaigns, Lackner *et al.* [111] have proposed a solution dubbed the round robin approach (RRA). The RRA takes advantage of advances in portable measuring equipment such as LiDARs (light detection and ranging) by measuring at several sites in a single year resulting in non-consecutive data sets from each site. The approach involves measuring at one site for a short time period, 30-60 days for example, before relocating the equipment to a second site for the same period. This procedure is repeated over the course of a full year to obtain multiple measurements at 2-3 sites. The study by Lackner *et al.* indicated that the RRA was able to assess the wind resource at 2-3 sites in a 12 month period with error statistics comparable to those obtained from measuring at a single site for a full 12 month period. Presumably such an approach is successful because seasonal trends are captured sufficiently across all sites and all seasons. The general focus of the study was large-scale wind resource assessment, and specifically, efficient use of portable LiDAR

equipment. However, the results are potentially of interest to small wind developers with access to either portable meteorological masts or LiDAR.

A related, although less detailed, study was also reported by Lopez *et al.* [124] where a neural network was used to predict the annual mean wind at a single target site using data collected from the first 5 days of each month over a 12 month period. However, since both these studies only consider predictions of the mean wind speed, it is not yet clear how these approaches may impact prediction of the wind speed distribution and hence the wind power.

In the small-scale wind industry, designing novel measurement procedures that are better able to capture seasonal variability across multiple sites and seasons could potentially reduce uncertainties, even in cases where a RRA approach is applied over a period of less than 12 months. This issue is returned to briefly in Chapter 7 where random sampling of measured data across different seasons (partly analogous to RRA) is shown to significantly improve the efficiency of an MCP approach.

3.2.6 Summary of approaches

A broad, though by no means exhaustive, overview of MCP approaches has been provided. While it is clearly not possible or desirable to test all available techniques in this thesis, from the above discussion, several stand out in terms of either their proven reliability or future promise. These techniques have been selected for further study and application to small-scale wind resource assessment as described below:

(i) The approaches of linear regression and variance ratio are established techniques that are simple to implement and widely used as baseline approaches against which other techniques are compared. Hence, despite their theoretical limitations, they serve as a useful starting point for the application of MCP in the context of small-scale wind resource assessment. In addition, the conventional linear regression technique has the potential for substantial improvement through modelling the residual scatter.

(ii) The multivariate probability approaches, and specifically the approach based on the bivariate Weibull distribution, show future promise in providing a more rigorous theoretical basis for MCP. By taking account of the underlying distribution between the reference and target site wind speeds, the approach has the potential to better account for the stochastic nature of the wind flow and the resulting target site variability. Open questions remain regarding the performance of such an approach in the real-world, as well as the impact of short training periods, and this provides opportunities for further study.

(iii) The learning-based approach of Gaussian process regression is attractive in that it builds on the established linear regression techniques while introducing much greater flexibility including the use of a Bayesian framework. Such an approach could potentially offer advantages when applied to short training periods and reference/target site pairs with non-linear correlations. Since the approach has not previously been applied to MCP, there is a clear opportunity for development in this area.

The overarching aim of this thesis is the development of tools capable of predicting the wind resource rapidly and to a degree of accuracy that allows well informed investment decisions to be made. Thus, the approaches outlined above should be assessed in terms of their performance when using a range of training periods. As stated earlier, there is currently a lack of reliable, empirical data regarding the performance of MCP under the constraint of limited onsite measurements, assessed across multiple sites and training periods. Hence, a major aim of this work is to provide realistic estimates of the likely uncertainties when applying MCP methods subject to the pressures of seasonal, inter-annual and stochastic variability.

4 Evaluation of a Boundary Layer Scaling Approach to Wind Resource Assessment

4.1 Overview

As described in Chapter 3, in 2008, the UK Met Office developed a methodology based on boundary layer scaling (BS) of the mean wind to estimate the potential for small-scale wind energy in the UK [15]. In simple terms, the methodology applies a series of corrections to the NCIC reference wind climatology based on parameters which describe the average surface roughness on a local and regional scale. This results in a prediction of the spatially averaged mean wind speed at a specific location and height that, with an appropriate choice of wind speed distribution, may be used to predict the wind energy resource. Subsequently, the methodology was extended and a graphical user interface was developed to make the methodology publically available online in the Carbon Trust's Wind Yield Estimation tool (WYET) that could be used to assess the viability of proposed small turbine locations. In 2012, shortly after the completion of the work described in the present chapter, the WYET ceased to be supported by the Carbon Trust due to UK Government budget cuts. However, the underlying methodology represents a valuable tool in the assessment of the wind energy resource for small-scale installations.

The purpose of the study described in this chapter was three-fold: firstly, to make a quantitative evaluation of the performance of the methodology when estimating the wind energy resource at individual sites in a variety of terrains, since this had not previously been investigated, secondly, to investigate ways in which the implementation of the methodology may be improved while maintaining the core principles of the approach, and thirdly, to quantify the expected uncertainties resulting from errors in the model's input parameters. These aims may be summarised as follows:

(i) Performance evaluation by quantifying the errors between the predicted and measured wind speed and wind power density at 38 UK sites.

(ii) Evaluation of the effect of incorporating building morphology data. This is considered in a simple way through the use of building height data as measured by LiDAR across UK cities to establish the mean building height.

(iii) Investigation into the effect of increasing the fetch size and including directionally-dependent roughness parameters, on the accuracy of the predicted wind resource.

(iv) Investigation of the effect on power predictions of the choice of Weibull shape factor used to describe the wind speed distribution and recommendations for improving this choice.

(v) Estimation of the propagated errors in mean wind speed and wind power density due to uncertainties in the aerodynamic parameters and Weibull shape factor.

Based on the above, realistic estimates of the likely errors in the predicted wind speed and wind power density are calculated and recommendations are made regarding the implementation of the methodology to maximise the accuracy of site-specific predictions.

This Chapter is organised as follows: Section 4.2 outlines the core methodology of the BS model, highlights potential improvements and describes the model's implementation at UK sites. Section 4.3 compares the BS model predictions with the observed wind resource at 38 UK sites and investigates various methods for improving the model's implementation. Section 4.4 describes the results of uncertainty and global sensitivity analyses in order to better understand the propagation of errors and the relative contributions from the model's input parameters.

4.2 Methodology

The core methodology described in Section 3.1, including the additions incorporated in the WYET, was used as the basis for the current investigation. To investigate the approach in detail, an implementation of the methodology was developed (Model A) in the programming environment MATLAB. Since not all of the enhancements incorporated in the WYET were publically documented, differences between the WYET and the original methodology described in the Met Office report [15] were identified through consultation with one of the tool's developers [138] as well as through a series of tests.

As far as possible, Model A utilised the same principles and numerical values as the WYET, however, two specific differences should be highlighted with regard to the calculation of the regional aerodynamic parameters. Firstly, land cover data, (used as the basis for estimating z_{0eff} and d_{eff}) was obtained from a square region

centred on each site of interest, rather than simply using the closest OS grid square. Secondly, more recent estimates of UK land cover (2007 compared to 2000) were used [95].

4.2.1 Potential improvements to the WYET

There are a number of uncertainties introduced at each stage of the methodology which may combine to reduce the accuracy of the final wind resource prediction, as outlined in Section 3.1 and reference [15]. Improvements to the methodology could potentially be made by reducing these uncertainties wherever possible and quantifying the uncertainties that cannot easily be controlled. Of particular interest in this work are the uncertainties in the downscaling stages III and IV shown in Figure 3.1, (Section 3.1). The figure is repeated below for convenience (Figure 4.1) and a description of these uncertainties is provided.

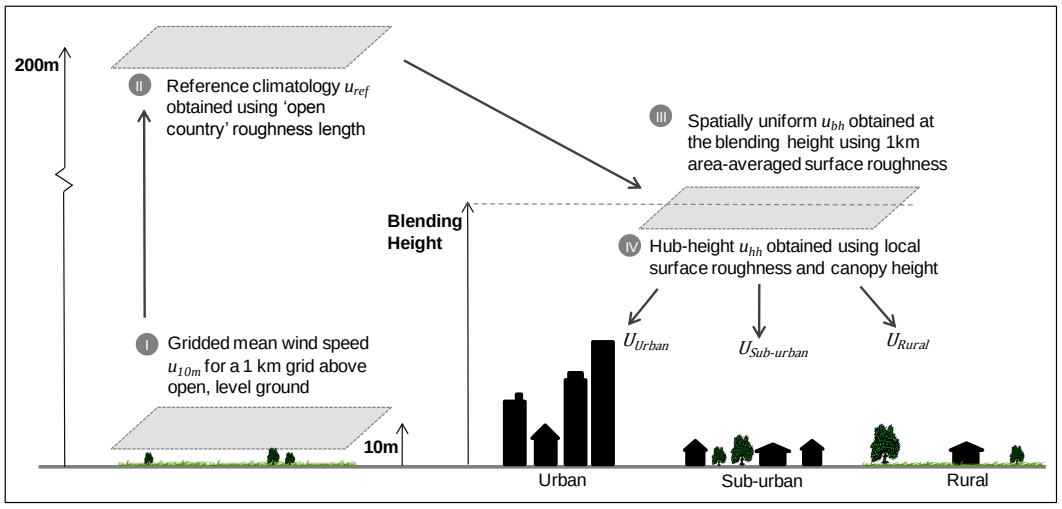

Figure 4.1: Outline of the methodology employed by the Met Office to predict the spatially averaged mean wind speed. Starting with an input mean wind speed (I) and culminating in a predicted mean wind speed at turbine hub height (IV).

Stage III: a blending method [65] is used to calculate an effective roughness length and displacement height using the fractionally weighted land cover types within the OS grid square of interest. However, if the site is at the edge of a particular OS grid square, the calculated land cover fractions may not be fully representative of the area surrounding the site. In addition, a single, 1 km grid square (fetch of ~500 m in each direction) may not be of sufficient size to accurately represent the regional aerodynamic parameters. Ideally, the fetch should be of sufficient distance to include all upwind surfaces which affect the wind profile throughout the boundary layer. Estimates of this distance vary depending on the model used and the magnitude of the roughness change [15] but central values indicate that a fetch of ~2 km should be considered assuming a boundary layer height of 200 m (Section

2.4) [139]. Additionally, in heterogeneous terrains, the roughness of the fetch is likely to be related to the wind direction and this is not accounted for in a single-valued average.

Stage IV: determining aerodynamic parameters representative of the local area can be a significant challenge. Assuming the local area is defined by a 250 m square surrounding the measurement site, the local topography can be relatively easily identified from aerial photographs. However, even with a general knowledge of this, estimating the aerodynamic parameters is non-trivial, particularly in urban environments [75]. A further factor is that the local values of z_0 and d, as well as the transition point between a logarithmic and exponential wind speed profile, are calculated as a function of the canopy height[2]. Hence, the estimated canopy height will be of particular significance in the final downscaling of the mean wind speed. Finally, the choice of Weibull shape factor used to represent the predicted wind speed distribution can have a large effect on the final wind power prediction and hence methods to optimise this choice should be investigated.

These factors are addressed in the following sections and where modifications were successful they were incorporated into a modified methodology (Model B). This approach allowed a direct comparison between predictions made using the WYET and Models A and B.

4.2.2 Estimation of the regional aerodynamic parameters

As described in Section 3.1, both regional and local aerodynamic parameters are required in order to implement the downscaling approach.

To apply the blending method (Section 2.4.4) of Mason [65] in the calculation of the regional aerodynamic parameters, land cover data are required detailing fractional coverage in the region surrounding the measurement site. These data were obtained from the Land Cover Map 2007 (LCM), compiled by the Centre for Ecology and Hydrology [95] at a resolution of 25 m and used to calculate fractional land cover for grid squares centred on each test site. The LCM uses 23 classes of land cover and these were grouped into the 8 classes shown in Table 4.1. The parameters of z_0, h_m and d associated with each land cover class reflect those implemented in the WYET methodology [15].

---

[2] The definition of the canopy height is somewhat ambiguous. In the WYET methodology it is defined as 'the average height of the tallest buildings' while in the scaling of the aerodynamic parameters it also represents the mean building height.

Parameters	Land cover							
	Broad-leaved trees	Needle-leaved trees	Grass/ Crops	Shrubs	Water	Soil	Sub-urban	Urban
z_0 (m)	0.95	1.08	0.14	0.18	$3x10^{-4}$	$3x10^{-4}$	0.70	1.60
h_m (m)	19.0	20.0	1.40	1.80	0	0	6.00	12.0
d (m)	12.7	13.3	0	0	0	0	3.10	7.00

Table 4.1: Eight land cover categories used to determine the regional values of roughness length z_{0eff} and displacement height d_{eff}. Adapted from [15].

Land cover fractions of <1% were ignored due to their potential to distort the regional estimates. The grid box average for the regional roughness length was calculated from these data using the expression introduced in Section 2.4.4 [15]:

$$\left[\ln\left(\frac{z_{bh}-d_{eff}}{z_{0eff}}\right)\right]^{-2}=\sum_i F_i\left\{\left[\ln\left(\frac{z_{bh}-d_i}{z_{0,i}}\right)\right]^{-2}\right\}$$

Equation 4.1

where z_{bh} is the blending height (the larger of 10 m or twice the maximum canopy height within the grid square), d_{eff} is the effective displacement height (the maximum of those identified within the grid square), F_i is the fraction of land cover class i within the grid square and d_i and $z_{0,i}$ are the aerodynamic parameters associated with land cover class i.

Note that z_{bh} is subject to a number of uncertainties. The value used is based on the assumption that it should be 2-5 times the height of the roughness elements [15, 66]. However, the height of the roughness elements is itself non-trivial to determine. Compounding this is the fact that, strictly, z_{bh} will also be a function of the distance from the roughness change (Section 2.4.4), although this may be impractical to implement in a simple automated approach designed to be applied to multiple sites. In addition, there is no simple way to aggregate the effect of multiple surface patches when estimating d_{eff}. Hence, defining d_{eff} as the maximum value of d_i within the grid square is a further source of uncertainty. The effect of these assumptions on the final predictions is investigated in Section 4.4.

4.2.3 Estimation of the local aerodynamic parameters

Local aerodynamic parameters, required for predictions within the built environment, are estimated using the method of Raupach [83]. Using this method,

z_0 and d are calculated as a function of the mean height h_m, and frontal area density λ_f, of the roughness elements using the expressions [15]:

$$\frac{d}{h_m} = 1 - \frac{1 - \exp\left(-\sqrt{15\lambda_f}\right)}{\sqrt{15\lambda_f}}$$

Equation 4.2

$$\frac{z_0}{h_m} = \left(1 - \frac{d}{h_m}\right)\exp\left(-\frac{\kappa}{\min(\sqrt{0.003 + 0.3\lambda_f}, 0.3)} + 0.193\right)$$

Equation 4.3

where $\kappa = 0.41$ is the von Karman constant.

In practice, values of h_m and λ_f are estimated by first categorising the test site in terms of the local site character, here defined as a 250 x 250 m square centred on the site. In this study, the six site categories implemented in the WYET are used. Each site class is linked to values of h_m and λ_f based on the values recommended by Grimmond and Oke [75] as shown in Table 4.2. As mentioned previously, while these estimates represent what could be considered typical values, they are subject to considerable uncertainty [75]. No sites in the current study were identified as belonging to categories 2 or 6, although they are included in Table 4.2 for completeness.

Category	**Description**	h_m**(m)**	λ_f	z_0/h_m	d/h_m
1. Open countryside	Rural areas with little urban infringement.	-	-	Fixed z_0 = 0.14	Fixed d = 0
2. Woodland	Mature trees.	19.5	0.53	0.05	0.67
3. Urban low height and density	Small buildings and tress that are not closely spaced.	6	0.15	0.10	0.48
4. Urban medium height and density	Mixed height buildings of 2-4 stories and trees.	9	0.2	0.12	0.52
5. Urban high height and density	Closely spaced buildings of 4-6 stories and trees.	12	0.3	0.13	0.59
6. Urban very high height and density	Tall towers of different heights in dense urban areas.	25	0.3	0.13	0.59

Table 4.2: Local site categorisation and the associated mean height of the roughness elements h_m, and frontal area density λ_f. The normalised aerodynamic parameters are calculated using the method of Raupach.

4.2.4 The NOABL-MCS method

While the model described above is referred to as a simple boundary layer scaling approach it is worth noting that it is considerably more sophisticated than the current recommended best practice for small wind turbines. The most recent UK microgeneration installation standard [13] sets out requirements that must be met in order for installed small wind turbines to receive financial support via the feed-in-tariff. The requirements for wind resource assessment involve a simple scaling of the NOABL 10 m wind speed using the following procedure:

(i) The site is assigned to one of five possible terrain categories ranging from flat grassland to dense urban.

(ii) Significant obstacles are identified within two zones upwind and downwind of the turbine based on the prevailing wind direction.

(iii) The height of the highest obstacle is used as a correction factor to the height of the turbine using $hc = ht - 0.8ho$, where hc is the 'corrected' turbine height, ht is the actual turbine height above ground level and ho is the height of the highest obstacle within the zones.

(iv) The mean wind speed prediction is then obtained from a table of corrected NOABL 10 m wind speeds for heights between 1 and 100 m where the height is defined as hc.

Note that step (iii) is roughly equivalent to defining a displacement height while step (iv) applies an approximately logarithmic wind speed profile based on the local terrain category. While this procedure provides a starting point to correcting the NOABL 10 m wind speeds it is somewhat simplistic.

For example, while defining a displacement height based on the height of the largest significant obstacle may be reasonable for sites located in relatively uniform urban areas, for sites in more open terrain with isolated obstacles, this approach is likely to lead to large underestimates of hc. This will in turn lead to large errors in the scaling factor chosen for the NOABL wind speed, which is in itself also a broad approximation. In addition, only obstacles in the prevailing wind direction are accounted for and no representation of the regional terrain is included. To highlight these issues, the NOABL-MCS method was applied to the sites considered in the current study as a comparison with the BS model.

4.2.5 Meteorological measurements

In order to compare the predicted wind resource with wind measurements at a variety of locations, recordings of mean wind speeds were collated from 38 sites

throughout the UK in a variety of terrains. Of these sites, 36 were at independent geographical locations, the remaining two were at the same location with different anemometer heights. Since the current thesis is concerned with predicting the wind resource available to small-scale wind turbines, it is important that the wind data used for validation and testing is obtained at heights similar to the hub heights of small turbines. From Table 2.1, (Section 2.1.2), typical hub heights are in the range 10-35 m. At these relatively low heights, the local surface and small-scale features may have a significant effect on the wind flow. Hence, with the exception of one site mounted above a tall building in an urban area, the data used for this study were obtained at heights below 35 m. This ensured that the resource assessment approaches could be evaluated in realistic scenarios representative of where small turbines are likely to be installed.

The majority of these sites (23) were selected from the Met Office anemometer network which archives long-term surface wind speed measurements in the Met Office Data Archive System (MIDAS) [140]. Data were extracted using the online CEDA Web Processing Service hosted by the British Atmospheric Data Centre [141]. These data consisted of values of mean wind speed averaged over a complete hour with resolution 1 knot = 0.51 ms^{-1}. The data had undergone a number of quality control procedures during the archiving process [142]. Over the years, several wind measuring devices have been used in the Met Office network, with the cup anemometer and wind vane the most common. Depending on the installed equipment, the anemometer can have poor response characteristics at low wind speeds, although since 1998, improved anemometers have been installed at most sites [142]. Historical changes to instrumentation as well as changes in measurement height and local land cover have in some cases introduced biases in the long-term MIDAS data record [98]. Although some meta data have been recorded in the MIDAS archive, this is generally not of sufficient detail to fully identify all such changes [142] further complicating the identification and correction of such biases. However, given the relatively short data period (5 years) used in the current study, these uncertainties are unlikely to significantly affect the overall conclusions.

After extracting the MIDAS data, further checks were performed to remove duplicate entries and any data identified as suspect by the Met Office quality control procedures. A final quality check was carried out to ensure no extreme wind values remained. To reduce the effect of inter-annual variation, wind speed statistics covering a five year period from August 2006 - July 2011 were selected. While variations in the mean wind speed will also occur on longer timescales, data from

other sources used in the study covered shorter periods and hence a balance was sought between the different data lengths available from all sources.

An additional 10 sites were selected from the Energy Saving Trust (EST) field trial of domestic-scale wind turbines [11]. These data consisted of five minute mean wind speeds (resolution 0.1 ms^{-1}) recorded by ultrasonic anemometers at the location of small wind turbine installations during the one year period April 2008 to March 2009. The data were obtained directly from a database administered by the UK trade association RenewableUK. While specific details regarding quality control procedures are not available for these data, they have been the subject of several previous peer reviewed studies [21, 22]. Given the relatively short data collection period, it was felt that a relatively strict 95% data coverage criterion should be applied to avoid seasonal biases. While 57 sites were involved in the original field trial, only the 10 used in the current study achieved this criterion.

The remaining 5 test locations were research installations. Four of these were located in the city of Leeds and consisted of ultrasonic anemometers mounted at two different heights at two geographical locations, one administered by the University of Leeds and a second by Leeds City Council. The Leeds data consisted of 15 minute mean wind speeds (resolution 0.01 ms^{-1}) recorded over the period July 2008 - May 2010. Since the four Leeds sites were located within approximately 2.5 km, consistency checks between the data recorded at each site were used for quality control purposes. The final site was a research installation in the city of Manchester administered by the University of Manchester Centre for Atmospheric Sciences. Data were obtained directly from the University of Manchester as 15 minute mean wind speeds (ultrasonic anemometer, resolution 0.03 ms^{-1}) for the period January 2008 to December 2008. A real-time, automated quality control procedure, implemented by means of the instrument firmware, was used for this site to prevent data spikes due to water or icing.

To avoid seasonal biases, data from all sites covered an integer number of years. While a 95% data coverage criterion was applied to the EST sites, this criterion was relaxed slightly for sites with more than one year of data. Of the sites with more than one year of data, all surpassed the 95% data coverage criterion except three, namely, SU1 (90%), U5 (85%) and U6 (85%). These three sites were deemed suitable for inclusion since the data period covered between two and five years. To ensure consistency in derived parameters related to the wind speed distributions and wind power, all data were converted to hourly means before further analysis. Since the non-MIDAS sites included data of differing lengths, a correction factor (<

5%) was applied to the mean wind speeds based on the average inter-annual variation at the remaining 23 MIDAS sites as shown in Figure 4.2. The resulting corrected mean wind speeds, $\bar{u}_{obs}$, are thus representative of the five-year mean wind speed over the period August 2006 to July 2011.

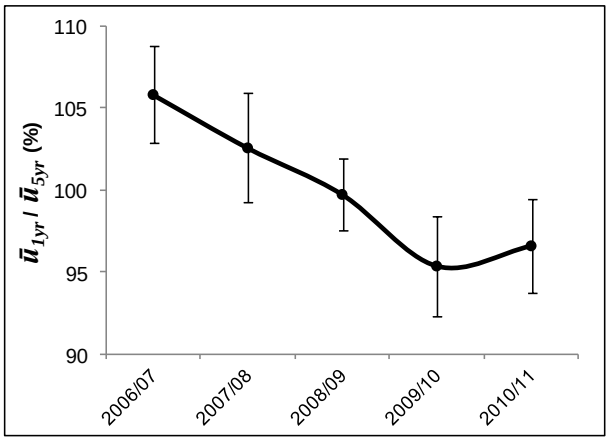

Figure 4.2: Annual mean wind speeds as a percentage of the five-year mean calculated at 23 UK sites. The annual periods run from August to July. Error bars represent +/- one standard deviation across the 23 sites, the solid line is a guide to the eye.

Note that the boundary layer scaling methodology utilizes reference climate data from the NCIC database as outlined in Chapter 23. These data are based on UK land surface observations over the 30 year period 1971 – 2000. Hence, due to long-term variability, the average wind climate during the NCIC data period may differ from that during the five year period used in the present study. To determine the size of this effect, UK annual wind indices for the period 1971 – 2000 were compared with the period 2006 – 2011 using the BADC-7, 55 year wind indices reported by Watson *et al.* [98]. The average wind indices were found to differ by only 1% for the two periods. Given the very small discrepancy between the two indices and the fact that these indices also exhibit geographical variability [98], it was felt that applying a correction factor would not be appropriate in the current study.

4.2.6 Terrain classification

To investigate the effect of terrain on prediction accuracy, sites were specifically chosen to represent four general categories defined as: *urban* (9 sites), *sub-urban* (8 sites), *coastal* (11 sites) and *rural* (10 sites). Classification of these sites was made from a subjective analysis of satellite images of an area covering several kilometres centred at the test site. In general, the terrain classification will match the local site category (defined in Table 4.2). However, given that the local site

categories are defined on a local area of just 250 x 250 m, in some cases the terrain class will differ from the local site category. For example, a site located in a large park within a sub-urban area would have a terrain classification of sub-urban but may have a local site category of open countryside. A further factor of note is that the coastal classification is not exclusive since coastal sites will also belong to one of the three remaining classes, (urban, sub-urban or rural). Details of all sites are given in Table 4.3 and their approximate geographical locations are shown in Figure 4.3.

Site	Cat.	OS grid	H (m)	$\overline{u}_{obs}$ (ms^{-1})	Data (yrs)	Site	Cat.	OS grid	H (m)	$\overline{u}_{obs}$ (ms^{-1})	Data (yrs)
U1* (M)	4	SJ8396	20.6	3.2	5	**SU1 (M)**	1	NJ8712	10.0	4.5	5
U2* (M)	4	SU4210	22.5	4.4	5	**SU2 (M)**	3	SK5045	10.0	3.5	5
U3* (E)	4	TQ4676	9.0	2.4	1	**SU3 (M)**	1	SU8554	10.0	3.7	5
U4* (E)	4	NT2575	12.0	2.3	1	**SU4 (M)**	3	SU1344	10.0	3.6	5
U5* (R)	5	SE2934	21.0	3.2	2	**SU5 (M)**	3	SU1740	10.0	4.6	5
U6* (R)	5	SE2934	24.4	3.8	2	**SU6 (M)**	3	SD8812	10.0	2.1	4
U7 (R)	4	SE3032	32.0	4.0	2	**SU7 (M)**	1	SP3180	10.0	3.1	5
U8 (R)	4	SE3032	12.0	2.9	2	**SU8* (E)**	3	SU9505	11.3	2.1	1
U9* (R)	5	SJ8497	43.3	3.4	1	**-**		**-**	-	-	**-**
C1 (M)	3	NK1345	10.0	5.4	5	**R1 (M)**	3	NH8914	10.0	2.6	5
C2 (M)	3	NU2514	10.0	5.1	5	**R2 (M)**	1	SE5238	10.0	4.3	5
C3 (M)	3	TA1967	10.0	4.9	5	**R3 (M)**	1	SK5026	10.0	3.5	5
C4 (M)	3	NM8834	10.0	3.9	5	**R4 (M)**	1	SO9749	10.0	3.5	5
C5 (M)	1	SN2452	10.0	6.6	5	**R5 (M)**	1	SU7349	10.0	4.6	5
C6 (M)	1	SX9456	10.0	6.1	5	**R6 (M)**	1	NS8264	10.0	5.9	5
C7 (M)	1	SZ2984	10.0	8.0	5	**R7 (E)**	1	SD7517	7.0	4.3	1
C8 (M)	3	SD3000	10.0	5.8	5	**R8 (E)**	1	SD8823	7.0	4.6	1
C9* (E)	3	TM2320	7.7	2.9	1	**R9* (E)**	3	SY9795	9.8	2.4	1
C10* (E)	3	SZ6899	5.0	3.6	1	**R10 (E)**	1	SS3315	8.0	4.9	1
C11* (E)	3	SX0653	7.7	1.8	1	**-**		-	-	-	**-**

Table 4.3: Summary of sites used to compare predicted and measured mean wind speeds. Sites are defined as Urban, Sub-Urban, Coastal or Rural. The letters in parenthesis indicate the data source, MIDAS (M), Energy Saving Trust (E) or research site (R). The site category (Cat.) refers to Table 4.2 and *H* is the anemometer height above ground level. *Building mounted mast.

Anemometers mounted above buildings are denoted by stars in Table 4.3. Note that the majority of urban sites are building mounted as well as a small number from the remaining categories. It is well known that for building mounted masts, the presence of the building can significantly affect the wind flow, indeed this in itself is an area of intense research [14, 17, 143]. Since the current study is concerned with evaluating methods of predicting the spatially averaged wind flow, using simple parameterisations of the local surface, detailed information regarding mounting position has not been considered. This is in line with the Met Office methodology

which the current study is designed to evaluate. However, building-scale effects can be considered as an additional source of uncertainly, particularly in urban areas.

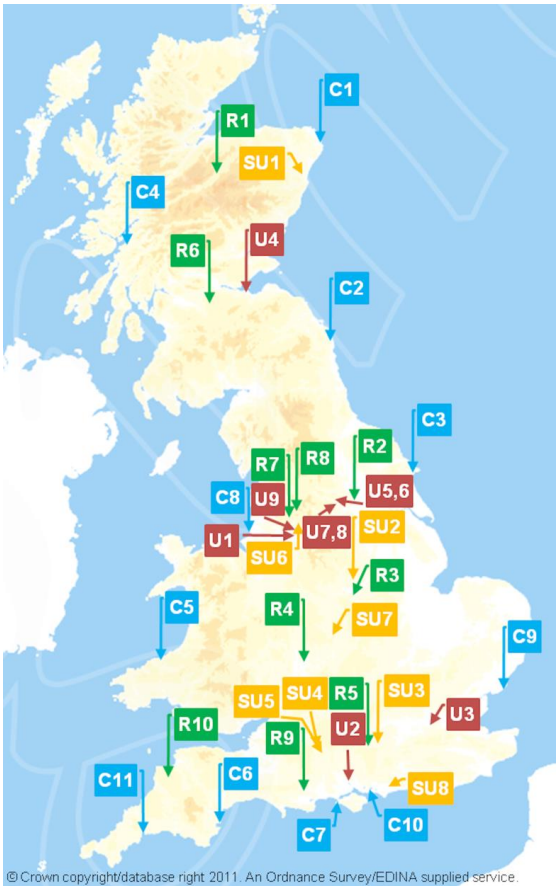

Figure 4.3: Approximate geographical locations of the sites used to compare predicted and measured mean wind speeds. Sites are defined as Urban, Sub-Urban, Coastal or Rural.

4.2.7 Error metrics

In making comparisons between prediction methodologies and terrain types, three main error metrics have been used:

The mean absolute error (MAE) defined as:

$$\mathrm{MAE} = \sum_{i}^{n} \left| \bar{u}_{obs,i} - \bar{u}_{pred,i} \right| / n$$

Equation 4.4

where $\bar{u}_{pred}$ is the predicted mean wind speed, i refers to the i^{th} site and n is the total number of sites.

The root mean square error (RMSE) given by:

$$\mathrm{RMSE} = \sqrt{\sum_{i}^{n} \left|\bar{u}_{obs,i} - \bar{u}_{pred,i}\right|^2 / n}$$

Equation 4.5

The average absolute percentage error (%Error) defined as:

$$\%\mathrm{Error} = 100 * \sum_{i}^{n} \frac{\left|\bar{u}_{obs,i} - \bar{u}_{pred,i}\right|}{\bar{u}_{obs,i}} / n$$

Equation 4.6

A fourth error metric, the mean bias error (MBE) was also used when analysing the overall predictions of mean power density:

$$\mathrm{MBE} = \sum_{i}^{n} (\bar{p}_{d,pred,i} - \bar{p}_{d,obs,i}) / n$$

Equation 4.7

where $\bar{p}_d$ is the mean Betz power density (Equation 2.9), i refers to the i^{th} site and n is the total number of sites.

Since the different metrics provide different sensitivities, it is useful to use a combination of these when making comparisons. For example, the MAE may be thought of as an estimate of the absolute error in $\bar{u}_{pred}$, however, its utility will depend on the number of sites used in its calculation, as well as the relative magnitude of the wind speed at the sites. The RMSE metric is similar to the MAE but more sensitive to large outliers. The percentage error metric is normalised by $\bar{u}_{obs}$ at each site. Hence there is an implicit assumption that the error scales with $\bar{u}_{obs}$. If such scaling is not present, this metric may not be appropriate for sites with particularly high or low $\bar{u}_{obs}$.

4.3 Results and Discussion I – performance and recommended improvements

The predicted and observed mean wind speeds at all 38 sites are compared in Figure 4.4 for predictions obtained using Model A. Similar predictions were obtained directly from the WYET, although differences were observed at coastal sites likely due to the differences in the estimation of the regional aerodynamic parameters from land cover data. As a benchmark, Figure 4.4 also shows wind speeds from the

NCIC database, used as input for the boundary layer scaling model. Note that the NCIC wind speeds represent $\bar{u}$ at a height of 10 m over open terrain. In cases where the anemometer height differed from 10 m, the NCIC values were scaled to the correct height using a logarithmic wind profile (Equation 2.19) with no correction for the local site characteristics (i.e. using aerodynamic parameters representative of open country z_0= 0.14 m and d = 0 m). In addition, predictions obtained using the simple NOABL-MCS method as set out in the installer standard [13] are also shown.

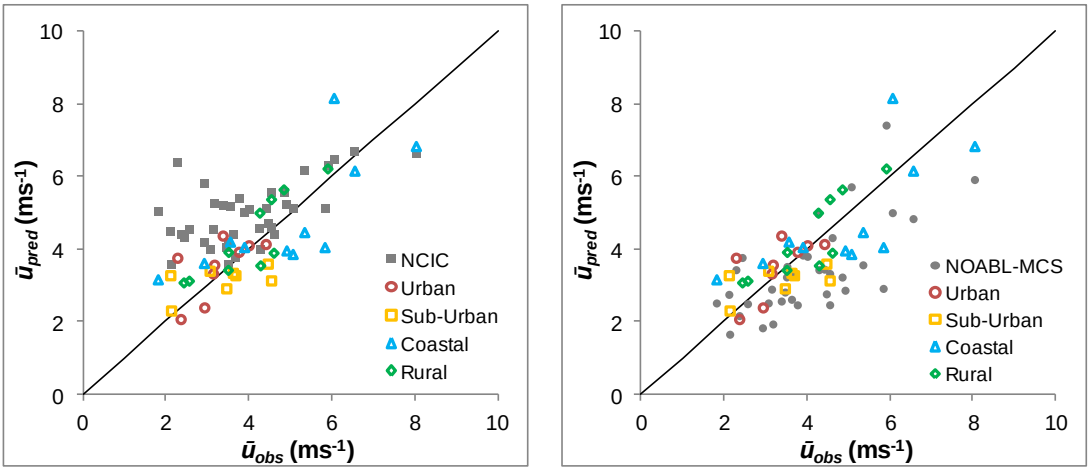

Figure 4.4: Predicted versus observed mean wind speeds for all sites using model A (coloured symbols). Benchmark predictions obtained directly from the NCIC database (left) and NOABL-MCS method (right) are shown in grey. The solid line represents a one-to-one relationship.

The Model A predictions can be seen to reproduce the general wind speed trends across the sites, although the scatter indicates relatively large site-specific errors. Unsurprisingly, the uncorrected NCIC wind speeds serve as poor predictors of $\bar{u}_{obs}$, particularly for low wind speed sites, since they do not take account of the local surface roughness. In the majority of cases this results in significant overestimates of $\bar{u}_{obs}$. The NOABL-MCS predictions show some improvement over the uncorrected NCIC estimates, however, the method frequently results in substantial underestimates of $\bar{u}_{obs}$, (29 underestimates compared to 7 overestimates). In addition, no prediction was possible for two sites since the corrected height, *hc,* was less than zero. Note that in the work by James *et al.* [21], which reported overestimates when using the NOABL-MCS method, an earlier version [144] of the NOABL-MCS method was employed. Overall, Figure 4.4 demonstrates that despite site-specific errors, a simple scaling approach (Model A) can add significant value by taking account of the local surface characteristics.

Table 4.4 compares the error metrics for the WYET and Model A using both the NCIC and NOABL wind speed databases as input. While the NCIC database is purported to be more accurate (as noted in Chapter 3), it is not publically available, and hence in practice, it may be necessary for users of the methodology to resort to the public NOABL database. Across all metrics, the largest errors are observed at coastal sites. This is not surprising since the large change in roughness at the land-sea boundary is not well represented by single-valued, directionally-independent aerodynamic parameters. As discussed in Section 2.4.5, the wind speeds at coastal sites will be particularly sensitive to the presence of sea or land in the fetch, and will be further affected by complex stability conditions, orography, and the height of the IBL. In addition, the majority of these sites also belong to the sub-urban category and hence the issues with identifying appropriate local aerodynamic parameters are compounded. In general, the smallest errors are observed at rural sites. These sites tend be well exposed and less subject to abrupt changes in roughness or local flow perturbations due to buildings and hence they conform more closely to the idealised model in Figure 4.1. On average, the absolute error in the predicted mean wind speed ranges from ~0.5 - 1.0 ms^{-1}.

	Methodology	Urban	Suburban	Coastal	Rural	All
MAE (ms^{-1})	WYET	0.54	0.60	0.82	0.51	**0.63**
	Model A (NCIC)	0.48	0.65	1.03	0.57	**0.70**
	Model A (NOABL)	0.59	0.72	0.92	1.06	**0.84**
RMSE (ms^{-1})	WYET	0.70	0.74	1.07	0.62	**0.81**
	Model A (NCIC)	0.65	0.78	1.17	0.62	**0.85**
	Model A (NOABL)	0.75	0.86	1.07	1.28	**1.02**
%Error	WYET	19.0	19.1	21.5	14.0	**18.4**
	Model A (NCIC)	17.1	19.9	24.1	15.0	**19.1**
	Model A (NOABL)	19.5	24.7	20.0	27.0	**22.7**

Table 4.4: Wind speed error metrics compared over four terrain types using the WYET and Model A.

Further insight can be gained by considering the residual errors in the predicted mean wind speed, $\varepsilon_{\bar{u}}$, defined by:

$$\varepsilon_{\bar{u}} = \bar{u}_{obs} - \bar{u}_{pred}$$

Equation 4.8

The distribution of residuals is shown in Figure 4.5 for the WYET and Model A. The results indicate that the predictions made using the WYET have a small negative bias (tendency to underestimate), while the opposite is true for Model A (NCIC). Predictions made using Model A (NOABL) exhibit a stronger positive bias, this is

likely due to a positive bias in the NOABL reference climatology which has been noted previously [15]. Due to the better performance of NCIC compared to NOABL, in the subsequent analysis predictions are made using solely the NCIC database.

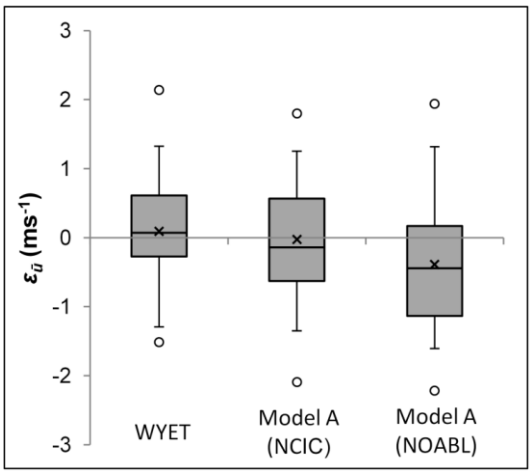

Figure 4.5: Distribution of residual errors in the mean wind speed for different implementations of the prediction methodology. The shaded regions represent the interquartile range, the horizontal lines and crosses represent the median and mean respectively, circles represent the maximum and minimum and the error bars are the 5th and 95th percentiles.

4.3.1 Sensitivity to canopy height

For urban and sub-urban sites, in addition to being used to scale the local aerodynamic parameters (Figure 4.5, Stage IV), the canopy height is also the lower bound of the log law, (Equation 2.20). For a surface with heterogeneous building heights (most real urban and sub-urban surfaces), this lower bound should ideally be defined by an 'effective mean building height'. This is a height scale that accounts for the disproportionate effect of isolated tall buildings on the drag, and hence, may be higher than the mathematical mean of the building heights [85]. In simple terms this effective height can be considered as the mean height of a uniform array that would give rise to the same aerodynamic parameters as the heterogeneous array. In the current methodology this distinction is not made and the canopy height is used for both the scaling of the local aerodynamic parameters and as the lower limit for the log law. Even without considering the complexity of height heterogeneity however, an improved estimate of the canopy height could potentially increase the accuracy of the scaled aerodynamic parameters. In addition, since at heights below the canopy top an exponential profile is matched to the log law and extended downwards, the estimated canopy height has further implications for $\bar{u}_{pred}$ close to the canopy top.

The first approximations to the canopy height are the default values used in the WYET (Table 4.2) ranging from 6 – 25 m for urban environments. These values are linked to descriptions of typical UK cities which are chosen based on the user's impression of the local site character. In an attempt to improve on these subjective estimations, local building heights, as measured by LiDAR with a horizontal resolution of 0.5 m, (rasterised to 1 m), were obtained using the Landmap Spatial Discovery database [145] for sites categorised as urban. These data, which are available for a number of UK cities, allow the local mean building height (h_m) to be estimated without the need for detailed ground based measurements.

Figure 4.6 shows $\bar{u}_{pred}$ versus $\bar{u}_{obs}$ for urban sites using the default canopy height, as well as the deviations introduced by setting the canopy height equal to h_m calculated within a 250 m square centred on each measurement site. In all cases, h_m was found to be greater than the default canopy height based on the descriptions given in Table 4.2, resulting in a reduction in $\bar{u}_{pred}$. In all but two cases, the error in the predicted mean wind speed is either worsened or not significantly affected when the canopy height is set to h_m.

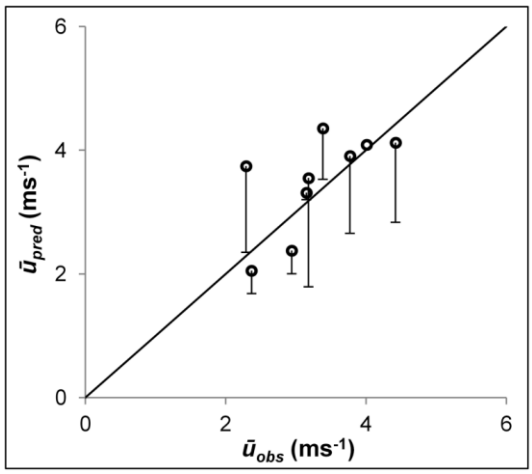

Figure 4.6: Predicted and observed wind speeds in the urban category. The data points represent $\bar{u}_{pred}$ using the default canopy height. The bars represent the deviation in $\bar{u}_{pred}$ when the canopy height is set to h_m as calculated from LiDAR. The solid line represents a one-to-one relationship.

There are several possible reasons for the increased error in $\bar{u}_{pred}$ when the canopy height is set to h_m. Firstly, it was observed in this work that on a local scale there are discrepancies between the building heights as measured using ground based observations and the LiDAR measurements within the Landmap database. In particular, for buildings with non-uniform roof heights, the database returns the height of the highest part of the roof, leading to overestimates of the average height [86]. Since the anemometer heights were measured using ground based

observations, the estimate of the anemometer height in relation to the canopy top will also contain discrepancies. Hence the scaling of z_0 and d, as well as the height at which the exponential profile (Equation 3.5) is matched to the log law, will be affected.

Secondly, there is some indication that the scaled values of z_0 used in the current methodology, which are calculated as a function of the frontal area density of the roughness elements using the method of Raupach [83], may be higher than those typically found in UK cities [86]. Since for all the sites shown in Figure 4.6, h_m was higher than the default canopy height, this will magnify any overestimation in z_0 and hence reduce $\bar{u}_{pred}$.

These issues may be improved by developing new methods for relating roughness parameters to building morphologies, as well as by extracting more accurate building heights from LiDAR data. In particular, building heterogeneity should be taken into account in the calculation of local aerodynamic parameters and in the definition of the effective mean building height [75, 86]. Recent studies [87, 146] have investigated more sophisticated approaches that incorporate detailed building height and footprint data in the calculation of not only the mean building height but also the local aerodynamic parameters. While these approaches require more detailed model inputs, results from these studies have indicated that they can lead to significant improvements in predictions of the mean wind speed in urban areas.

4.3.2 Sensitivity to wind direction and size of fetch

In both the WYET and Model A, a relatively small fetch (1 km square) is used in the first downscaling stage, (Figure 4.1, stage III). In addition, no account is taken of the directional dependence of the upwind roughness.

In order to investigate the sensitivity to these factors, the regional grid was increased to 4 km x 4 km squares, centred on each measurement site. Each grid square was further subdivided into a sub-grid of four, representing the upwind fetch from the 90^0 angular sectors of North-East, South-East, South-West and North-West. The LCM 2007 was then interrogated for each sub-grid square to obtain regional aerodynamic parameters as described previously. This process allowed regional aerodynamic parameters to be calculated for each of the four wind direction sectors and these were used to calculate a directionally-dependent wind speed at the blending height based on the roughness of the upwind fetch. Note that the regional parameters are still grid box averages but in this case there are four estimates, one for each grid square, rather than a single estimate. A final

downscaling (Figure 4.1, stage IV) was applied, based on the local site character, to calculate the mean wind speeds at hub height. The directionally-dependent mean wind speeds were then combined into a single value using a weighted sum, given by:

$$\bar{u}_{pred} = \sum_{k=1}^{4} f_k . \bar{u}_{pred,k}$$

Equation 4.9

where f_k is the probability of wind directions from the k^{th} angular sector, obtained from the onsite wind rose and $\bar{u}_{pred,k}$ is the directionally-dependent predicted mean wind speed at hub height.

This methodology is referred to as Model B. By way of example, Figure 4.7 compares the implementation of the WYET/Model A and Model B for a single site C8, with an observed mean wind speed $\bar{u}_{obs}$= 5.8 ms^{-1}. In both cases the same local aerodynamic parameters are used since the increased fetch only impacts the regional parameters of z_{0eff} and d_{eff}. It can be seen that Model A poorly represents the fetch resulting in a large error in $\bar{u}_{pred}$ (31% error) due to unrepresentative regional estimates of z_{0eff} and d_{eff}. In contrast, Model B is able to account for a larger fetch as well as the higher frequency of south-westerly winds which experience reduced drag due to the presence of the sea. This results in a higher estimate of $\bar{u}_{pred}$ that is much closer to the observed value (9% error).

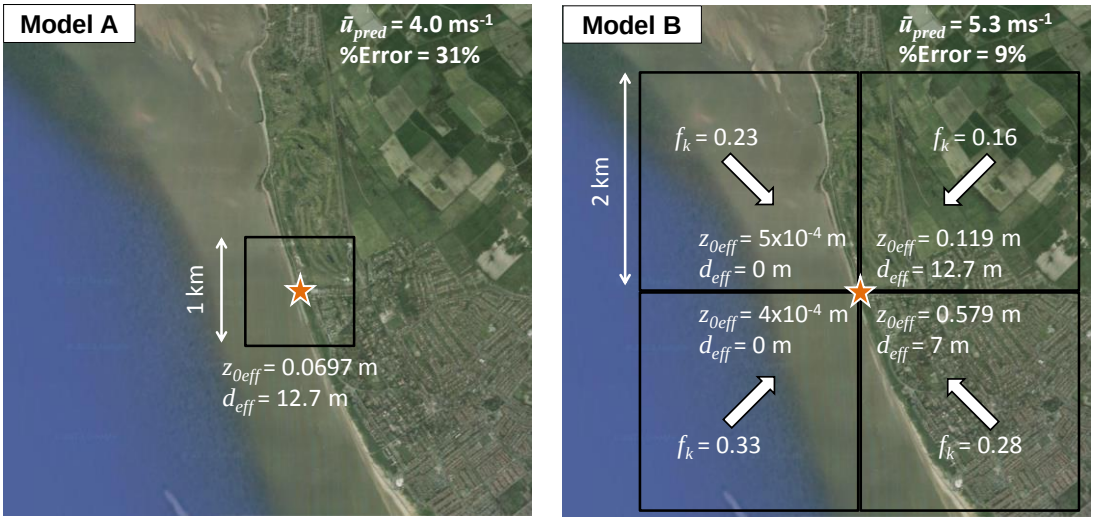

Figure 4.7: Comparison between methodologies for estimating the regional aerodynamic parameters and mean wind speed at test site C8. Left: the standard approach as implemented in WYET/Model A, Right: The enhanced approach of Model B incorporating a larger fetch and weighted directionally-dependent regional aerodynamic parameters. The star marks the anemometer location. © Google maps.

Figure 4.8 compares the distribution of residuals over all sites grouped into terrain type using Models A and B. For all terrains, the inter-quartile range and the bias are reduced when using the larger fetch and directionally-dependent aerodynamic parameters. The improvement is largest for coastal sites where the inter-quartile range is reduced by 30% and the large negative bias becomes a small positive bias. Intuitively one would expect that a larger fetch is most important in areas where the land cover is changing rapidly with distance. This is particularly true of coastal sites where the presence of the sea may cause an abrupt change in roughness, leading to a strong dependence of the degree of roughness on both distance and the direction of the prevailing wind.

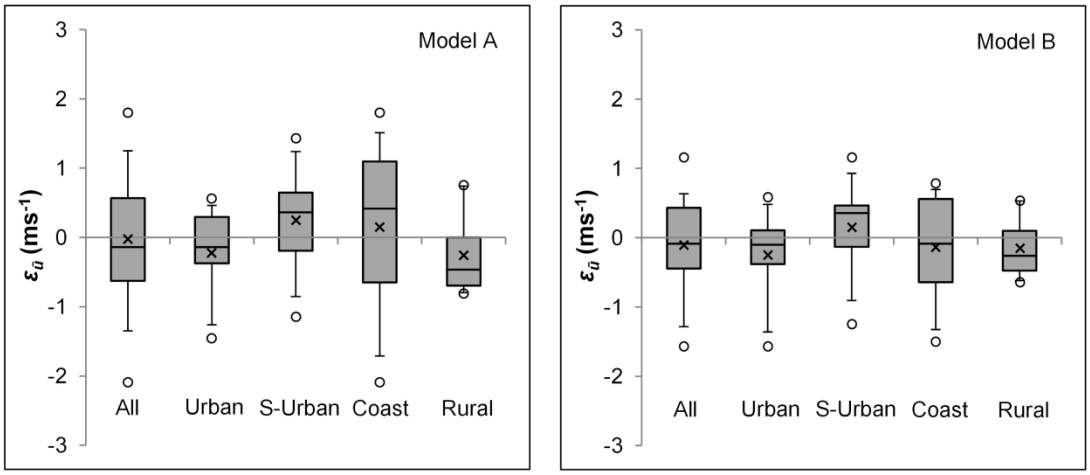

Figure 4.8: Distribution of residual errors in the mean wind speed for different terrain types. Left: Model A, fetch defined by a 1 km square, Right: Model B, fetch defined by a 4 km square with directionally-dependent regional aerodynamic parameters. The horizontal lines and crosses represent the median and mean respectively, the circles represent the maximum and minimum values and the error bars are the 5th and 95th percentiles.

Table 4.5 summarises the error metrics obtained using Models A and B. An improvement can be seen when using Model B for all error metrics and terrain types except *urban*, where there is no significant change. This is likely because the LCM description of the regional land-cover for urban sites is simply either 'urban' or 'sub-urban'. Hence, using this rather simplistic descriptor of roughness, there are unlikely to be large roughness changes with wind direction or fetch size. Instead, the urban predictions are expected to be more sensitive to the descriptors of local roughness (Figure 4.1, Stage IV) which may be improved by more sophisticated models that take into account building morphology [85].

	Methodology	Urban	Suburban	Coastal	Rural	All
MAE (ms^{-1})	Model A	0.48	0.65	1.03	0.57	**0.70**
	Model B	0.47	0.55	0.64	0.39	**0.52**
RMSE (ms^{-1})	Model A	0.65	0.78	1.17	0.62	**0.85**
	Model B	0.68	0.68	0.76	0.44	**0.65**
%Error	Model A	17.1	19.9	24.1	14.9	**19.1**
	Model B	17.2	18.0	18.9	10.9	**16.2**

Table 4.5: Wind speed error metrics compared over four terrain types using Models A and B.

In the implementation of Model B, a regional grid square of size 4 km was chosen since this implies a fetch of ~2 km in each direction. Given that the regional wind speed is downscaled from the top of an IBL assumed to be of height 200 m, this fetch distance corresponds to the approximate 1/10 IBL height-to-fetch ratio discussed in Section 2.4.3, and is close to central values proposed by other authors

[15, 139]. However, it is informative to examine the impact of this choice of fetch distance on the accuracy of predictions made using Model B. To investigate this, Model B was applied using regional grid squares in the range of 1 - 5 km and the accuracy of the predicted mean wind speeds were assessed using the error metrics described previously. Note that even for a grid square of 1 km, the error metrics may differ from those obtained using Model A, since Model B also subdivides the region into a grid of four to account for wind direction. Figure 4.9 shows the resulting error metrics averaged across all measurement sites as a function of the size of the regional grid square.

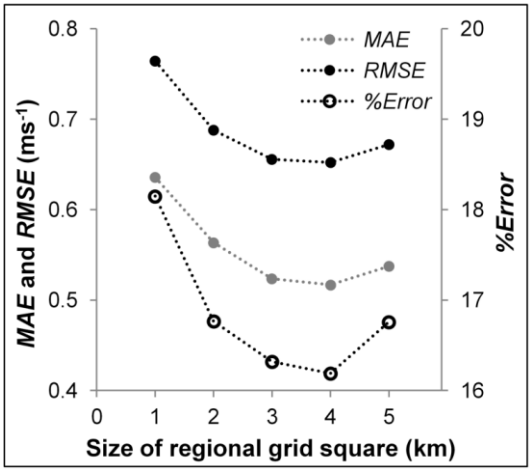

Figure 4.9: Wind speed error metrics averaged across all sites as a function of the size of the regional grid square. Dotted lines are included as a guide to the eye.

Across the three error metrics, the lowest values are achieved for a regional grid square of size ~4 km indicating that on average, a fetch of ~2 km results in the most accurate predictions when using Model B, in agreement with previous studies [15, 139]. The apparent increase in the error metrics for grid squares larger than 4 km is to be expected using a simple grid-box averaging approach when calculating regional aerodynamic parameters. Using this method, the inclusion of land-cover fractions at large distances from the measurement site will distort the calculation of the aerodynamic parameters since no distance weighting is applied. Hence, while these results provide empirical support for the choice of a regional grid square size of 4 km, care should be taken when extending this result to cases where the reference height differs from 200 m or where the aerodynamic parameters are calculated using more complex distance weighted approaches rather than grid box averages [67].

Finally, it should be noted that if Model B were to be implemented in practice, data regarding the onsite wind rose, as required in Equation 4.9, would not be available.

However, this could be replaced by data from a nearby reference site or even generic wind roses representative of regional UK climates.

4.3.3 Sensitivity to the Weibull shape factor

While an accurate estimate of $\bar{u}$ is the starting point in predicting the wind resource, in the absence of onsite wind speed measurements, some assumption must also be made regarding the distribution of wind speeds. Typically, the two-parameter Weibull distribution is used for this purpose since it has been shown to be representative of a large number of wind regimes [41]. The distribution is defined by a mean wind speed and a shape factor k which describes the spread of wind speeds about the mean. Due to the non-linear relationship between wind speed and wind power, the value of k will impact on the predicted power in the wind. In the case of the WYET, a default value of k = 1.8 is used with an option for the user to define an alternative value. To investigate the effect of the shape factor on the predicted wind power, it is useful to express the Betz mean power density $\bar{p}_d$, in terms of the Weibull distribution parameters as [40]:

$$\bar{p}_d = 0.5\left(\frac{16}{27}\right)\rho\left[\frac{\bar{u}}{\Gamma(1+1/k)}\right]^3\Gamma(1+3/k)$$

Equation 4.10

where the factor 16/27 represents the Betz limit, ρ is the air density and Γ is the gamma function.

A question worthy of investigation is the effect on the predicted $\bar{p}_d$ of the assumption that k = 1.8, for cases where the observed $k \neq 1.8$. Using Equation 4.10, the normalised power density can be written as:

$$\bar{p}_{d,norm(k_{obs}/k_{ass})} = \frac{\bar{p}_{d(k_{obs})}}{\bar{p}_{d(k_{ass})}} = \left[\frac{\Gamma(1+1/k_{ass})}{\Gamma(1+1/k_{obs})}\right]^3\frac{\Gamma(1+3/k_{obs})}{\Gamma(1+3/k_{ass})}$$

Equation 4.11

where $\bar{p}_{d,norm(k_{obs}/k_{ass})}$ represents the wind power density calculated using the observed shape factor, k_{obs}, normalised by that calculated using the assumed shape factor k_{ass}. This expression can be used to investigate the sensitivity of $\bar{p}_d$ to k_{ass}, independent of the mean wind speed. The curve described by Equation 4.11 is shown in Figure 4.10 for k_{ass} = 1.8.

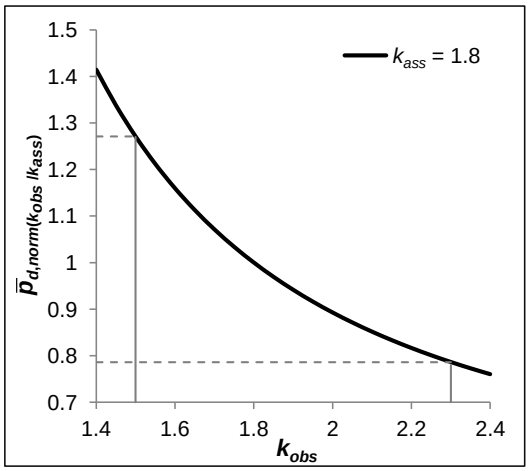

Figure 4.10: Normalised wind power density as a function of k_{obs} for k_{ass} = 1.8.

For each of the 38 measurement sites, the method of maximum likelihood (MML) was used to fit Weibull distributions to the observed wind speeds resulting in a range of shape factors, k_{obs} = 1.5 - 2.3. For these sites, Figure 4.10 shows that the fractional errors in the predicted $\bar{p}_d$ resulting from fixing k_{ass} = 1.8 are in the range of -0.3 to 0.2.

Errors in the predicted value of $\bar{p}_d$ will result from uncertainties in both k_{ass} and $\bar{u}_{pred}$ and these error sources are not independent. However, it is useful to estimate the relative magnitude of the fractional error in $\bar{p}_d$ resulting from these two sources. In the case of uncertainties in $\bar{u}_{pred}$, the fractional error in $\bar{p}_d$ will be a function of $\bar{u}_{obs}$. Given the range k_{obs} = 1.5 to 2.3 and the resulting fractional error in $\bar{p}_d$ of -0.3 to 0.2, we can define upper and lower bounds for the uncertainty in $\bar{u}_{pred}$ at which point the fractional error in $\bar{p}_d$ due to uncertainties in $\bar{u}_{pred}$ becomes comparable to that due to uncertainties in k_{ass}. Put another way, we can consider how accurately we must be able to predict the mean wind speed in order for the exact choice of shape factor to become significant in terms of the predicted wind power density. Figure 4.11 shows $\bar{u}_{pred}$ using Model B (i.e. the most accurate predictions based on the modifications suggested in this work) along with the upper and lower bounds for the uncertainty in $\bar{u}_{pred}$. For many sites, $\bar{u}_{pred}$ is outside of these bounds implying that in these cases, uncertainties in $\bar{u}_{pred}$ will be more significant than those in k_{ass}. This does not imply that improvements in k_{ass} will not improve the accuracy of the predicted $\bar{p}_d$, rather, it is simply an indication of the relative contribution of these uncertainties given the typical errors in $\bar{u}_{pred}$.

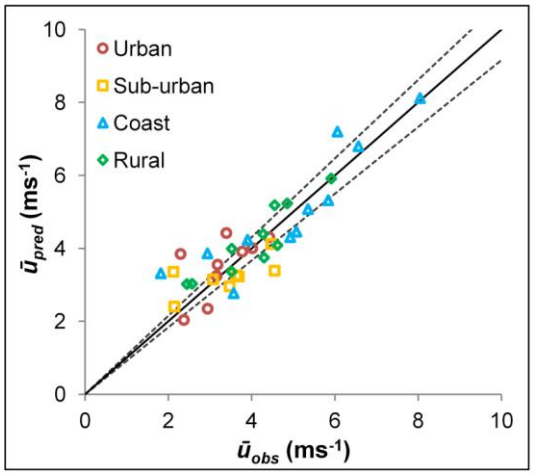

Figure 4.11: Predicted versus observed mean wind speeds using Model B, the solid line represents a one-to-one relationship. The dotted lines mark the upper and lower bounds at which the fractional error in $\bar{p}_d$ due to uncertainties in $\bar{u}_{pred}$ becomes comparable to that due to the assumption that k_{ass} = 1.8.

A question still remains as to how an optimum value for k_{ass} should be chosen. Since the relationship shown in Figure 4.10 is non-linear, the mean value of k is not necessarily equal to the optimum value. Instead it is possible to calculate the optimum value of k_{ass} that results in the smallest fractional error in $\bar{p}_d$ when combined across all measurement sites. Given a set of observed values $k_{obs,i}$ across i sites and an assumed constant shape factor k_{ass}, invoked to calculate the power density at the i^{th} site, the combined fractional error in $\bar{p}_d$ across all sites, $\varepsilon_{\bar{p}_d,tot}(k_{ass})$ is given by:

$$\varepsilon_{\bar{p}_d,tot}(k_{ass}) = \sum_i \left| 1 - \left[\frac{\Gamma(1+1/k_{ass})}{\Gamma(1+1/k_{obs,i})} \right]^3 \frac{\Gamma(1+3/k_{obs,i})}{\Gamma(1+3/k_{ass})} \right|$$

Equation 4.12

Equation 4.12 can be minimised iteratively with respect to k_{ass} to find the optimum value given the set of observed values $k_{obs,i}$. In the case of the 38 sites investigated in this study, the optimum value for k_{ass} is 1.89, which due to the clustering of k_{obs} values around 1.90, is also very close to the mean value.

Finally, it should be noted that the above discussion assumes that the data may be accurately represented by a Weibull distribution. This was found to be true for the 38 sites in this study, with an average difference of 5% in the $\bar{p}_d$ calculated using the best Weibull fit and the raw data. Hence, in the absence of site-specific information to the contrary, the assumption of a Weibull distribution appears to be justified.

4.3.4 Overall performance of the prediction methodology

The overall performance of a wind resource prediction methodology may ultimately be judged by its ability to accurately predict the available wind power, as tested by comparison with onsite measurements. In the current study, an adjustment factor was applied to the measured mean wind speeds for sites with less than five years' data to account for inter-annual variation. Hence, in order to calculate the observed $\bar{p}_d$, a Weibull distribution was fitted to each of the sites using MML and the extracted value of k_{obs} was used in conjunction with the adjusted mean wind speed to calculate a value representative of the observed wind power density using Equation 4.10.

The error metrics detailed in Equation 4.4 to Equation 4.6 were then used along with an analysis of the residuals to investigate the success of the prediction methodologies. Figure 4.12 compares the distribution of residual errors in the wind power density ($\varepsilon_{\bar{p}_d}$) for the WYET along with Models A and B, where Model B also incorporates the optimised value k_{ass} = 1.89. For the WYET, the inter-quartile range over all sites is -14 to 21 Wm^{-2}. Since this represents the range covered by 50% of the data, this is a useful guide to the likely error in the predicted value of p_d using this methodology but note should also be taken of the 5th and 95th percentiles, which range from -35 to 49 Wm^{-2}. If Model B is used, incorporating the increased fetch, taking account of wind direction and using the optimised k_{ass}, the inter-quartile range is reduced to between -13 and 16 Wm^{-2} and the 5th and 95th percentiles to between -37 and 32 Wm^{-2}.

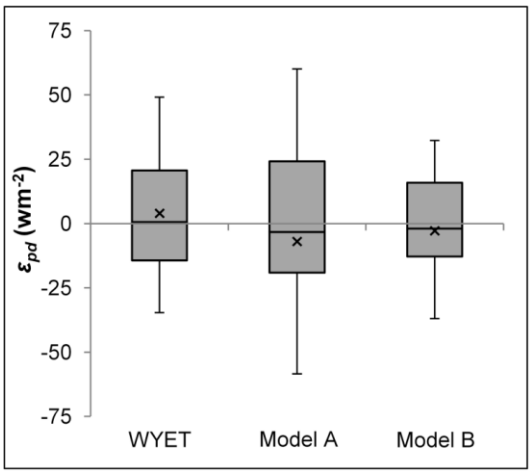

Figure 4.12: Distribution of residual errors in the wind power density for different implementations of the prediction methodology. The horizontal lines and crosses represent the median and mean respectively and the error bars represent the 5th and 95th percentiles.

Table 4.6 compares the error metrics for the WYET and Models A and B. On average, these metrics show a clear improvement in the accuracy of the predictions when using Model B compared to the other approaches. Table 4.6 also includes the mean bias error (MBE). The MBE is a useful indicator as to whether on average the predicted $\bar{p}_d$ is an over- or underestimate. Averaged across all terrains, the MBE is similar for Models A and B and slightly worse for WYET. However, when split into terrain categories, Model B shows a clear improvement, reducing the negative bias at coastal sites and positive bias at rural sites that are seen in both Model A and WYET.

	Methodology	Urban	Suburban	Coastal	Rural	All
MAE (wm^{-2})	WYET	16	15	33	25	**23**
	Model A	14	17	68	30	**35**
	Model B	13	14	33	17	**20**
MBE (wm^{-2})	WYET	6.29	-6.88	-25.8	12.8	**-4.06**
	Model A	8.31	-8.76	-10.3	19.0	**2.14**
	Model B	7.25	-7.51	2.44	7.34	**2.77**
RMSE (wm^{-2})	WYET	20	19	48	33	**33**
	Model A	19	22	102	37	**59**
	Model B	18	18	43	21	**28**
%Error	WYET	93	64	70	53	**70**
	Model A	88	63	83	57	**73**
	Model B	85	59	73	37	**63**

Table 4.6: Power density error metrics compared over four terrain types using the WYET and Models A and B.

It is noteworthy here that due to the cubic relationship between wind power and wind speed, over-estimates of $\bar{u}_{pred}$ carry a higher penalty in terms of the predicted $\bar{p}_d$ than under-estimates of the same magnitude. Hence, reduced errors in $\bar{u}_{pred}$ do not automatically result in reduced errors in the predicted $\bar{p}_d$. This effect can be seen in the coastal sites where Model B performs better than the WYET in terms of $\bar{u}_{pred}$ but except for the MBE metric, these improvements are not evident in the predicted $\bar{p}_d$. This is because for the WYET, the errors in $\bar{u}_{pred}$ at coastal sites are skewed towards under-prediction and hence on average they have a disproportionately smaller effect on the predicted $\bar{p}_d$.

Table 4.6 also demonstrates that errors in $\bar{u}_{pred}$ can result in significant errors in the predicted $\bar{p}_d$ highlighting the need for caution when using such models to assess site viability. In spite of this, the improved predictions obtained using Model B demonstrate that with even relatively simple changes, a consistent improvement in the accuracy of the predicted wind resource can be achieved.

4.4 Results and Discussion II – propagation of errors and sensitivity analysis

Given the inherent uncertainties in the BS model, several questions are worthy of further investigation:

(i) How might the propagation of errors in the predicted wind resource be quantified?

(ii) What are the most significant sources of error in the BS model?

(iii) Can such approaches be considered of sufficient accuracy to be used as site screening tools?

These questions are explored in more detail in the following sections.

4.4.1 Propagation of uncertainties

While the analysis in Section 4.3 allows the average prediction errors to be quantified given measurements at specific sites, it is useful to be able to estimate the likely prediction errors in advance by quantifying the propagated uncertainties.

Uncertainties in the final mean wind speed and wind power density predictions arise firstly from assumptions and simplifications inherent in the model itself and secondly from uncertainties in the model input parameters. In the following discussion we are concerned with the uncertainties arising from the latter. In each of the five stages of the model implementation, appropriate input parameters must be chosen. These include the input reference wind speed, the regional and local aerodynamic parameters, the estimated blending height and the Weibull shape factor required to construct a distribution of wind speeds. Uncertainties in the values of these parameters combine to produce uncertainties in the final model predictions. Since accurately estimating the regional and local aerodynamic parameters is known to be particularly challenging, these are likely to be the dominant error source in the mean wind speed prediction. These errors combine with the uncertainty in the Weibull shape factor to produce errors in the final mean wind power density prediction.

To quantify the effect of these uncertainties, a quasi-random Sobol sampling approach, implemented in the MATLAB programming environment, was applied to each individual site prediction. A Sobol [147] sequence is a low discrepancy numerical sequence that allows a multi-dimensional parameter space to be filled efficiently with minimal gaps. Given a model output based on a number of input parameters, sampling using a Sobol sequence allows an efficient estimation of the

overall output uncertainty related to various combinations of input parameters. For the mean wind speed prediction, a five dimensional Sobol sequence was used to sample a range of values for the four aerodynamic parameters of regional and local z_0 and d as well as the blending height z_b. Preliminary tests at four sites indicated that a Sobol sequence of length 1024 was sufficient for convergence of the output mean and variance. An approximate uncertainty in the default z_0 and d values for each terrain of +/- 35% was estimated, based on the ranges recommended in a comprehensive study by Grimmond and Oke [75]. The same uncertainty range was applied to z_b since it was found that larger ranges resulted in parameter combinations within the Sobol sequence that were not physically viable.

The sampling approach gives a distribution of predicted mean wind speeds for each site which reflects the propagation of uncertainties within the input parameters to the outputs. Presented in Figure 4.13 are the mean wind speed predictions from these distributions with uncertainties represented by plus or minus twice the standard deviation across the 1024 samples (+/- 2σ represents an approximate 95% confidence interval assuming a normal distribution). On average, these uncertainties represent around +/- 0.4 ms^{-1} (11%) in $\bar{u}_{pred}$ which is a comparable order of magnitude to, although somewhat smaller than, the average observed error in $\bar{u}_{pred}$ of 0.5 ms^{-1} (16%) shown in Table 4.5

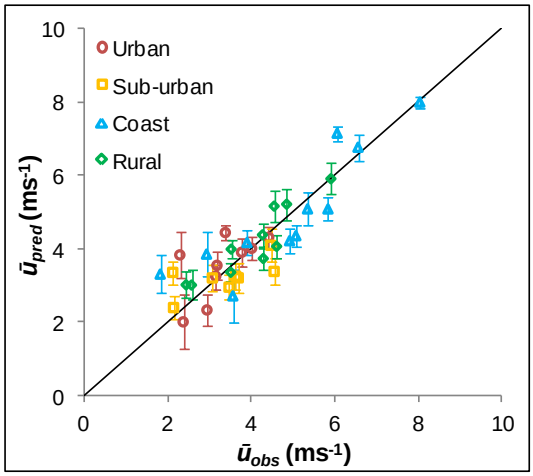

Figure 4.13: Predicted versus observed mean wind speed for all sites using Model B. Symbols represent the mean prediction using a five-dimensional Sobol sampling sequence to account for uncertainties in the input aerodynamic parameters. The error bars represent the mean prediction +/- 2σ. The solid line represents a one-to-one relationship

While this approach provides a first estimate as to the uncertainty in $\bar{u}_{pred}$, Figure 4.13 shows that the uncertainty is underestimated since many of the error bars do not cross the observed values (represented by the solid line). This is not surprising

since this analysis only accounts for uncertainties in the aerodynamic input parameters and does not consider the specific model assumptions, errors in the NCIC input climatology, or the fact that the mean wind speed predictions are a spatial average which may be affected by local perturbations to the flow.

For the mean wind power density prediction, a six dimensional Sobol sequence of length 1024 was used, with the Weibull shape factor as the sixth sampling parameter. Weibull shape factors in the range k = 1.5 - 2.3 were employed based on the range of Weibull shape factors observed at the 38 sites.

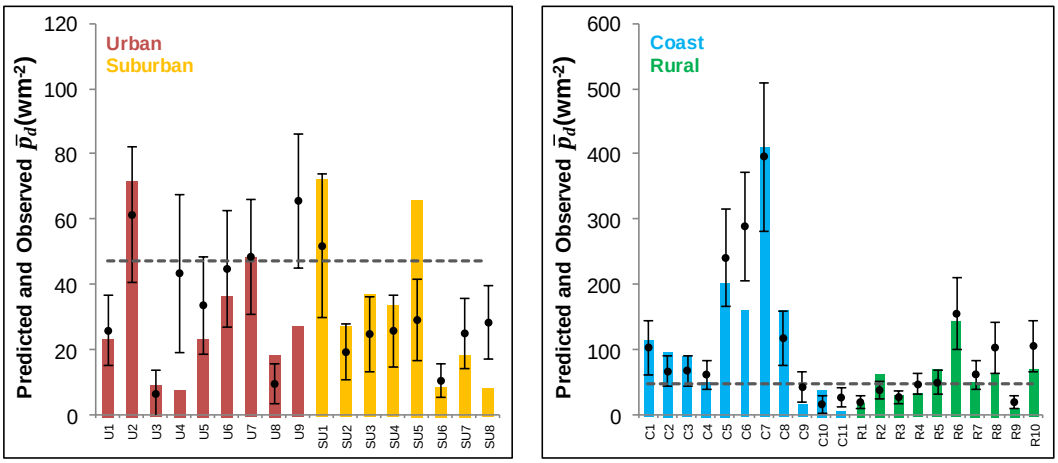

Figure 4.14: Predicted (symbols) and observed (bars) mean wind power density for all sites using Model B. The two y-axis scales are used for clarity. Symbols represent the mean prediction using a six-dimensional Sobol sampling sequence to account for uncertainties in the input aerodynamic parameters and Weibull shape factor. The error bars represent the mean prediction +/- 2σ. The dotted horizontal line represents a viability criterion of $\bar{p}_d$ ≥ 47 Wm^{-2} (Section 4.4.3).

Figure 4.14 shows the observed wind power density at each site along with the mean predictions from the Sobol samples with uncertainties represented by +/- 2σ. On average, these uncertainties represent approximately +/- 24 Wm^{-2} (61%) of the predicted $\bar{p}_d$ at each site. As with the predicted mean wind speed, this uncertainty is similar to the average observed error in $\bar{p}_d$ of +/- 20 Wm^{-2} (63%) presented in Section 4.3.4. The estimated percentage uncertainties are highest for urban sites (82%) and lowest for rural sites (47%) in line with the observed percentage errors in Table 4.6. For around a third of the sites the error bars do not cross the observed values. Once again this highlights that some uncertainties are not fully accounted for and hence the estimated uncertainties should not be taken as fully representative of 95% confidence intervals. Figure 4.14 also demonstrates that for the sites and anemometer heights considered in this study, the best wind resource

tends to be at the coastal and rural sites while the sub-urban and urban sites tend to have a considerably smaller available resource.

4.4.2 Global sensitivity analysis

In addition to estimating the overall prediction uncertainty using a sampling approach, it is also informative to understand the relative contributions to this uncertainty from each of the six input parameters. Ultimately, such information can be used to improve the model by obtaining better estimates of the most significant parameters. To investigate these contributions, a global sensitivity analysis was conducted using the GUI-HDMR (graphical user interface – high dimensional model representation) software tool in MATLAB, which has successfully been applied in a number of environmental modelling contexts [148]. Given a range of specified input parameters, (in the present case these are sampled from a six-dimensional Sobol sequence as described above), and the corresponding model output (predicted wind power density) the GUI-HDMR applies a variance-based sensitivity analysis to estimate the relative importance of each of the input parameters.

The output sensitivity to each parameter is quantified by means of a first order sensitivity index S_i, which is a measure of the fractional contribution of the i^{th} parameter to the output variance. In the current study, only first order S_i estimates were considered in detail, since higher order sensitivities, which represent interactions between input parameters, were found to be small. A detailed description of the of the GUI-HDMR is available in reference [148]. Table 4.7 shows the first order S_i estimates for each of the six model input parameters at the target sites with respect to the output prediction $\bar{p}_d$. Parameters with larger index values (lighter shading) indicate that the output is more sensitive to that parameter.

Parameter	U1	U2	U3	U4	U5	U6	U7	U8	U9	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
d_{reg} (m)	0.19	0.13	0.02	0.09	0.18	0.20	0.16	0.10	0.05	0.05	0.02	0.03	0.03	0.08	0.02	0.00	0.00	0.05	0.01	0.05
$z_{0,reg}$ (m)	0.23	0.04	0.01	0.06	0.13	0.15	0.18	0.09	0.15	0.02	0.04	0.02	0.02	0.03	0.01	0.01	0.01	0.02	0.00	0.01
z_{bh} (m)	0.06	0.02	0.01	0.01	0.01	0.01	0.04	0.09	0.00	0.07	0.07	0.04	0.07	0.04	0.00	0.03	0.25	0.01	0.01	0.01
d_{local} (m)	0.01	0.03	0.63	0.34	0.12	0.05	0.00	0.29	0.00	0.16	0.08	0.09	0.11	0.00	0.00	0.00	0.03	0.34	0.59	0.35
$z_{0,local}$ (m)	0.01	0.05	0.21	0.19	0.08	0.04	0.00	0.17	0.00	0.17	0.09	0.10	0.13	0.03	0.02	0.00	0.03	0.26	0.22	0.28
k	0.45	0.68	0.07	0.25	0.41	0.49	0.60	0.21	0.80	0.51	0.68	0.68	0.63	0.81	0.94	0.95	0.66	0.27	0.13	0.26

Parameter	SU1	SU2	SU3	SU4	SU5	SU6	SU7	SU8	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	All
d_{reg} (m)	0.12	0.10	0.12	0.09	0.09	0.11	0.12	0.11	0.11	0.11	0.11	0.07	0.11	0.11	0.10	0.11	0.10	0.11	**0.09**
$z_{0,reg}$ (m)	0.05	0.04	0.06	0.03	0.04	0.06	0.16	0.04	0.05	0.04	0.04	0.07	0.04	0.05	0.03	0.04	0.04	0.03	**0.06**
z_{bh} (m)	0.34	0.05	0.40	0.02	0.02	0.15	0.27	0.03	0.16	0.19	0.19	0.11	0.21	0.15	0.13	0.20	0.06	0.19	**0.10**
d_{local} (m)	0.00	0.16	0.00	0.17	0.16	0.13	0.00	0.11	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	**0.11**
$z_{0,local}$ (m)	0.03	0.20	0.02	0.21	0.20	0.17	0.01	0.17	0.17	0.03	0.04	0.02	0.04	0.03	0.07	0.07	0.21	0.05	**0.10**
k	0.43	0.42	0.37	0.44	0.45	0.34	0.42	0.50	0.33	0.61	0.60	0.71	0.58	0.64	0.64	0.57	0.39	0.59	**0.51**

Table 4.7: First order sensitivity indices S_i for the six BS model input parameters of regional displacement height and roughness length (d_{reg}, $z_{0,reg}$), blending height (z_{bh}), local displacement height and roughness length (d_{local}, $z_{0,local}$) as well as Weibull shape factor (k) at each target site. Indices are calculated with respect to the output of predicted wind power density ($\bar{p}_d$). Shading indicates the relative contribution of each parameter to the uncertainty, from smallest (dark) to largest (light). The average values across all sites are also shown.

It is clear that the relative sensitivities vary depending on the site, however, some general trends can be observed. Based on the average S_i values across all sites, the first order sensitivities account for 97% of the output variance, hence higher order effects do exist but they are relatively small on average. The most significant single parameter is k which accounts for around half of the output variance. Note that the remaining five parameters all relate to $\bar{u}_{pred}$ and account for the remaining output variance. The results of Section 4.3.3, based on actual observations of the prediction errors, indicated that the predicted $\bar{p}_d$ is often more sensitive to $\bar{u}_{pred}$ than k. However, the sensitivity analysis presented here is unable to account for all the uncertainties in $\bar{u}_{pred}$, while it does fully account for the uncertainty in k by using the full range of observed values at the target sites. Although this may result in a slightly exaggerated estimation of the sensitivity to k, the Weibull shape factor is clearly an important parameter. In addition, since improvements to $\bar{u}_{pred}$ require better estimates of at least five input parameters, obtaining site-specific estimates of k is potentially an efficient way of improving the model performance. A possible means of obtaining improved estimates of k is the use of regional forecast data in the form of long-term time series of wind speeds, from which, wind speed distributions can be calculated. This topic is returned to in more detail in Chapter 8.

Of the remaining parameters, on average, the predicted $\bar{p}_d$ is most sensitive to the local aerodynamic parameters d_{local} and $z_{0,local}$. This is particularly the case at urban and sub-urban sites, and by extension, coastal sites since many of these

include a sub-urban onshore fetch. As noted previously, these parameters are notoriously challenging to estimate and more sophisticated methods, such as the use of site-specific building height data may be required in order to obtain noteworthy improvements [86]. The z_{bh} parameter tends to be more significant at sites where there are large differences between the regional and local aerodynamic parameters. For rural sites, the sensitivity to z_{bh} was consistently high due to the fact that d_{local} is nominally set to zero (with a few exceptions), while on a regional scale there are frequently patches of forestry that result in a high value of d_{reg} and hence a large difference between the local and regional parameters. In cases where the regional and local wind profile are significantly different, it may be appropriate to obtain more precise estimates of z_{bh} using a boundary layer growth law as discussed in Section 2.4.3.

The above analysis offers suggestions for obtaining improved estimates of some of the most critical input parameters. It should be borne in mind however, that obtaining such estimates requires detailed, site-specific information, and this negates some of the advantages of the approach, which is aimed at obtaining rapid predictions at multiple sites using simple parameterisations of the surface.

4.4.3 Boundary layer scaling as a site screening tool

One way of leveraging the advantages of rapid implementation at multiple sites is to apply the tool in a site screening context to test sites against some criterion of viability. Sites passing the criterion would be judged worthy of further investigation using more detailed approaches such as onsite measurements.

Defining a viability criterion is non-trivial since one may consider environmental viability (the ability of a turbine to produce sufficient energy to repay its embedded carbon) and financial viability, (the ability of a turbine to produce sufficient energy to repay the financial investment). These will vary greatly depending on the materials and costs of specific turbines as well as the availability and level of government sponsored financial incentives. In the context of small-scale wind turbine installations, recent studies concerned with assessing city-wide wind energy potential [146, 149] have used a mean wind speed viability criterion of 4-5 ms^{-1}, these values are also in line with industry advice offered by the UK trade body RenewableUK [150].

While a minimum mean wind speed is a useful starting point, the available wind power will depend both on the mean wind speed and the distribution of wind speeds at the proposed site. Hence, it is useful to express this criterion in terms of a

minimum wind power density. Assuming a Weibull distribution of wind speeds, the mean Betz power density in the wind $\bar{p}_d$, can be obtained from Equation 4.10. Using a minimum mean wind speed of 4 ms^{-1} and a Weibull shape factor k = 1.9, as representative of UK sites [16], this equates to a viability criterion of $\bar{p}_d \geq$ 47 Wm^{-2}.

An intuitive feel for this number can be gained through some simple calculations. Assuming 50% of the available Betz power can be converted to electrical power [11], (this is a broad approximation since efficiency will vary with wind speed and turbine design), for a small-scale turbine with a blade diameter of 2 m, this equates to an average power production of 74 W and an annual energy production of 647 kWh. For a larger turbine with a blade diameter of 6 m, this equates to an average power production of 664 W and an annual energy production of 5821 kWh. Note that efficiencies may be reduced for building mounted turbines in urban areas due to turbulent wind flows. In the following analysis, the minimum power density criterion of $\bar{p}_d \geq$ 47 Wm^{-2} is applied to screen for viable sites.

To implement the site pre-screening process we wish to identify sites which are likely to be unsuitable as judged by the viability criterion of $\bar{p}_d \geq$ 47 Wm^{-2}. To reduce the likelihood of mistakenly excluding viable sites, only sites where even the most optimistic wind power prediction is below the viability criterion should be excluded. Taking this approach, all sites where the predicted $\bar{p}_d$ plus the associated uncertainty (top of each error bar) is below the viability criterion should be deemed unsuitable. These sites can be identified directly from Figure 4.14. The observed and BS predicted viability for each site are compared in Table 4.8, in addition, the predictions obtained using the NOABL-MCS approach are also shown. Crosses indicate that a site fails the viability criterion, ticks indicate that it passes. Correct viability predictions are highlighted in green, incorrect predictions are red.

Site	Obs	BS Model	NOABL-MCS	Site	Obs	BS Model	NOABL-MCS	Site	Obs	BS Model	NOABL-MCS	Site	Obs	BS Model	NOABL-MCS
U1	×	×	×	SU1	✓	✓	×	C1	✓	✓	×	R1	×	×	×
U2	✓	✓	×	SU2	×	×	×	C2	✓	✓	✓	R2	✓	✓	×
U3	×	×	×	SU3	×	×	×	C3	✓	✓	×	R3	×	×	×
U4	×	✓	×	SU4	×	×	×	C4	✓	✓	×	R4	×	✓	×
U5	×	✓	×	SU5	✓	×	×	C5	✓	✓	✓	R5	✓	✓	✓
U6	×	✓	×	SU6	×	×	×	C6	✓	✓	✓	R6	✓	✓	✓
U7	✓	✓	×	SU7	×	×	×	C7	✓	✓	✓	R7	✓	✓	✓
U8	×	×	-	SU8	×	×	×	C8	✓	✓	×	R8	✓	✓	×
U9	×	✓	×					C9	×	✓	×	R9	×	×	×
								C10	×	×	-	R10	✓	✓	×
								C11	×	×	×				

Table 4.8: Comparison of observed and predicted viability for each site based on a criterion of $\bar{p}_d \geq$ 47 Wm^{-2}. Crosses indicate non-viable sites, ticks indicate viable sites. Correct predictions are highlighted in green, incorrect predictions are red.

Table 4.8 shows that the BS predicted viability is correct in 32 out of 38 cases. In a site screening context, it is particularly important that promising sites are not excluded from further investigation based on an incorrect viability prediction, (represented by red crosses in Table 4.8). Using the BS model, 14 non-viable sites would correctly be excluded from further investigation with only one incorrect exclusion. This is a promising result given that the BS model requires no onsite wind measurements. In comparison, the NOABL-MCS approach makes correct viability predictions in only 25 out of 38 cases and incorrectly excludes 11 viable sites. These results indicate that the BS model can be a valuable tool when used in a site screening context and can add significant value compared to the NOABL-MCS method. A final point worthy of note is that Table 4.8 indicates that the BS model may result in overly optimistic predictions at urban sites since four sites in the urban category were incorrectly predicted to be viable. This is possibly related to the fact that the approach does not account for local scale building effects which can significantly affect the wind resource, particularly for turbines mounted close to roof height. This issue appears to be less significant when using NOABL-MCS due to the tendency for the approach to underestimate the wind resource.

4.5 Conclusions

A boundary layer scaling model for estimating the wind energy resource has been investigated by comparing its predictions to wind speed data collected from 38 UK sites located in a range of terrains and suggestions have been made for improving its implementation. The MAE, RMSE, %Error and distribution of residuals were used to quantify the errors associated with the methodology and to judge attempts to improve its implementation. For the original methodology, as implemented in the WYET, averaged across all terrain types, the MAE and %Error were found to be 0.63 ms^{-1} and 18%, respectively.

For urban sites, the effect of defining the canopy height as the local mean building height, as obtained from a database of LiDAR measurements, was investigated. However, it was found that canopy heights based on a set of generic values performed better. While this may be partly due to discrepancies between the measurement procedures used for the anemometer and building heights, it also highlights the need for more accurate descriptors of local roughness in urban areas. For all terrains, the effect of an increased fetch and the use of directionally-dependent regional roughness parameters on the accuracy of $\bar{u}_{pred}$ was also

investigated. Averaged across all sites, it was found that taking account of these factors reduced the MAE and %Error to 0.52 ms^{-1} and 16%, respectively.

The analysis was extended to investigate the success of the methodology in predicting $\bar{p}_d$ based on the assumption of a Weibull distribution with a fixed shape factor. An expression was derived to describe the fractional error in $\bar{p}_d$ resulting from the use of a fixed value of k_{ass} and the significance of this error in relation to errors in $\bar{u}_{pred}$ was quantified. Using the values of k_{obs} extracted from each test site, this expression was minimised in order to obtain an optimum value of $k_{ass} = 1.89$.

To investigate the overall performance of the approach in predicting the wind energy resource, three implementations of the methodology were used to predict $\bar{p}_d$ at each site and these predictions were compared to the measured values. Using the modified methodology, Model B, it was found that the MAE and %Error in $\bar{p}_d$ could be reduced from 23 Wm^{-2} to 20 Wm^{-2} and from 70% to 63%, respectively, when averaged across all sites. These results indicate that while the approach clearly has limitations, improvements can be achieved even with relatively simple modifications to the methodology.

An uncertainty analysis using a Sobol sampling procedure was used to quantify the combined uncertainties in $\bar{u}_{pred}$ and the predicted $\bar{p}_d$ arising from uncertainties in the regional and local aerodynamic parameters, blending height and Weibull shape factor. The analysis revealed average predicted uncertainties of +/- 0.4 ms^{-1} (11%) in $\bar{u}_{pred}$ and +/- 24 Wm^{-2} (61%) in the predicted $\bar{p}_d$. While these values offer useful first approximations, they are likely to be underestimates since they do not account for all the possibly uncertainties involved in the implementation of the modelling approach. The sampling procedure was extended to include a global sensitivity analysis using the GUI-HDMR software tool. The results indicated that on average the predictions of $\bar{p}_d$ are most sensitive to the values of the local aerodynamic parameters and the Weibull shape factor.

Due to the costs and timescales involved in onsite measurement campaigns, semi-empirical models such as the one presented here, can be of great value in assessing the wind resource quickly and cheaply. However, such models should be applied with an awareness of their inherent uncertainties. Given the magnitude of the average errors in the predicted mean wind speed and wind power identified in this study, such models are perhaps better applied in a scoping context to identify test sites against a viability criterion. At promising sites, these predictions can then be supplemented by onsite measurements in order to minimise the investment risk.

However, since long-term onsite measurement campaigns are unlikely to be accessible to small-scale wind energy developers, there is a clear case for the development of data-driven approaches that are capably of accurately predicting the wind resource using short-term onsite measurements. The viability of such approaches will be discussed in the following chapters.

5 Data Efficient Measure-Correlate-Predict Approaches to Wind Resource Assessment

5.1 Overview

Since simple physical modelling approaches may not provide sufficiently accurate wind resource estimates, as outlined in Chapter 4, the current chapter investigates how short-term onsite measurements may be best used to add value to the site assessment procedure. Making the assumption that, due to financial and practical constraints, the onsite measurement period may be severely restricted, a key question is how best use may be made of any onsite wind data that can be obtained.

The measure-correlate-predict (MCP) approaches discussed in Chapter 3 represent an established strategy for increasing the value of onsite measurements by correcting for inter-annual variations. However, MCP has received little attention in the context of small-scale wind energy where the reduced measurement period may necessitate corrections for both intra- (seasonal) and inter-annual variations. While long-term measurements are clearly desirable in reducing uncertainty, the focus of this work is to establish the feasibility of applying MCP approaches to measurement periods of much less than one year. This is with the aim of determining whether data-driven resource assessment can be made more accessible to the small-scale wind industry. Such an approach should not be seen as an alternative to long-term measurements, which can significantly improve accuracy but rather, as a tool that can be used to increase the value of short-term measurements in the absence of long-term, onsite data.

As outlined in Section 3.2.5, although a limited number of previous studies have considered the effect of short measurement periods on MCP performance, there is a clear need for more detailed work in this area. Many studies are restricted to a small number of sites, a specific terrain type or a relatively short data record used for evaluation. In addition, the effects of seasonal variability, residual scatter and the use of training data from different years are often not explicitly investigated. Furthermore, MCP performance is in some cases judged using limited criteria such as the predicted mean wind speed rather than parameters related to the full wind speed distribution.

The contribution of the work described in this chapter is a rigorous investigation of MCP performance using very short training periods at a large number of potential small turbine sites in a range of UK terrains. Three MCP approaches were investigated using multiple 3 month onsite training periods shifted throughout an 11 year data record using a sliding window technique. This allowed the effect of both seasonal and inter-annual variability in the error metrics to be fully investigated and the calculation of robust error statistics. Residual scatter in the regression relationships was explicitly considered and metrics were applied to quantitatively assess the average errors in the predicted mean wind speed, mean wind power density, standard deviation of wind speeds and Weibull shape factor over the multiple test and training periods.

The main objectives of this study can be summarised as follows:

(i) Development of a sliding window procedure to facilitate robust error estimates that account for both intra- and inter-annual variations.
(ii) Quantification of the typical prediction errors that may be expected when applying the MCP approach to very short onsite measurement periods and hence an assessment of the added value of a short measurement campaign.
(iii) Comparison of the performance of different linear MCP techniques when applied to very short measurement periods and an investigation of the effect of accounting for the residual scatter.
(iv) An assessment of the magnitude and sign of seasonally dependent prediction errors and recommendations regarding the best measurement season.
(v) Comparison of the average prediction errors using the MCP approach and the boundary layer scaling approach detailed in Chapter 4.

This chapter is organised as follows: Firstly, a description of the sliding window technique used to obtain multiple training and test periods is given. Next, the meteorological measurements, site selection procedure and error metrics used to assess the MCP approaches are described. This is followed by an investigation of the performance of the different MCP approaches, including a detailed consideration of seasonal effects. Finally, the MCP error statistics are compared to those obtained using the BS model described in Chapter 4 and overall conclusions are drawn.

5.2 Methodology

In this work, three of the linear MCP techniques described in Section 3.2 have been implemented, namely, linear regression (LR), linear regression with a Gaussian scatter model (LR2) and the variance ratio method (VR). As discussed in Section 3.2, while LR2 is not a new approach (in effect it is simply a more sophisticated implementation of LR), it has not been widely applied in the MCP literature despite its apparent advantages. Hence in this work, an explicit distinction is made between LR and LR2 in order to compare their performance.

5.2.1 The sliding window approach

It is typical for MCP studies [125, 151] to identify a fixed training period, during which the reference/target site relationship is estimated, and a fixed test period, during which the MCP approach is used to make long-term predictions. Replicating the implementation of a real MCP assessment procedure, the training period may be chosen as the most recent 12 months of wind observations and the test period as the long-term historical observations, as shown in Chapter 3, Figure 3.3. While this seems logical, it may not be the most appropriate approach for rigorously comparing the performance of MCP techniques.

Due to the inter-annual variability of synoptic-scale weather events, a single training year may not be fully representative of the types of reference/target site relationships that the MCP approach will be required to resolve. In the current study this issue is significantly compounded by the fact that the training period is the length of a single season. Hence, seasonal variability in the reference/target site relationships will not be accounted for unless multiple training periods are used.

To resolve these issues, a sliding window technique was implemented using multiple 3 month training and 10 year test periods throughout an 11 year data record. The approach is shown schematically in Figure 5.1 and can be summarized as follows:

(i) A training window spanning a full year is first defined and this is shifted in steps of one month throughout the 11 year reference and target site data records resulting in a total of 120 steps. The remaining data not covered by the training window are designated as the test data, which always cover a combined period of 10 years.

(ii) At each step, a three month training period at the start of the training window is used to extract the regression parameters for the MCP approaches of LR, LR2 and

VR. This represents the short-term onsite measurement period proposed in this study.

(iii) The MCP approaches are then applied to the 10 year test data at the reference sites in order to predict a concurrent time series of wind speeds at the target sites. This represents the long-term predictions extrapolated from the short-term onsite measurements.

(iv) Error metrics are calculated at each step by comparing the predictions with the observed values over the test period at each target site.

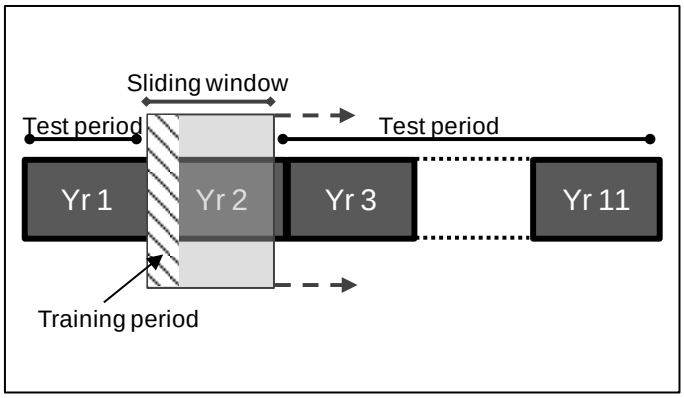

Figure 5.1: Schematic diagram of the sliding window technique used to test the MCP predictions across the entire data record. The test periods move with the training window such that the two never overlap.

Note that while the training period spans just 3 months, a training window covering a full year is used to ensure that the test period always covers an integer number of years thus avoiding seasonal variations in the test data. This approach results in a total of 120 predictions of the 10 year wind resource spanning all training seasons and years within the 11 year data record.

A 3 month training period was chosen for this study based on empirical observations of the average data length required to establish regression relationships between the hourly wind speeds at the reference/target site pairs. In addition, a 3 month period was found to be convenient for investigating seasonal variability based on the nominal seasons of autumn, winter, spring and summer. The impact of the training length on the accuracy of MCP predictions is returned to in more detail in Chapter 7.

5.2.2 Implementation of the MCP techniques

A description of each of the three MCP approaches is given in Chapter 3, Section 3.2.2.2 (LR), Section 3.2.3.1 (LR2) and Section 3.2.2.3 (VR).

Each approach was applied sector-wise to data which was first binned according to the reference site wind direction. The sector approach was used to account for the directional dependence of the upwind roughness which previous work [16] has shown affects the scaling between the reference and target site wind speeds. In line with previous studies [107, 115], angular bins of width 30º were used resulting in 12 separate regressions for each reference/target site pair. For angular bins with less than 20 data entries, the regression parameters were found to behave erratically and hence in these cases, the parameters were obtained by applying a global fit to data from all bins.

5.2.3 Meteorological measurements

In order to assess the performance of the MCP approaches, long-term meteorological data (hourly averaged wind speed and direction) were obtained from the Met Office anemometer network (MIDAS) [140] at a total of 37 UK sites for an 11 year period between August 2001 and July 2012. These sites consisted of 22 target sites (locations where we wish to predict the wind resource) and 15 reference sites (used to implement the MCP approaches). The target sites formed a subset of the sites considered in Chapter 4, although in the present study they cover a longer period. The general characteristics of these data are described in Section 4.2 and further details of specific relevance to the MCP approach are described below.

As outlined in Chapter 4, the long-term MIDAS data record may be subject to biases due to historical changes to measurement equipment, mast siting and surrounding land usage. Such changes are not always recorded in the MIDAS metadata record. However, for the 11 year data record used in the present study, three relevant changes were recorded, namely, replacement masts at sites C8 and R2, and felling of trees at site Rf15. For these sites, the wind data record was examined close to the dates where the changes occurred but no clear anomalies were detected. To further screen for changes not recorded in the metadata, time series plots of the monthly mean wind speeds over the entire 11 year data record were visually examined at all reference/target site pairs. The purpose of this was to check for significant deviations between the trends recorded at each of the site pairs. No significant anomalies were detected during this screening.

After performing the quality control steps detailed in Section 4.2, it was necessary to ensure the data from all sites were time-aligned since the MCP approach is applied to concurrent time series. This was achieved by padding missing entries in the time series with NaNs to ensure a consistent data record across all sites covering the entire 11 year data period. Further precautions were required in

dealing with lulls (entries with zero wind speed and/or wind direction). The presence of lulls presents several difficulties when applying an MCP approach. Firstly, when binning wind speeds according to angular sectors, it is not possible to assign lulls to a particular wind direction bin, although in some studies lulls have been randomly distributed between all bins [18]. Secondly, when applying Weibull fits to the data, and hence when implementing the bivariate Weibull MCP approach (discussed in Chapter 7), lulls cannot generally be accounted for due to the mathematical form of the distribution. It should also be noted that lulls recorded in the MIDAS data record may not correspond to a true period of zero wind speed due to the relatively coarse data resolution (1 knot = 0.51 ms^{-1}) as well as the poor response characteristics of some cup anemometers at low wind speeds [142].

Hence, in order to apply a consistent approach, all lulls were removed from the data before application of the MCP techniques. Since concurrent data are required, removing a lull (or NaN) at a site, also requires removing the concurrent observation at the site's pair. To better understand the effect of this procedure, the target site mean wind speed $\bar{u}_{tar}$ and wind power density $\bar{p}_d$ were calculated using both the raw data and the data with lulls removed over the full 11 year period. It was found that removing the lulls resulted in a relatively small average increase in $\bar{u}_{tar}$ of 2.2% and $\bar{p}_d$ of 2.7%. For consistency, the MCP error metrics were calculated after the removal of NaNs and lulls to ensure that the predictions were compared against the same data on which they were based. Figure 5.2 shows a schematic illustration of the processing of raw data to remove NaNs and lulls, including concurrent data points at the reference or target sites.

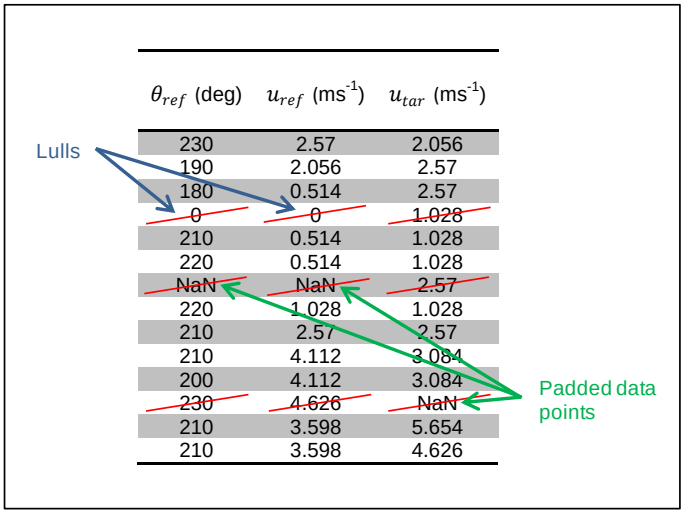

Figure 5.2: Schematic diagram illustrating the removal of lulls and NaNs including concurrent observations at the reference or target sites.

Before time alignment and subsequent processing as described above, the reference and target site observations had an average availability of 98% with a minimum of 94%. Post processing, the average availability was 94% with a minimum of 87%.

5.2.4 Site selection

In order to fully account for inter-annual variations in the reference/target site correlations and to obtain robust error statistics, only sites with a data record covering at least 11 years were chosen for this study. This, along with the MCP requirement of nearby long-term reference sites, limited the total number of available sites. The target sites were specifically chosen to be representative of different terrain types, urban, sub-urban, rural and coastal and this classification was made through examination of satellite images as described in Section 4.2.6. Since some of the sites studied in Chapter 4 were excluded from this study due to their short data record, it was not possible to include a sufficient number of sites from each terrain category to enable robust conclusions to be made regarding the effect of terrain on MCP performance. Instead, the main intention of including different terrain types was to reflect the range of scenarios which may be encountered by small-scale wind energy developers in real-world site assessments. However, for some terrain types, there were sufficient sites to observe general trends.

The standard Met Office observational practice is the location of anemometers at a height of 10 m above open ground, although this is not always possible in urban areas. In the case of the urban sites used in this study, the anemometers were known to be building mounted and the heights of these anemometers are given in Table 5.2. In a real-world MCP scenario, short-term measurements at the target site would be used with long-term historical data from the reference site. In this study however, long-term data at both the reference and target sites are used to allow the predictions to be tested against long-term measurements.

Wherever possible, reference sites were located in areas of simple, exposed terrain, or in coastal locations when paired with coastal sites. Depending on geographical location, some reference sites served multiple target sites. Clearly it is desirable to choose reference sites on the basis of the strength of their correlation to the target site under consideration, often judged by means of the linear correlation coefficient between concurrent wind speeds. However, in the current study, the choice of reference sites was generally limited by the availability of concurrent, long-term data, and the restrictions of exposed terrain and proximity to

the target sites. In certain cases this resulted in reference/target site pairings that were not ideal. For example, the rural target site R1, which is located in complex rural terrain, was paired with a coastal reference site 50 km away due to the lack of available reference data in the area. Such cases are still informative in a research context however as they allow MCP approaches to be tested at the limit of their applicability. Finding suitable reference sites for coastal target sites presented a significant challenge. As described in Section 2.4.5, due to the complexity of coastal climates, conditions can vary considerably even over short distances in the coastal zone. Due to data availability, even the closest coastal reference sites were separated from the target sites by a distance of 30 km, which presents the risk of decoupling due to differences in fetch and atmospheric stability conditions. However, such issues are representative of the challenges faced by small-scale wind developers in the real-world where access to suitable long-term reference data may be limited. This issue is returned to in Chapter 8. Details of all sites used in this work can be found in Table 5.1 and approximate geographical locations are shown in Figure 5.3.

Target sites			Reference sites					
Site	OS grid	Elev (m)	Site	OS grid	Elev (m)	$\bar{u}_{tar}/\bar{u}_{ref}$	d (km)	r_u
U1*	SJ8396	33	Rf1	SD6614	440	0.49	25	0.79
U2**	SU4210	26	Rf2	SU5501	9	0.72	16	0.87
SU1	NJ8712	65	Rf3	NO4620	10	0.94	101	0.55
SU2	SK5045	117	Rf4	TF0049	63	0.67	49	0.82
SU3	SU8554	65	Rf5	SU3039	90	0.92	58	0.85
SU4	SU1344	132	Rf5	SU3039	90	0.90	17	0.88
SU5	SU1740	126	Rf5	SU3039	90	1.13	13	0.92
SU6	SD8812	110	Rf1	SD6614	440	0.35	22	0.73
SU7	SP3180	119	Rf6	SP2186	96	0.87	12	0.81
C1	NK1345	15	Rf7	NJ2169	7	1.06	96	0.51
C2	NU2514	23	Rf3	NO4620	10	1.06	133	0.66
C3	TA1967	15	Rf8	TA0243	7	1.20	30	0.68
C4	NM8834	3	Rf9	NR6622	10	0.64	113	0.70
C5	SN2452	133	Rf10	SM8905	44	1.27	59	0.79
C6	SX9456	58	Rf11	SX4952	50	1.17	46	0.67
C8	SD3000	9	Rf12	SD3131	10	1.10	31	0.88
R1	NH8914	228	Rf13	NJ0662	5	0.59	51	0.53
R2	SE5238	8	Rf14	SE4961	14	1.13	24	0.88
R3	SK5026	43	Rf4	TF0049	63	0.67	55	0.79
R4	SO9749	35	Rf6	SP2186	96	1.03	44	0.85
R5	SU7349	118	Rf5	SU3039	90	1.14	45	0.86
R6	NS8264	277	Rf15	NT2302	236	1.68	74	0.73

Table 5.1: Summary of the meteorological monitoring sites used in this study. Target sites are defined as Urban, Sub-Urban, Coastal or Rural, reference sites are denoted as Rf. The elevation above sea level (Elev), ratio of target and reference site mean wind speeds ($\bar{u}_{tar}/\bar{u}_{ref}$), distance between target and reference sites (d) and linear correlation coefficient between the reference and target site wind speeds (r_u) are also shown. Building mounted anemometers: *H = 20.6 m, **H = 22.5 m above ground level.

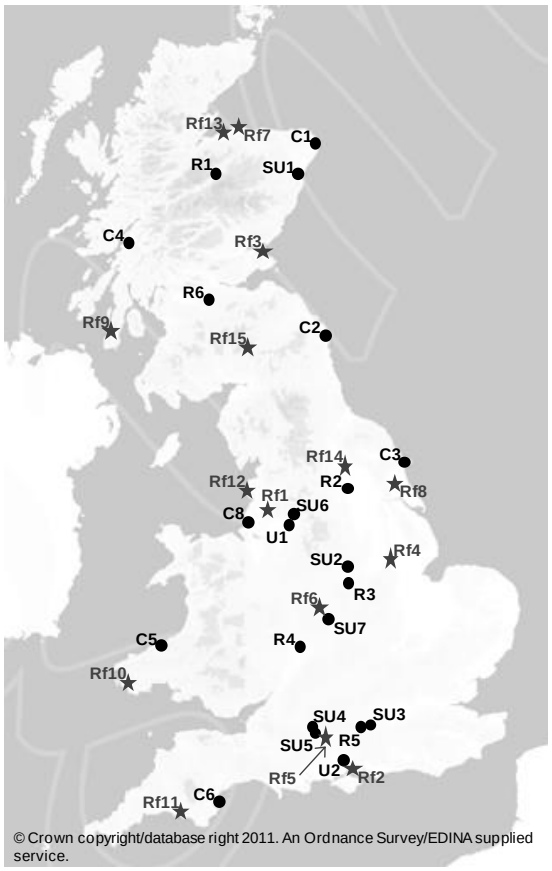

Figure 5.3: Approximate geographical locations of the meteorological monitoring sites used in this study. Target sites are defined as Urban, Sub-Urban, Coastal or Rural, reference sites are denoted as Rf.

5.2.5 Error metrics

The MCP approaches were used to predict time series of hourly averaged wind speeds over multiple 10 year prediction periods at each test site. From this time series, a number of parameters related to the wind resource were extracted. Of particular importance are the predicted mean wind speed $\bar{u}_{pred}$ and the mean Betz wind power density $\bar{p}_d$ (defined by Equation 2.9). Also of interest are parameters that give insight into the form of the predicted wind speed distribution. Two such parameters are considered in this study, the predicted Weibull shape factor k (defined by Equation 2.6 and estimated from the predicted wind speeds using the method of maximum likelihood) and the standard deviation of predicted wind speeds σ defined by:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(u_{i,pred} - \bar{u}_{pred})^2}$$

Equation 5.1

where $u_{i,pred}$ is the i^{th} wind speed prediction, $\bar{u}_{pred}$ is the long-term mean wind speed prediction and N is the total number of instantaneous wind speed predictions. The σ parameter is useful in that, unlike k which is evaluated through a fitted Weibull distribution, σ is simply a measure of the spread of the wind speeds about the mean without the assumption of a specific wind speed distribution.

Quantitative comparisons were made between predicted and observed values of these parameters across all 22 sites using the metrics of mean absolute percentage error (%Error), mean absolute error (MAE) and mean bias error (MBE). These are defined below for the predicted mean wind speed, equivalent error metrics may be defined for the remaining parameters of interest.

$$\%\text{Error} = 100 \sum_{j}^{n} \frac{\left|\bar{u}_{obs,j} - \bar{u}_{pred,j}\right|}{\bar{u}_{obs,j}} / n$$

Equation 5.2

$$\text{MAE} = \sum_{j}^{n} \left|\bar{u}_{obs,j} - \bar{u}_{pred,j}\right| / n$$

Equation 5.3

$$\text{MBE} = \sum_{j}^{n} (\bar{u}_{pred,j} - \bar{u}_{obs,j}) / n$$

Equation 5.4

where j represents the j^{th} site, $\bar{u}_{obs}$ and $\bar{u}_{pred}$ are the long-term observed and predicted mean wind speeds respectively and n is the total number of target sites.

Using the sliding window approach, 120 predicted time series, and hence 120 values of the error metrics (Equation 5.2 - Equation 5.4) corresponding to each window position, were obtained for each target site. The final error statistics for a site were calculated as the average of these error metrics across all training/test periods. These averages were then combined into a single value across all sites to provide overall error statistics.

It is important to note that the error metrics have two types of variability about these single-valued averages. Firstly, there is the variability across the different training and test periods which is linked to seasonal and inter-annual effects. The sliding window approach uses partially overlapping 3 month training periods in order to provide a large number of training/test cases as described above. This means that

the errors for each training period are not independent and hence, while the standard deviation of the error metrics across the training periods may be a useful measure of variability, care should be taken in interpreting this variability in terms of confidence intervals. Note that even for training periods that are non-overlapping, serial correlations are still likely to violate a strict independence assumption. Secondly, there is also variability in the error metrics across the 22 different target sites which is linked to site-specific performance of the MCP approaches. Both of these types of variability are considered in the following investigation.

5.3 Results and Discussion

Figure 5.4 shows hourly averaged target and reference site wind speeds for a single 30º angular sector from the reference/target site pair Rf11/C6. Since there is a large amount of variability at each site and within each angular sector, these results are intended as an example rather than to be fully representative. Wind speed observations are recorded with a resolution 0.51 ms^{-1}, hence, in order to fully represent the discretized data, the wind speeds are presented as two dimensional density plots where the shading indicates the frequency of observations at each wind speed. In the case of the predicted wind speeds, (Figure 5.4C and D), the data are also discretized using wind speed bins of width 0.51 ms^{-1} for consistency with the recorded data.

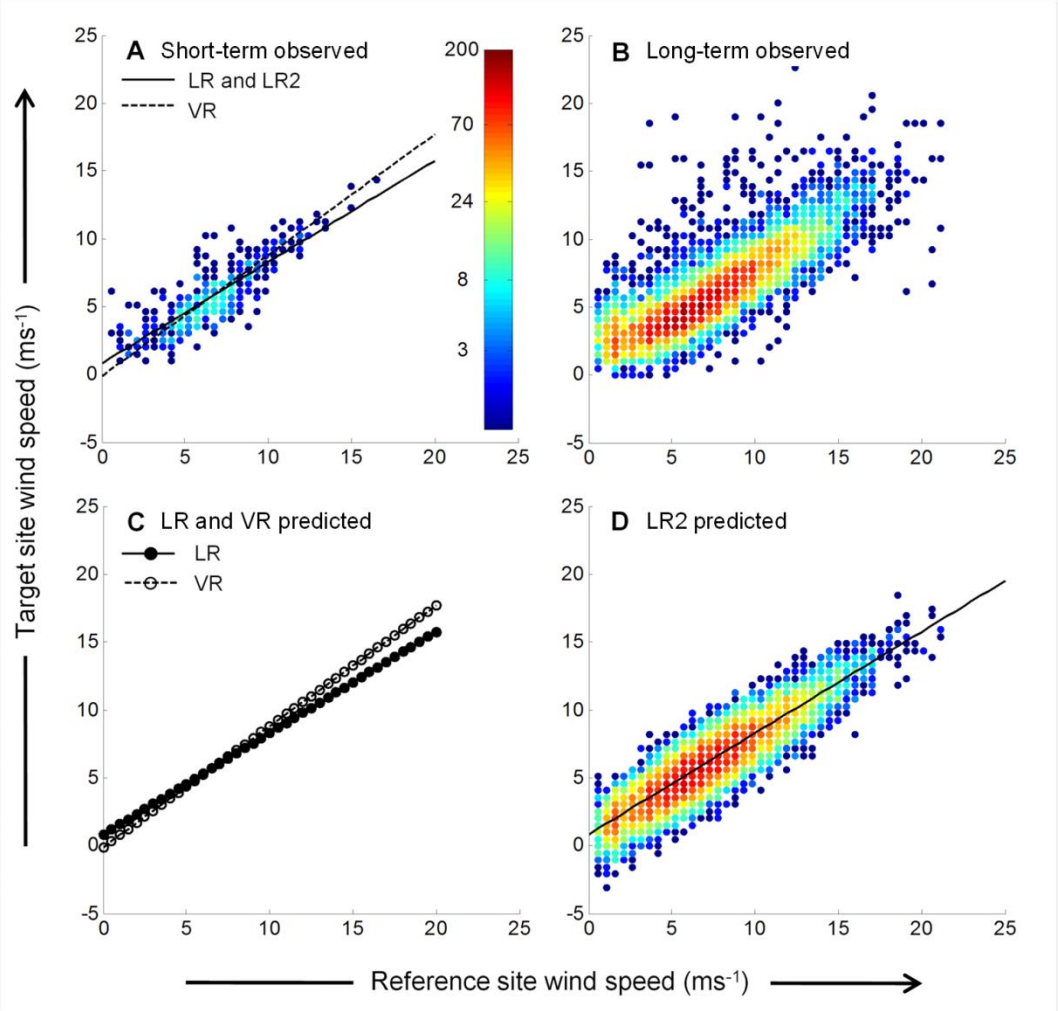

Figure 5.4: Target and reference site wind speeds for a single 30º angular sector from the reference/target site pair C6-Rf11. A) Observations over the 3 month training period along with the LR, LR2 and VR fits, B) 10 year observations, C) 10 year predictions using LR and VR, D) 10 year predictions using LR2. The solid lines represent the mean prediction, the dots show the observed or predicted scatter and the shading represents the frequency.

Figure 5.4A shows the observed wind speeds over the three month training period along with the three regression fits. The fits are identical for LR and LR2 since the approaches only differ in the prediction phase. It can be seen that VR results in a steeper gradient than LR due to the forced increase in $\sigma(\hat{u})$ discussed in Section 3.2.2. Figure 5.4B shows the recorded wind speeds over the entire 10 year prediction period and Figure 5.4C and D show the attempt to predict these observations using the MCP approaches. The predictions using LR and VR (Figure 5.4C) all lie along the straight lines defined by the regression equations. It can be seen that these predictions differ markedly from the instantaneous observations (Figure 5.4B) which exhibit considerable scatter. In contrast, the LR2 approach (Figure 5.4D) is capable of reproducing at least the general form of the scatter about the mean prediction. However, the predicted scatter has a narrower range

than that observed in Figure 5.4B, implying that σ_{res} (Equation 3.13), as modelled from the short-term training data, is lower than the long-term observed value.

It can be seen from Figure 5.4D that at low reference site wind speeds, the predicted target site wind speed may be less than zero due to the effect of the residual scatter term. Previous studies have dealt with negative predictions arising from simple linear models by removing the values [125] or setting them to zero [99]. However, where an attempt is made to model the residual scatter, negative predictions are more frequent and simply removing them will reduce the number of entries at low wind speeds, resulting in a positive bias in the predicted mean wind speed and mean wind power. The opposite will be true if the values are simply set to zero. In this work a compromise was used whereby negative predictions were set to the mean value of the function before the residual scatter term is applied. In the small number of cases where the mean value of the function is also less than zero, the value was removed from the prediction.

As discussed in Section 3.2.3, the Gaussian scatter model implemented in LR2 is a simplification. If the reference and target site wind speeds follow correlated Weibull distributions, the scatter will not strictly be Gaussian and σ_{res} is unlikely to be constant with wind speed. To demonstrate this, Figure 5.5 shows the distribution of residuals (or scatter) about the mean regression predictions for the data shown in Figure 5.4. The residuals are given by $\varepsilon = u_{tar} - \hat{u}_{tar}$, where u_{tar} and $\hat{u}_{tar}$ are the hourly averaged, observed and predicted target site wind speeds respectively.

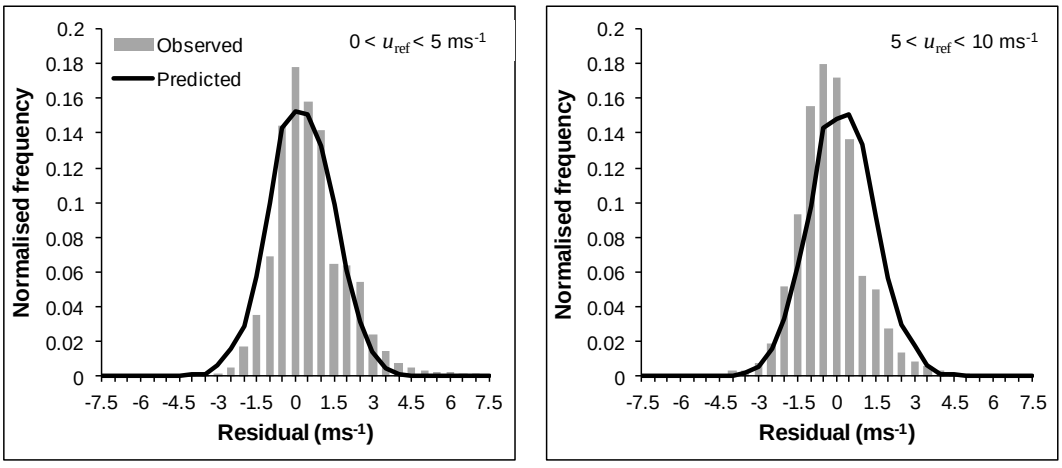

Figure 5.5: Observed (bars) and predicted (line) distributions of target site wind speed residuals for a single 30º angular sector from the reference/target site pair C6-Rf11 based on a 10 year data period. The distributions are split into a low (left) and high (right) wind speed regimes based on the reference site wind speed.

The predicted residuals are those obtained after application of the Gaussian scatter model, with variance conditioned on the short-term training period (Figure 5.4D) and hence represent samples drawn from an ideal Gaussian distribution. The observed residuals are the actual deviations from the linear regression over the full 10 year test period (Figure 5.4B) which do not necessarily follow a Gaussian distribution. The distributions have been split into a low and high wind regime with respect to the reference site wind speed to demonstrate changes in the scatter distribution with wind speed. Note that these distributions represent only a snapshot since they were taken from a single reference/target site pair using a single training period and angular sector.

From visual inspection, the distributions of observed residuals appear to deviate from ideal Gaussian behaviour since they feature both skew and increased peakedness, particularly in the high wind speed regime. This deviation from Gaussian behaviour was confirmed at the 95% confidence level based on measures of skewness and kurtosis. For the example shown in Figure 5.5, the skew is most pronounced in the higher wind speed regime which features a relatively high frequency of small negative residuals and occasional but relatively large, positive residuals. This pattern is also visible in Figure 5.4B and is likely due to localised climate effects at the target site. However, such deviations do not necessarily preclude the use of a Gaussian scatter model as a first approximation. The overall degree to which the assumption of Normally distributed residuals impacts on the accuracy of the wind resource predictions is best assessed through recourse to average error statistics over multiple sites and training periods. These statistics are considered in the following sections and alternative methods of representing the residual scatter are considered in more detail in Chapter 7.

5.3.1 The added value of MCP

Since the purpose of undertaking an MCP approach is to improve the long-term assessment of the wind resource, it is appropriate to first estimate the added value of correlating to a long-term reference site. In the absence of MCP, direct target site wind observations could be used to make some long-term estimate of $\bar{u}$ and $\bar{p}_d$. However, due to seasonal variability, it would be expected that such an estimate would be subject to considerable error when using onsite measurement periods of less than 12 months.

Figure 5.6 compares the %Error in $\bar{u}$ and $\bar{p}_d$ for direct target site observations and the LR2 approach. While the current study is predominantly concerned with a fixed 3 month onsite measurement period, Figure 5.6 compares prediction errors for

multiple training periods of between 1 and 12 months. For each training period, the error metrics are averaged across all sites and 120 sliding window positions.

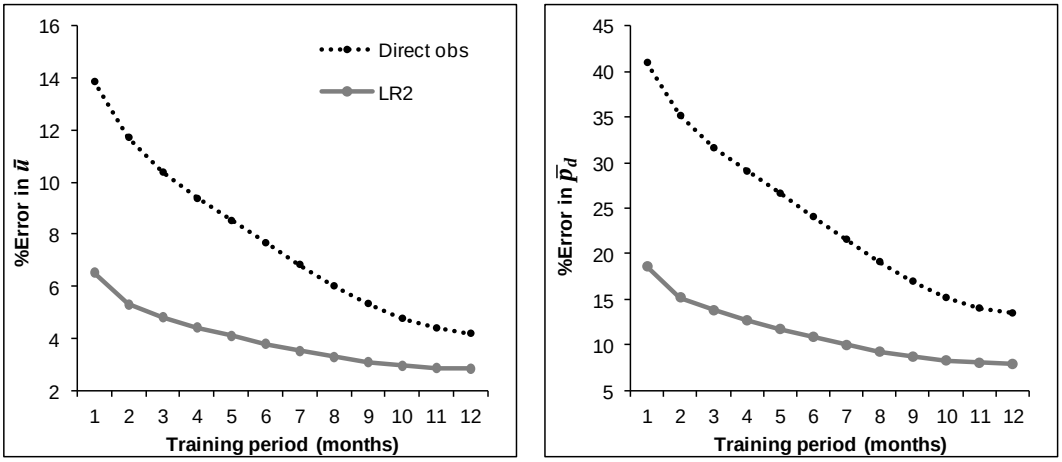

Figure 5.6: %Error as a function of training period for the wind resource parameters of $\bar{u}$ and $\bar{p}_d$ using direct observations and the linear MCP approach of LR2. The data represent the mean values averaged across 22 site pairs and 120 starting months.

Figure 5.6 demonstrates that the LR2 approach offers large reductions in error compared to direct observations for training periods of 12 months or less. The added value of MCP is particularly pronounced for short training periods where direct observations are subject to large seasonal variability, while for longer training periods, direct measurements become more representative of the long-term wind resource. These results highlight the potential value in using MCP approaches, particularly in cases where the onsite measurement period is less than 12 months.

5.3.2 Overall error statistics

To compare the performance of the regression MCP approaches implemented in this study, Figure 5.7 shows the average $\bar{u}_{pred}$ versus $\bar{u}_{obs}$ 10 year mean wind speeds at the 22 target sites using the three MCP approaches. Note these are averages over all training periods and hence only systematic biases will be visible since seasonal and inter-annual biases will be smoothed out. The long-term mean wind speeds appear to be well predicted at all sites using each of the MCP approaches. However, there appears to be a small tendency for VR to overestimate the mean wind speed compared to the other two approaches. This is not surprising since the VR method enforces a steeper gradient compared to LR. Note that the LR and LR2 approaches result in very similar values for $\bar{u}_{pred}$ indicating that the residual scatter term does not significantly affect the mean wind speed prediction.

Figure 5.7 also shows the average $\bar{u}_{pred}$ versus $\bar{u}_{obs}$ 10 year mean wind speeds at the 22 target sites using just the LR2 approach highlighted with respect to terrain type. The error bars represent plus-or-minus twice the standard deviation (+/-2σ) across the 120 training periods. The error bars highlight the intra- and inter-annual variations in the predictions, which appear to be particularly pronounced for coastal regions. As mentioned previously, the error bars cannot strictly be treated as 95% confidence intervals since the predictions are not independent due to the structure of the sliding window approach. On average, the 2σ values represent around +/- 0.5 ms^{-1} (~17%), while for the most variable sites they can be greater than 1 ms^{-1}. Note that while these 2σ values are larger than those estimated for the boundary layer scaling (BS) model in Chapter 4, (0.4 ms^{-1}, 11%) the estimated errors where shown to not fully account for all contributions to the prediction errors in the BS case.

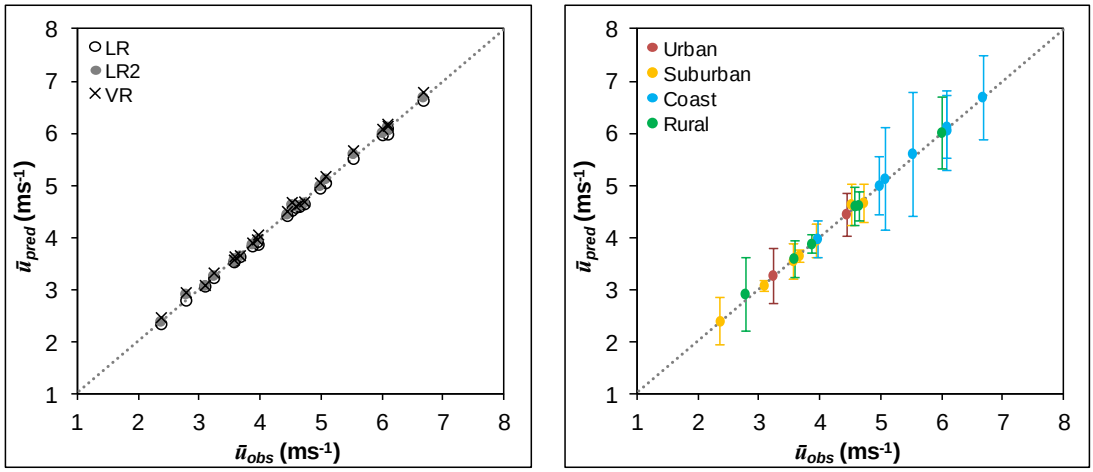

Figure 5.7: Left - Predicted and observed 10 year mean wind speeds at 22 target sites using three MCP approaches averaged across all training/test periods. Right – Equivalent predictions using only the LR2 approach highlighted by terrain type. Error bars represent +/- 2σ across the 120 training periods. The dotted line shows a one-to-one relationship.

Table 5.2 shows the error metrics of %Error, MAE and MBE averaged across all target sites for the three MCP approaches. Predictions obtained using the BS model are also included and these are discussed further in Section 5.3.5. The results indicate that on average, $\bar{u}$ and $\bar{p}_d$ can be predicted to within 4.8% and 14% respectively using just three months onsite measurements with the best performing MCP approach. Note that individual predictions can exhibit larger or smaller errors depending on the measurement season as highlighted by Figure 5.7.

For the predicted $\bar{u}$ there is no clear difference between MCP approaches while for $\bar{p}_d$, the LR2 approach performs best, closely followed by VR with the largest errors observed for LR. In the case of σ and k, VR performs best indicating that this

approach is more successful at predicting the wind speed distribution, again the largest errors are observed for LR. The bias errors indicate the degree to which the approaches systematically overestimate or underestimate a particular parameter. It is noteworthy that while LR results in very low bias in $\bar{u}$, the approach underestimates σ and overestimates k resulting in a narrower predicted wind speed distribution that in turn leads to a large underestimate of $\bar{p}_d$. LR2, which accounts for the residual scatter, results in the lowest bias for $\bar{p}_d$, followed closely by VR. Overall, LR2 tend to underestimates $\bar{p}_d$ while VR overestimates it.

3 M	**Method**	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	BS	9.5	26	NA	NA
	LR	4.7	19	17	23
	LR2	4.8	14	6.2	7.8
	VR	4.8	15	5.3	4.3
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	BS	0.42	22	NA	NA
	LR	0.21	15	0.44	0.42
	LR2	0.21	11	0.16	0.14
	VR	0.21	11	0.13	< 0.1
MBE	BS	-0.17	-9.9	NA	NA
	LR	< 0.1	-13	-0.43	0.42
	LR2	< 0.1	-2.8	-0.10	0.13
	VR	< 0.1	5.7	< 0.1	< 0.1

Table 5.2: Error metrics averaged across 22 target sites for three MCP approaches, (LR, LR2 and VR), as well as a boundary layer scaling model (BS).

In addition to the average error metrics, it is useful to consider the distribution of residual errors across all sites. The residual percentage errors $\varepsilon_{\%}$ in the predicted mean wind speed at a particular site may be expressed as:

$$\varepsilon_{\%} = 100 * \left[\frac{\bar{u}_{obs} - \bar{u}_{pred}}{\bar{u}_{obs}}\right]$$

Equation 5.5

Note that a positive $\varepsilon_{\%}$ represents an underestimate. Similar expressions may be obtained for the remaining parameters of interest.

Figure 5.8 shows $\varepsilon_{\%}$ across 22 sites. Note that the error distributions represent the range of errors across the different sites after averaging across all 3 month training periods, hence seasonal variations are smoothed out as described in Section 5.2.5. The error distributions for $\bar{u}$ are very similar for all approaches although VR shows a slight tendency to overestimate. For $\bar{p}_d$, which is perhaps the most significant parameter given that the aim is to predict the wind energy resource, LR can be seen to exhibit a strong tendency to underestimate. This is expected due to the

failure of LR to represent the residual scatter. On average, VR has a significantly lower bias than LR but the error distribution is negatively skewed indicating a tendency to overestimate. The best predictions (small bias and low error range) are obtained using LR2 highlighting the value of explicitly accounting for the residual scatter term.

These observations are also reflected in the error distributions for of σ and k. LR underestimates σ and overestimates k leading to a narrower predicted wind speed distribution. LR2 does significantly better at estimating these parameters although it still leads to an underestimate of the width of the wind speed distribution, likely because the short training period makes it challenging to accurately estimate σ_{res} (Equation 3.13). VR does particularly well at estimating σ and k resulting in a very small bias and a low error range. It is likely that the tendency of VR to slightly overestimate $\bar{u}$, which may have a more significant effect on the power density compared to k [16], prevents this approach from outperforming LR2 in terms of predicted $\bar{p}_d$.

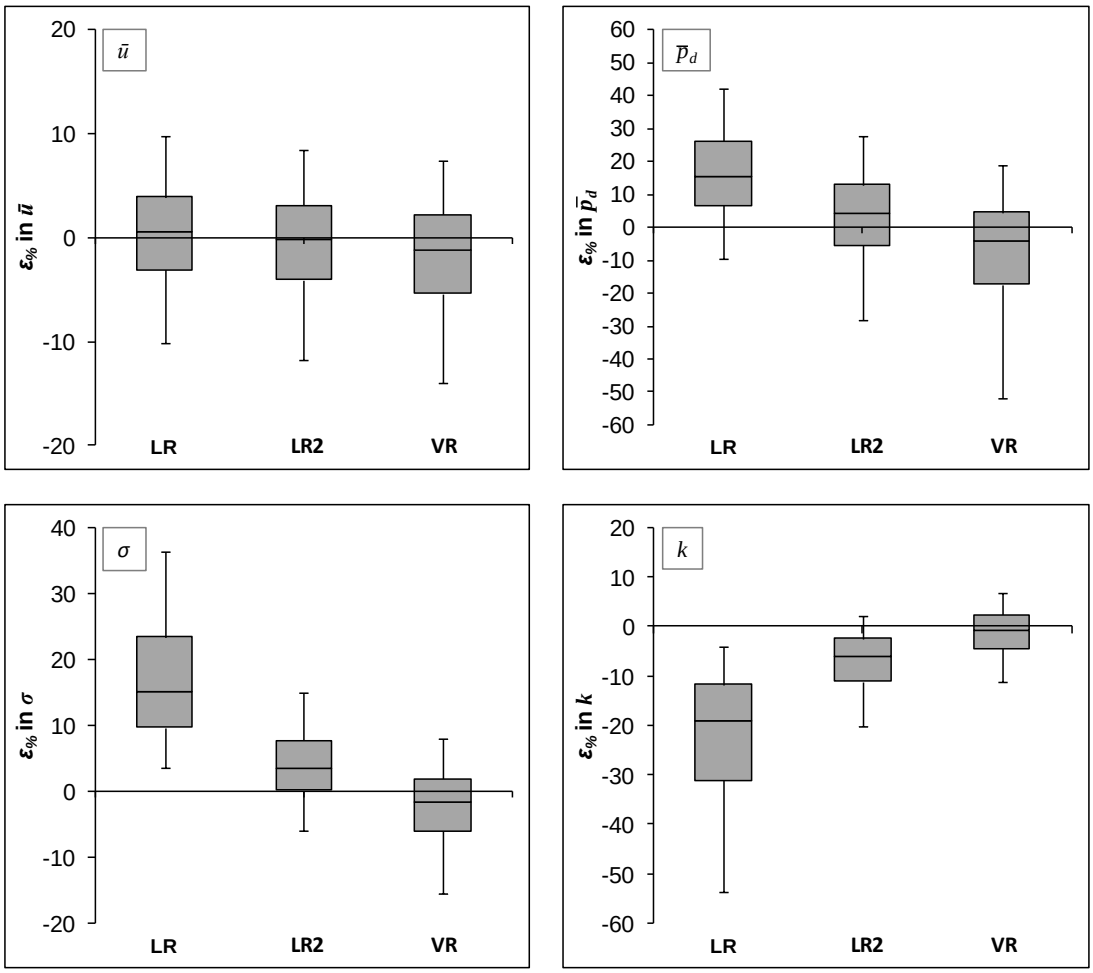

Figure 5.8: Residual percentage error distributions across 22 target sites for $\bar{u}$, $\bar{p}_d$, σ and k. The error bars represent the 5th and 95th percentiles, the shaded regions encloses the interquartile range.

5.3.3 Variability in the bias error

While Table 5.2 shows the overall MBE statistic (Equation 5.4), as averaged across all 22 sites and 120 sliding window positions, it is of interest to consider the variability in the bias error. For a fixed 3 month training period, there are a total of 2640 estimates of the bias error, representing prediction errors for 22 sites and 120 sliding window positions. Figure 5.9 shows the distributions of these errors for $\bar{u}$ and $\bar{p}_d$ using each of the three MCP approaches.

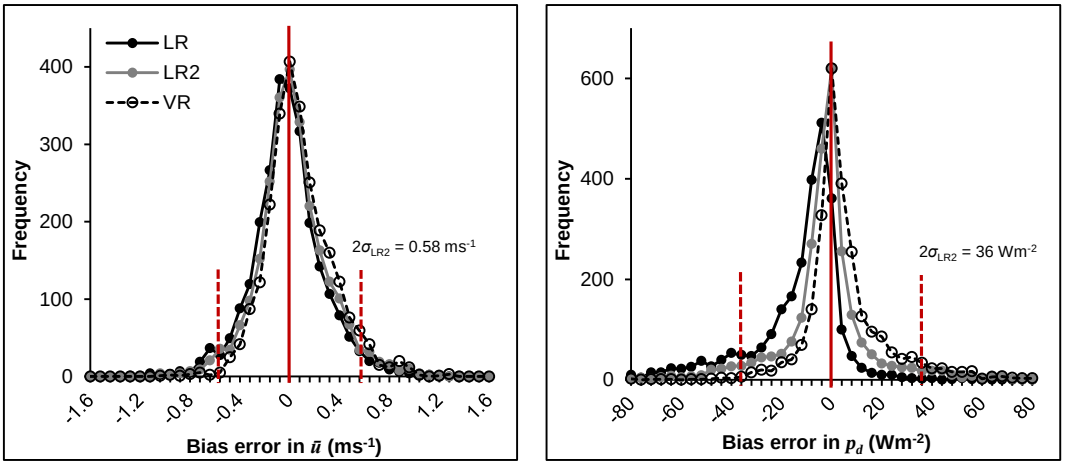

Figure 5.9: Distributions of bias errors in $\bar{u}$ and $\bar{p}_d$ for the three MCP approaches. The 2σ values for the LR2 approach are marked by the vertical, dotted red lines.

For $\bar{u}$, the bias errors are very similar for all MCP approaches and close to symmetric about zero. For $\bar{p}_d$, the LR errors are skewed in the negative direction, demonstrating the increased tendency to underestimate this parameter, while for LR2 and VR they are more symmetric about zero. Note that the distributions encompass the variability across all sites, training seasons and years. In the case of independent, normally distributed errors, the standard deviations of these distributions can be used as an estimate of the prediction uncertainty, with twice the standard deviation (+/- 2σ) interpreted as approximate 95% confidence intervals. For LR2, the 2σ values were calculated as 0.58 ms^{-1} and 36 Wm^{-2} for $\bar{u}$ and $\bar{p}_d$ respectively. Very similar values were obtained for LR and VR.

As stated previously, the overlapping training periods violate the condition of independent errors. To investigate this further, the distributions of bias errors were calculated using only the 40 non-overlapping sliding window positions. However, this was found to have little effect on the shape of the distributions and the calculated 2σ values were within 2% of those obtained from the full data. From visual inspection, there was also some concern as to whether the distributions in Figure 5.9 approximate Normal distributions, particularly in the case of the errors in $\bar{p}_d$ where the distributions appear to be strongly peaked. The kurtosis can be used as a simple measure of the peakedness of a distribution and is defined by [152]:

$$kurtosis = \frac{\mu_4}{\sigma^4}$$

Equation 5.6

where μ_4 is the fourth moment about the mean and σ is the standard deviation.

A Normal distribution has a kurtosis of 3 with higher values indicative of a more peaked distribution with heavier tails, termed leptokurtic [152]. Using the MCP approach of LR2, the kurtosis was found to be 4.6 and 7.2 for the error distributions for $\bar{u}$ and $\bar{p}_d$ respectively. This indicates that, particularly in the case of $\bar{p}_d$, the variability is not fully captured in the 2σ values and the predictions may be subject to occasional, large outliers, as characterised by a heavy tailed distribution. Thus, while MCP applied to short-term measurements can significantly aid investment decisions, the possibility of large outliers at specific sites should not be ignored.

5.3.4 Seasonal effects

Given that the MCP approaches presented in this study propose a training period of just three months, it is particularly important to consider seasonal effects in relation to the error estimates. The error metrics presented thus far have been averaged across all training periods, and while these provide robust statistics, they do not give information as to how the magnitude and sign of the errors may vary with the measurement season. Such information is important in making a more precise estimate of the likely error given a specific training season as well as in determining if prediction errors can be minimized through choosing an optimum season in which to collect the short-term onsite wind measurements.

To investigate these sensitivities, the average error statistics were decomposed into seasonal averages. This was achieved by averaging the error metrics for equivalent three month training periods across all years in the data record resulting in seasonal averages across a full 10 years. Due to the large bias present in the predictions obtained using LR, as described above, only the LR2 and VR approaches were selected for more detailed study.

5.3.4.1 Seasonal variation in the percentage error

Figure 5.10 shows the variation in the average %Error for $\bar{u}$, $\bar{p}_d$, σ, and k using different three month training periods throughout the calendar year. The vertical lines mark training periods corresponding to the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). Training data periods between these points include months which overlap more than one season. Clear seasonal variations can be seen in the prediction errors for all parameters for both LR2 and VR. For the key parameters of $\bar{u}$ and $\bar{p}_d$, the largest errors occur close to winter and summer while the smallest errors occur close to autumn and late winter/early spring (Feb-April). These observations are similar to those noted by Oliver and Zarling, [137], described in Section 3.2.5. The results indicate that on

average, large reductions in the error of the predicted $\bar{p}_d$ can be achieved through choosing optimum seasons in which to obtain the short-term onsite measurements. For LR2, the best season results in a %Error in $\bar{p}_d$ of ~10%, compared to ~18% for the worst season. Similarly for VR, the best season results in a %Error in $\bar{p}_d$ of ~12%, compared to ~20% for the worst. The slightly poorer performance of VR in winter compared to LR2, with respect to the %Error in $\bar{p}_d$, may be related to the tendency of VR to slightly overestimate the mean wind speed, as shown in Figure 5.8, since overestimates of wind speed carry a higher penalty than underestimates when evaluating wind power. The LR2 approach may also benefit from some cancellation of errors between estimates of $\bar{u}$ and the distribution parameters of σ, and k. Interestingly, the LR2 approach exhibits a stronger seasonal variation in σ and k compared to the VR approach, with the largest errors occurring in summer.

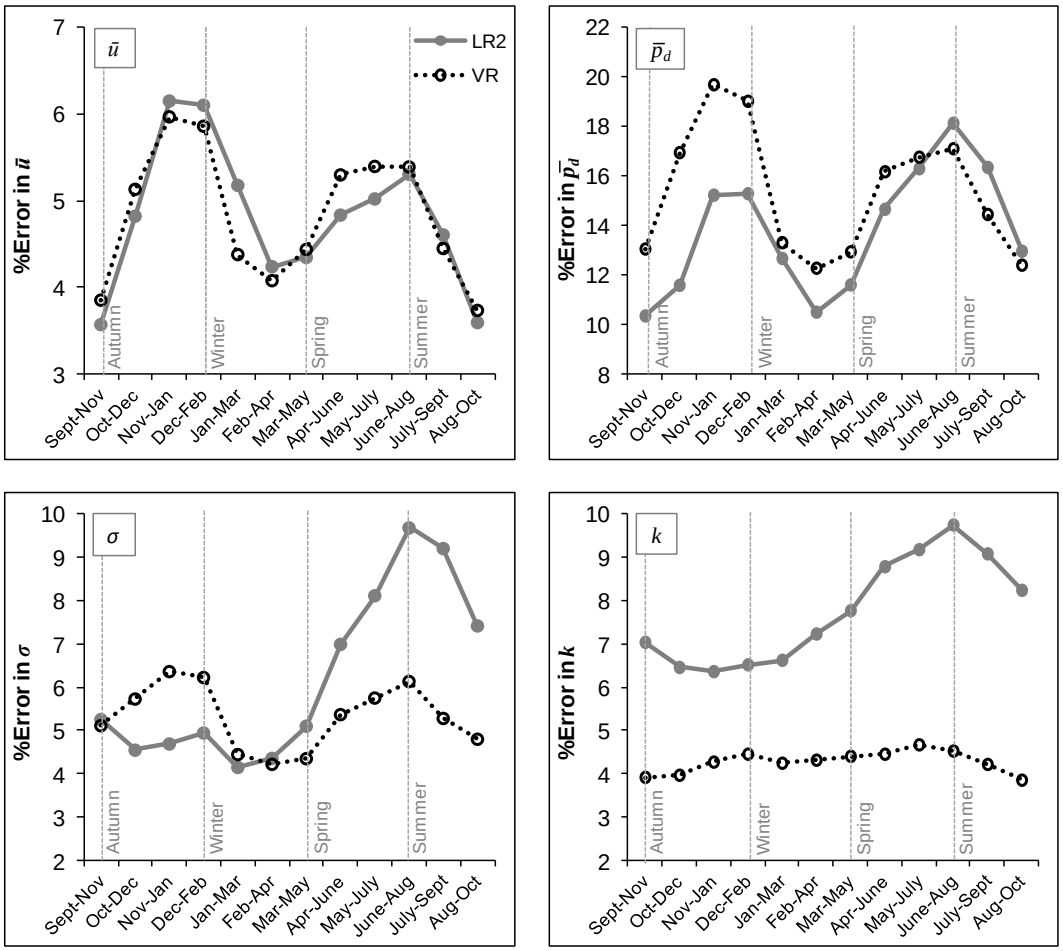

Figure 5.10: Seasonal variation of %Error in $\bar{u}$, $\bar{p}_d$, σ and k averaged across 22 target sites using two MCP approaches. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training.

The observed seasonal variations in the error metrics indicate that the regression parameters extracted from the short-term training data vary according to seasonally dependent weather patterns. If the training period features a high frequency of atypical weather patterns, the extracted regression parameters will not be representative of the long-term reference/target site relationship and this will lead to errors in the long-term wind predictions. Seasonal weather patterns could introduce a number of factors which contribute to such effects. For example, since a sector approach is used in the implementation of the MCP algorithms, it is expected that training periods that include wind directions that adequately populate each of the 30º angular sectors will result in improved long-term predictions. Of interest in this regard is a recent study by Earl *et al.* [153] which investigated the variability in UK surface winds over a 30 year period using a network of 40 anemometer stations. While the UK surface winds are dominated by winds from the southwest, relatively large seasonal variations in wind direction were observed. In particular, spring was found to have a more significant north-easterly component leading to a more even spread of wind directions compared to other seasons.

For the 15 reference sites used in the current study, Figure 5.11 compares the percentage frequency of wind directions during spring and summer as averaged across the 11 year data record. The data shows similar seasonal variations in wind direction to those observed by Earl *et al*. with more significant north-easterly and easterly components during spring compared to summer. As the MCP approaches are applied sector-wise, it is likely that seasons which have a greater spread of wind directions will result in more accurate estimates of the regression parameters since the sectors will be more uniformly populated. Conversely, seasons with a strongly dominant wind direction could result in poor estimates of the regression parameters for certain poorly represented sectors. Note that the effects of wind direction may be further complicated at coastal sites due to the presence of sea breezes which may result in localised diurnal changes in the predominant wind direction. While any such effects could in principle be reduced by using a single regression fit to data from all wind directions, it was found that on average the prediction errors were lower when using the sector approach.

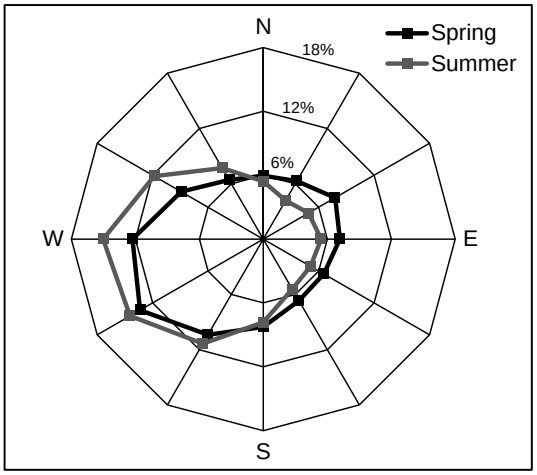

Figure 5.11: Percentage frequency of wind directions in 30º sectors during spring (Mar-May) and summer (June-Aug) averaged across the 15 reference sites and 11 year data record used in the current study. The lines are included as a guide to the eye.

Another point of interest is the increased errors in σ and k observed for LR2 during summer. For the sites considered in the current study, both the mean wind speed and the average variance of the wind speeds was lowest during this season. This could result in insufficient range over which to achieve an accurate least squares regression as well as impacting on the estimates of the long-term residual standard deviation σ_{res} required to reconstruct the residual scatter. In addition, periods of low wind speeds will be more subject to variability caused by local phenomena, as described in Section 2.4, and this may lead to decoupling between reference and target sites when they are separated by large distances.

Changes in atmospheric stability may be a further source of seasonally dependent decoupling between the reference and target sites. For example, in winter when there is a higher probability of stable conditions, a reduction in vertical mixing may cause the local wind flow to become decoupled from the large scale flow. This in turn can result in a reduced correlation between the reference and target sites, and hence, the regression parameters extracted during these periods may not be representative of the long-term relationship.

5.3.4.2 Seasonal variation in the mean bias error

To establish whether certain seasons are more likely to result on average in over or underestimates of the wind resource, Figure 5.12 shows the seasonal variation in the MBE for $\bar{u}$ and $\bar{p}_d$. For LR2, winter training periods are more likely to result in overestimates of the long-term wind resource while the opposite is true for summer training periods. For VR, the same trend is visible for winter training periods while for the remaining seasons the bias is close to zero. Since Figure 5.10 shows that

the percentage errors in $\bar{u}$ and $\bar{p}_d$ also peak for VR in the summer, this implies that for VR the sign of the error varies depending on the specific site during this season.

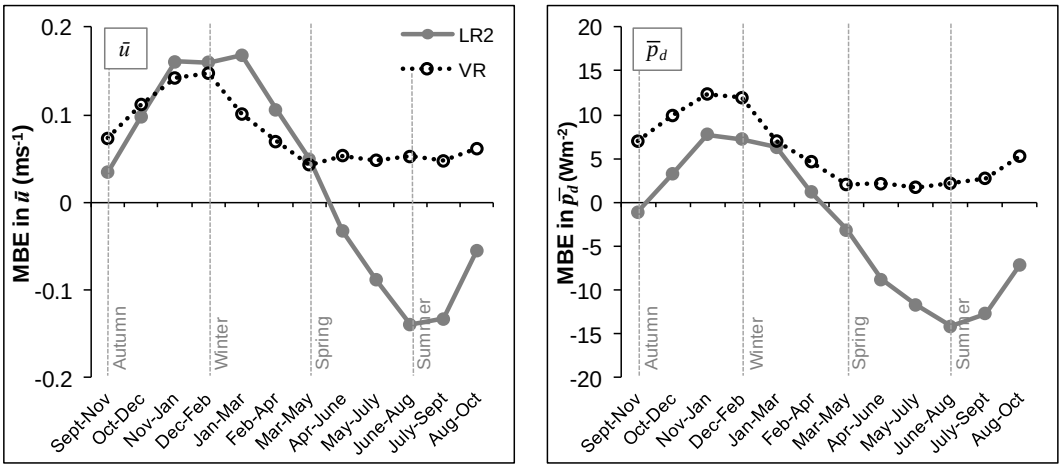

Figure 5.12: Seasonal variation of the MBE in $\bar{u}$ and $\bar{p}_d$, averaged across 22 target sites using two MCP approaches. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training.

5.3.4.3 Seasonal gains in prediction accuracy

It is of interest to examine the potential gains in prediction accuracy that may be achieved using the optimum season for onsite measurements on a site-by-site basis. Figure 5.13 compares the LR2 %Error in $\bar{p}_d$ for each of the 22 target sites using three different 3 month training periods chosen to represent the 'worst' (Nov-Jan or June-Aug) and 'best' (Feb-Apr) training periods as identified from Figure 5.10. The errors represent values averaged over the full data record. Several observations can be made from this analysis. Firstly, in all but two cases, the Feb-Apr training period results in lower errors than the summer June-Aug training period. Hence, it appears that the summer frequently features atypical coupling between the reference/target sites leading to erroneous long-term predictions. Secondly, there is also a strong preference for the Feb-Apr training period over the early winter Nov-Jan period, although in this case the results are more variable with six sites not conforming to this trend. However, perhaps the strongest trend emerging from Figure 5.13 is the dominance of coastal sites in the seasonal variability in the %Error. At some coastal sites, the %Error metric is on average three to four times greater when using the worst training season compared to the best.

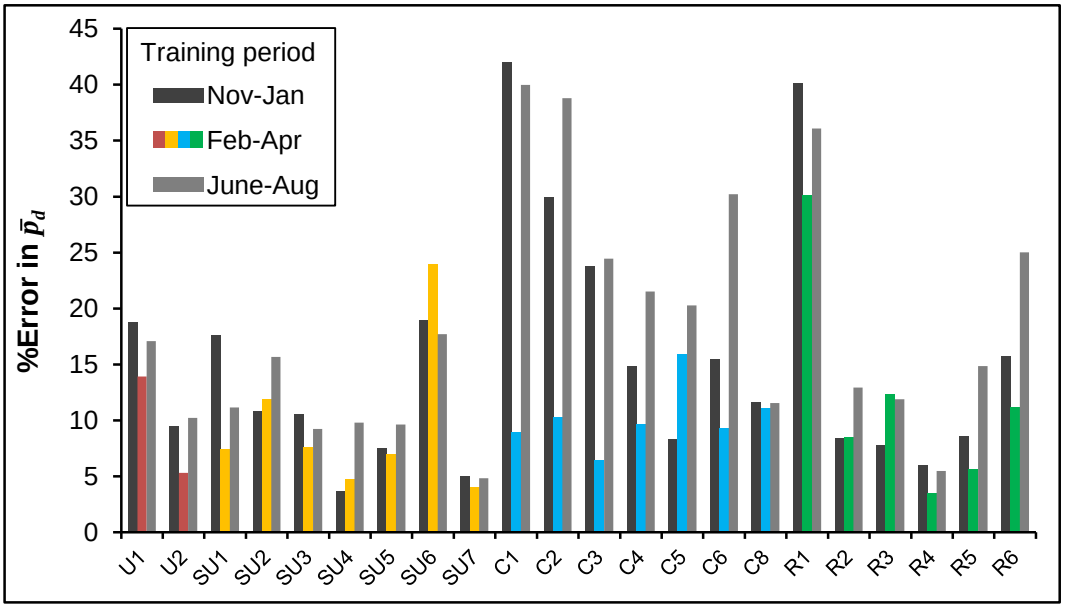

Figure 5.13: %Error metrics for each of the 22 target sites using the 'worst' (Nov-Jan or June-Aug) and 'best' (Feb-Apr) 3 month training periods based on the LR2 MCP approach. The results are averaged for these training periods across the entire data record.

Robust conclusions regarding the seasonal effects for each terrain type would require a greater number of sites to be considered, however, the strong seasonal effects observed at coastal sites are not surprising. As discussed in Section 2.4.5, coastal locations may be subject to complex atmospheric stability effects, which exhibit strong seasonal variability [70]. Seasonal weather patterns can result in rapid heating or cooling of the land surface, and these abrupt changes will not occur in the sea surface temperature. This may result in different stability conditions onshore and offshore, and hence, the coupling between reference and target sites may become particularly sensitive to the distance from the shoreline, as well as the presence of sea or land in the fetch. Since the coastal target sites used in this study were all located within 1 km of the shoreline, these effects are expected to be particularly pronounced. Thermally driven winds with diurnal cycles are likely to be a further source of seasonal dependent coupling between reference and target sites, since these may affect both the wind speed and wind direction on a very local scale. Due to the relatively large separation between the coastal reference and target sites, (greater than 30 km), these localised effects are likely to be a significant source of error. These issues are returned to in Chapter 8 where alternative sources of reference data are considered that may be of particular value in regions where the climate is highly localised.

5.3.4.4 Summary of seasonal effects

The above results indicate that on average, relatively large improvements in the accuracy of the predicted long-term wind resource, as estimated from short-term measurements, can be made through choosing optimum seasons in which to obtain onsite measurements. In addition, the results give an indication as to the average sign of the bias error as a function of the season used for training and the MCP approach used as well as indicating that these effects may be largest at coastal sites. While such results are potentially of great use in predicting the long-term wind resource based on very short measurement periods, the following factors should be borne in mind:

(i) These results are average statistics across a number of sites and years and hence individual predictions in any given year at a particular site may deviate from these trends.

(ii) The trends may vary considerably for non-UK sites that experience different climatic conditions.

(iii) Since the sample of 22 sites considered in this study is relatively small it may not be large enough to be fully representative of UK sites as a whole.

5.3.5 Comparison between modelling and data-driven approaches

Given that obtaining onsite wind speed measurements, even for a short time period, necessitates additional time and expense, it is useful to investigate to what extent a data-driven approach, based on a very short measurement period, may improve predictions of the wind resource compared to a modelling approach. The target sites used in this work formed a subset of the sites considered in the evaluation of the boundary layer scaling model (BS) in Chapter 4. Hence, it is possible to make a direct comparison between the accuracy of the two approaches for these sites. The average error metrics are compared in Table 5.2 for BS and the three MCP approaches. The BS predictions were obtained using the improvements suggested in Chapter 4 and reference [16] which included taking account of the angular dependent upwind roughness. Note that the error statistics reported here for the BS approach applied to 22 sites are somewhat smaller than those reported for the full 38 sites in Chapter 4. This is likely because the subset of sites used in the current chapter was selected from the Met Office anemometer network, and such anemometers are generally carefully sited so as to avoid local sheltering effects. Table 5.2 shows that despite the very short training period, all the MCP approaches

result in a clear improvement in all of the average error metrics. For example, using LR2 compared to BS reduces the average percentage error in the predicted $\bar{p}_d$ from 26% to 14% and halves the MAE from 22 Wm^{-2} to 11 Wm^{-2}.

It should be noted that modelling approaches are still of significant value in that they can be easily implemented in a scoping context with little prior investment. However, the results presented here indicate that the additional time and investment required for short-term onsite measurements accompanied by MCP analysis is well justified in cases where investors require greater confidence in the predicted wind resource, even when these measurements only cover a period of months.

5.4 Conclusions

The feasibility of predicting the long-term wind resource over 10 years at 22 UK sites using MCP approaches based on short-term onsite measurements covering just three months has been investigated. Using a sliding window approach over an 11 year data period, robust error statistics have been obtained which account for both inter-annual and seasonal variations. This work extends previous studies related to MCP with very short training periods by (i) considering a large number of sites in a variety of terrains, (ii) using multiple training and test periods over a long-term data record, (iii) investigating seasonal effects in detail and (iv) explicitly considering residual scatter. The results indicate that while short measurement periods introduce additional challenges including seasonal variations and reduced data coverage, the approach can be a valuable tool for wind resource assessment in the small-scale wind energy industry.

Three regression approaches, LR, LR2 and VR were compared, and it was found that all approaches were able to successfully predict the mean wind speed. However, due to the failure of LR to take account of the residual scatter, the predictions of wind power density showed significant bias when using this approach. On average, LR2 resulted in wind power predictions with the lowest bias and percentage error, closely followed by VR. The LR2 approach tends to slightly underestimate wind power while VR tends to overestimate it. VR was on average most successful at predicting parameters related to the wind speed distribution, σ and k, closely followed by the LR2 approach, while LR again resulted in large biases.

Analysis of the sensitivity of the wind resource predictions to the season in which the onsite wind speed measurements were obtained revealed clear seasonal

variations in both the sign and the magnitude of the prediction errors. The results indicate that on average in the UK, the lowest prediction errors are obtained when using either autumn or early spring as the training period, while the highest errors are obtained when using winter or summer. Seasonal effects appear to be larger for coastal sites compared to the other terrains considered. For a three month training period, choosing the optimum measurement season can result in an average improvement of 8 percentage points in the predicted wind power density compared to the worst season.

Comparison between the MCP approaches presented in this work and a previously developed semi-empirical model demonstrate that large improvements can be made in predicting the long-term wind resource using the MCP approaches even with just three months onsite wind speed measurements. Across 22 UK sites, the best performing MCP approach resulted in mean absolute and percentage errors of 4.8% and 0.21 ms^{-1} respectively for $\bar{u}$ and 14% and 11 Wm^{-2} for $\bar{p}_d$. By way of comparison, the modelling approach resulted in errors of 9.5% and 0.42 ms^{-1} for $\bar{u}$ and 26% and 22 Wm^{-2} for $\bar{p}_d$. It should be borne in mind, however, that these errors represent averages over multiple sites and training seasons and may vary considerably given a specific site and training period. This variability can be informed by considering the standard deviation of the bias errors, although care must to be taken in interpreting this parameter, as discussed in Section 5.3.3.

The following chapter will consider whether the data-driven techniques developed above can be extended through the use of non-linear MCP algorithms that allow for greater flexibility in describing the correlation between reference/target site pairs.

6 A Gaussian Process Regression Approach to Wind Resource Assessment

6.1 Overview

The results presented in Chapter 5 indicate that on average, MCP approaches based on linear functions are capable of producing reasonable wind resource estimates using onsite measurement periods as short as three months. However, it is not clear from this work whether linear functions are likely to provide the best representation of the relationship between the reference and target site wind speeds. Indeed it is possible to show that based on the assumption that wind speeds follow Weibull distributions at the target and reference sites, the correlation between the two sites will theoretically be non-linear [110, 116]. However, there is no guarantee that this non-linearity will emerge in practice from short-term, noisy measurements. A further question is whether probabilistic inference can improve on the least squares regression estimates when working with short-term, highly variable wind data. While this is a non-trivial question from a theoretical perspective, practical application of the two approaches may give some insights, at least with respect to their use in MCP.

Gaussian process regression (GPR) is an alternative regression approach which does not restrain the relationship between the target and reference site wind speeds to be any specific functional form. In addition, since it is a Bayesian learning approach, the input/output mapping is estimated probabilistically and the resulting predictions are distributions with a specified mean and variance rather than point values. Note the similarity here with the LR2 approach described in Chapter 5 where a distribution on the point predictions is constructed by computing the residual variance.

In the current chapter, GPR was applied in an MCP context to obtain long-term wind resource predictions from short-term onsite measurements. Specifically, the wind resource prediction errors using GPR were compared with those obtained using the linear approach detailed in Chapter 5 in order to establish whether GPR is likely to result in improvements. To facilitate this comparison, the MCP implementation strategy described in Chapter 5 was also used in the current chapter. However, the key difference was the methodology used to establish the relationship between the reference and target site wind data, namely the GPR

approach. Additionally, the GPR framework was applied in the more challenging context of orthogonal regression where the regression is applied separately to concurrent orthogonal wind vectors at the reference/target site pairs. While such correlations are more likely to be non-linear, the technique is attractive in that it is capable of also predicting the distribution of target site wind directions.

The contribution of the work described in the current chapter is the application of a new technique in the field of MCP (the GP framework) that to the best of our knowledge has not previously been applied in this context. Additionally, previous work related to orthogonal regression is extended through the use of flexible, non-linear functions in describing the correlation between reference and target site orthogonal wind vectors.

The main objectives of this study can be summarised as described below:

(i) Development of an MCP approach based on GPR and its application to predict the long-term wind resource at 22 sites.
(ii) Comparison of the success of the GPR and linear MCP approaches using a range of error metrics.
(iii) Investigation of the non-linearity of the GPR model across different sites.
(iv) Extension of the GPR framework to orthogonal regression in order to predict the distribution of target site wind angles.

The chapter is organised as follows: Firstly, a broad overview of the GPR methodology is given and this is followed by a more detailed mathematical description. The stages of implementation of GPR are then outlined and details of its application in two MCP approaches are given. The results are presented in two parts with the first detailing the performance of GPR as applied to regression of the scalar wind speeds and the second detailing the performance when applied to orthogonal regression. Finally, the overall conclusions and opportunities for further work are outlined.

6.2 Methodology

6.2.1 Introductory remarks

Conceptually, GPR differs somewhat from conventional linear regression approaches, hence, the aim of this section is to provide some intuition as to how GPR operates. While the field of Gaussian Processes is highly specialised, the present study is concerned with the practical application of GPR in the relatively simple one-dimensional case. Since there are well established procedures for

achieving this, the focus of this work is on the real-world performance of a simple implementation of GPR as applied to MCP. The results presented here should be seen as a first step in the application of GPR to MCP with the recognition that there is scope for expansion of this study.

In a conventional, one-dimensional regression problem, the aim is to find a functional relationship between some independent input variable $\mathbf{x} = [x_1, \ldots, x_n]$ and a dependent output variable $\mathbf{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$. A fixed function $f(\mathbf{x})$ is assumed to exist which, along with observational noise $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, describes the mapping of the observations $\mathbf{y}$ onto $\mathbf{x}$ by means of the expression [128]:

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon$$

Equation 6.1

For a linear model it is assumed that the function is fixed and may be parameterised by appropriate weights $\mathbf{w}$ giving [133]:

$$f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{w}$$

Equation 6.2

The weights may be determined by maximising the likelihood function with respect to $\mathbf{w}$, namely the probability density $p(\boldsymbol{y}|\mathbf{x}, \mathbf{w})$ which represents the probability of observing $\mathbf{y}$ given the data $\mathbf{x}$ and weights $\mathbf{w}$. For conventional linear regression, maximising this likelihood can be shown to be equivalent to the least squares fitting procedure.

In a Bayesian formulation, a prior distribution must first be defined over the weights and this prior is combined with the likelihood to give the posterior distribution according to Bayes theorem [133]:

$$p(\mathbf{w}|\mathbf{y}, \mathbf{x}) = \frac{p(\mathbf{y}|\mathbf{x}, \mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|\mathbf{x})}$$

Equation 6.3

where $p(\mathbf{w})$ is the prior probability density over the weights that encodes our prior assumptions, $p(\mathbf{y}|\mathbf{x}, \mathbf{w})$ is the likelihood, $p(\mathbf{y}|\mathbf{x})$ is a normalising factor and $p(\mathbf{w}|\mathbf{y}, \mathbf{x})$ is the posterior probability density over the weights given the data.

Note that in the Bayesian approach, the weights are not single-valued but are represented by the posterior probability density with some mean and variance. This is in contrast to conventional regression where we attempt to find the ‘true’ value of these parameters. Since the parameters are represented by a probability distribution, predictions at test points using such a model are obtained by

integrating over the posterior parameter distribution. Hence these predictions will also be represented by a distribution with the mean and variance reflecting the uncertainty in the weights.

The model described by Equation 6.2 is linear regardless of whether a maximum likelihood or a Bayesian formulation is used to obtain the weights. Using the idea of a 'feature space', the data maybe projected onto some new space using basis functions which allows non-linear relationships to be represented linearly in the feature space. Using this approach, a much wider variety of cases may be treated providing the basis functions themselves are independent of $\mathbf{w}$ [133]. However, the limitation of these approaches is that in general, some fixed functional form must be applied *a priori* thus restricting the range of possible outputs. In contrast, GPR does not restrict the regression to a fixed functional form. Instead, as shown by Rasmussen and Williams [133], the GPR can be thought of as a Bayesian linear regression model with an infinite number of basis functions thus offering much greater flexibility.

GPR does however, involve somewhat of a conceptual leap from conventional linear regression. In the Gaussian Process (GP) framework, the set of observations, $\mathbf{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$, are considered to be a sample drawn from an n-dimensional multivariate Gaussian distribution. This is, in effect, what is meant by the formal definition given in Section 3.2.4. Points within this n-dimensional space are related to each other through a covariance function, in fact, a GP is completely specified by a mean function and a covariance function [133]. Applying this framework, it is possible to develop a GPR approach based on the mathematical properties of multivariate Gaussian distributions. GPR is considered a Bayesian method, since first, a prior mean function and covariance function are specified that encode one's prior assumptions about the form of the regression. The prior is then conditioned on the observed data to provide a posterior distribution over functions and this is used for prediction. The following section presents the mathematical framework that is used to implement this approach.

6.2.2 Mathematical framework

Using a GP approach, a prior distribution over functions is first expressed using a mean function and a covariance function which encode prior assumptions regarding the general form of the relationship between the target and reference sites. Given a set of concurrent wind speed observations at the reference and target sites (the training data) this information may be incorporated by updating the prior. The new distribution is the posterior and is used as the basis for prediction of the target site

wind speeds given new wind speed observations at the reference site. Note that while the posterior is a distribution over functions, it can always be evaluated for a specific input value to obtain a prediction of the mean and variance at that point.

In developing a mathematical description of this process it is useful to make the following distinctions:

(i) The training data are defined as the short-term concurrent observations of wind speed at the target and reference sites. Short-term observations at the reference site are designated as the training inputs $\mathbf{x} = [x_1, \dots, x_n]$, while concurrent observations at the target site are designated as the training outputs $\mathbf{y} = [y_1, \dots, y_n]^{\mathrm{T}}$ where n is the number of training observations. The observed outputs $\mathbf{y}$ are considered to be noisy realizations of an underlying but unspecified function $f(\mathbf{x})$.
(ii) The test inputs are defined as the long-term historical observations of wind speed at the reference site. These are used as the basis for predicting the long-term time series of wind speeds at the target site and are represented by $\mathbf{x}_* = [x_{*1}, \dots, x_{*n_*}]$ where n_* is the number of test inputs.
(iii) Finally, the predicted function values at the location of the test inputs are defined as $\mathbf{f}_*$. These function values represent the predicted wind speeds at the target site given the test inputs $\mathbf{x}_*$. Since the prediction represents a distribution over functions, the mean function is denoted as $\bar{\mathbf{f}}_*$ and the covariance as $cov(\mathbf{f}_*)$.

Using these distinctions, the GP can be expressed as [133]:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right)$$

Equation 6.4

where $m(\mathbf{x})$ and $k\left(\mathbf{x}, \mathbf{x}'\right)$ represent the mean function and covariance function respectively. This notation simply expresses the existence of some function $f(\mathbf{x})$ in a space defined by a multivariate Gaussian of infinite dimensions, as described by a mean function and a covariance function.

In this work an affine mean function was used which encodes the prior assumption that the relationship between the reference and target site wind speeds is approximately linear. Note that the mean function is simply a starting point and does not exclude non-linear functions from the GP posterior. Similarly, the prior assumption that the covariance between data points will decrease with increasing

distance in the input space was encoded by defining a squared exponential covariance function. Here it should be noted that a wide range of covariance functions are possible, subject to certain restrictions, see for example Rasmussen and Williams [133], as well as Mackay [154]. If specific prior information is available regarding the expected relationship between the variables, such as periodicity or a decaying periodic trend, several covariance functions may be combined or a function may be tailor-made to incorporate this information [127]. Since such prior information is not available in the current application, a smoothly varying covariance function which decreases with increasing distance in the input space is more appropriate. The squared exponential is considered a suitable choice in this regard since it fulfils these requirements, is simple to implement and has been widely used in a number of regression applications [128, 133].

Assuming a one-dimensional input space, namely the reference site wind speeds, the covariance function may be expressed as [133]:

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left[-\frac{1}{2l^2}(x_p - x_q)^2\right] + \sigma_s^2 \delta_{pq}$$

Equation 6.5

where x_p and x_q denote individual training inputs, σ_f^2 is the variance of the underlying function $f(\mathbf{x})$, σ_s^2 is the variance of the noisy training outputs (where the training outputs are considered noisy realisations of the underlying function $f(\mathbf{x})$), l represents a length scale and δ is the Kronecker delta function. The variables σ_f^2, σ_s^2 and l along with the gradient and intercept of the affine mean function make up the *hyperparameters* of the GP. Optimal values for the hyperparameters are inferred from the training data by maximizing the log marginal likelihood. For a specific training data set, the likelihood can be expressed explicitly in terms of the training outputs, the covariance matrix and the mean function, allowing the likelihood to be evaluated for different values of the hyperparameters [128]. This optimization of the hyperparameters is referred to as the training phase of the GPR.

Given a finite set of training inputs, namely the short-term wind speed observations at the reference site, the GP reduces to a multivariate Gaussian of finite dimensions. This distribution is defined by the mean $m(\mathbf{x}) = [m(x_1) \ldots, m(x_n)]^{\mathrm{T}}$ and covariance matrix $\mathbf{K}(\mathbf{x}, \mathbf{x})$, evaluated at these finitely many points, namely [133]:

$$\mathbf{y}(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}) + \sigma_s^2 \mathbf{I})$$

Equation 6.6

where $\mathbf{I}$ is the identity matrix.

Since the aim of GPR is to make predictions at new test points based on previous observations from the training data, the joint distribution of the observed training outputs $\mathbf{y}$ and predicted function values $\mathbf{f}_*$ at the test inputs $\mathbf{x}_*$ is expressed in the following partitioned form [133]:

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\left(\begin{matrix}m(\mathbf{x})\\ m(\mathbf{x}_*)\end{matrix}, \begin{bmatrix}\mathbf{K}(\mathbf{x},\mathbf{x}) + \sigma_S^2\mathbf{I} & \mathbf{K}(\mathbf{x},\mathbf{x}_*)\\ \mathbf{K}(\mathbf{x}_*,\mathbf{x}) & \mathbf{K}(\mathbf{x}_*,\mathbf{x}_*)\end{bmatrix}\right)$$

Equation 6.7

where the matrix $\mathbf{K}(\mathbf{x},\mathbf{x}_*)$ represents the covariance matrix between the training and test inputs and $\mathbf{K}(\mathbf{x}_*,\mathbf{x}_*)$ represents the covariance matrix between the test inputs. Finally, the posterior distribution conditioned on the observed training outputs $\mathbf{y}$ may be expressed as the multivariate Gaussian [133]:

$$\mathbf{f}_*|\mathbf{x},\mathbf{y},\mathbf{x}_* \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, cov(\mathbf{f}_*)\right)$$

Equation 6.8

It is now possible to obtain the key predictive outputs from GPR. Although not proved here, standard theorems of multivariate Gaussians allow the mean and covariance of the posterior distribution conditioned on the training data to be written out explicitly. The predicted mean function values $\bar{\mathbf{f}}_*$ at the test inputs $\mathbf{x}_*$ can be expressed as [133]:

$$\bar{\mathbf{f}}_* = m(\mathbf{x}_*) + \mathbf{K}(\mathbf{x}_*,\mathbf{x})[\mathbf{K}(\mathbf{x},\mathbf{x}) + \sigma_S^2\mathbf{I}]^{-1}\left(\mathbf{y} - m(\mathbf{x})\right)$$

Equation 6.9

and the predicted covariance at the test inputs, $cov(\mathbf{f}_*)$ as [133]:

$$cov(\mathbf{f}_*) = \mathbf{K}(\mathbf{x}_*,\mathbf{x}_*) - \mathbf{K}(\mathbf{x}_*,\mathbf{x})[\mathbf{K}(\mathbf{x},\mathbf{x}) + \sigma_S^2\mathbf{I}]^{-1}\mathbf{K}(\mathbf{x},\mathbf{x}_*) + \sigma_S^2\mathbf{I}$$

Equation 6.10

For the present application, the vector of mean function values $\bar{\mathbf{f}}_*$, represents the predicted wind speeds at the target site given the reference site wind speed observations $\mathbf{x}_*$. Note that these predictions are conditioned on the short-term training data represented by the vectors $\mathbf{y}$ and $\mathbf{x}$. Similarly, the variability in the predictions is represented by the $cov(\mathbf{f}_*)$ matrix.

From the properties of the squared exponential covariance function described above, the term $\mathbf{K}(\mathbf{x}_*,\mathbf{x})$ in Equation 6.9 will tend towards zero for test points $\mathbf{x}_*$ that

are far from the training points $\mathbf{x}$. Note that what constitutes 'far' is controlled by the length scale l, which is one of the covariance function hyperparameters in Equation 6.5. In such cases, $\bar{\mathbf{f}}_*$ will tend towards the values $m(\mathbf{x}_*)$ described by the prior mean function. In practice, this means that for reference site wind speeds that occur during the long-term test period that are significantly different from those observed during the short-term training period, the GPR predictions at the target site will be close to those predicted by linear regression since the prior mean function was defined to be linear. However, in the vicinity of the training data, the GPR predictions will reflect any non-linearity inferred from the training data. This ability to reflect non-linearity inferred directly from the observations is a key advantage of the GPR approach compared to conventional linear regression.

The pointwise variances of the mean predictions are simply obtained from the diagonal elements of the $cov(\mathbf{f}_*)$ matrix. Note from Equation 6.10 that $cov(\mathbf{f}_*)$ consists of the prior covariance between the test points $\mathbf{K}(\mathbf{x}_*, \mathbf{x}_*)$, from which is subtracted a positive term related to the additional information contained within the training data [133]. Finally a noise term is added $\sigma_s^2\mathbf{I}$, in the case of noisy observations. As with the mean prediction $\bar{\mathbf{f}}_*$, for points far from the training observations the term $\mathbf{K}(\mathbf{x}_*, \mathbf{x})$ will tend towards zero and $cov(\mathbf{f}_*)$ will reduce to the prior covariance $\mathbf{K}(\mathbf{x}_*, \mathbf{x}_*)$ plus the noise term. Hence the σ_f^2 hyperparameter defined in Equation 6.5 can be seen to represent the function variance at test points far from the training data. For test points close to the training observations, the variance is reduced reflecting the additional information provided by the training data. In the current application however, the constant noise term σ_s^2 is expected to dominate the pointwise variances due to the stochastic nature of the wind speed observations.

6.2.3 Implementation of Gaussian process regression using the GPML toolbox

In this work, the GPR technique was embedded within an MCP approach to obtain long-term predicted time series of wind speeds and associated statistics at the target sites given short-term training data and long-term observations at the reference sites.

In order to implement GPR, the GPML (Gaussian Processes for Machine Learning) MATLAB Toolbox developed by Rasmussen and Nickisch [155] was used. The toolbox facilitates the implementation of GPR as described in Section 6.2.2, allowing the user to specify the required covariance and mean functions as well as

the inference method and likelihood function. The implementation procedure is outlined below:

6.2.3.1 Function specification and initialisation of hyperparameters

The first step in the implementation process is the specification of the mean and covariance functions. As outlined above, an affine mean function and squared exponential covariance function were used in the present study. The affine mean function is specified by a linear function plus a constant. The hyperparameters must also be initialised to provide starting values for the optimisation using maximum likelihood. The affine function requires the hyperparameters of gradient and y-axis intercept and these were both initialised at zero. These could equally be initialised using a linear fit to the data, although the initialisation values were found to have little impact on the optimised hyperparameters. The covariance function requires a length scale l, the function variance σ_f^2 and noise variance σ_s^2 to be specified in terms of their natural logarithm. All three of these hyperparameters were initialised at unity (i.e. at zero in terms of their logarithm). This simply provides a neutral, consistent starting point from which to optimise the parameters.

6.2.3.2 Learning the hyperparameters

If the hyperparameters were known *a priori*, the GPR could simply be implemented using the specified mean and covariance functions. However, a key step in GPR is to establish values of the hyperparameters that best represent the data under consideration; this is known as the training phase. The GPML toolbox is capable of dealing with both regression and more complex classification problems and hence includes a rich array of likelihood functions and inference methods. However, in the present application, only a Gaussian likelihood function is required and this allows exact (i.e. analytically tractable) inference. In the GPML, the hyperparameters are optimised given the training data by a call to minimize the negative log marginal likelihood (maximise the likelihood) using exact inference and a Gaussian likelihood function.

6.2.3.3 Prediction

The final stage in implementing the GPR is the prediction phase. Given the optimised hyperparameters, the posterior distribution described by Equation 6.9 and Equation 6.10 is calculated. The posterior is obtained by passing the hyperparameters, mean and covariance functions, training data and test points to the GPML prediction function. Like the prior, the posterior distribution is also multivariate Gaussian with mean $\bar{\mathbf{f}}_*$ and covariance $cov(\mathbf{f}_*)$. In practice this permits

a predicted mean and variance (or a univariate Gaussian) to be calculated for each test input. In the present application this corresponds to a predicted target site wind speed given a concurrent reference site wind speed.

6.2.4 The MCP techniques

6.2.4.1 Regression of the scalar wind speeds

The main aim of the current study was to investigate the performance of an MCP approach based on predicting the sector-wise scalar wind speeds using GPR as compared to linear regression. Thus, the principle MCP methodology employed in this chapter was equivalent to the LR2 approach described in Chapter 5, with the exception that a non-linear GPR framework was used in place of linear regression to describe the relationship between the reference and target site wind speeds.

The same Gaussian scatter model employed in the LR2 approach was also employed in the current chapter although here the pointwise variance was extracted directly from the GPR (Equation 6.10) rather than inferred from the variance of the residuals. This scatter model is particularly intuitive when using the GPR framework since the GPR predictions are in fact Gaussian distributions centred on some mean value. Strictly, the variance computed from Equation 6.10 contains contributions from the function variance (the uncertainty in the underlying fit) and the noise variance. Since in the present application, no attempt was made to model the function uncertainty, only the noise variance was used in the implementation of the Gaussian scatter model. Hence the scatter was modelled using the following distribution of residuals:

$$\varepsilon \sim \mathcal{N}(0, \sigma_s^2)$$

Equation 6.11

For consistency, the GPR approach was applied using the same data processing procedures outlined in Chapter 5. Hence, the regression was applied sector-wise using 30º angular sectors and the MCP approach was applied to multiple training and test periods using the sliding window technique.

6.2.4.2 Orthogonal regression

As mentioned previously, while there is some basis for expecting that the scalar wind speeds at the reference and target sites may be better described by a non-linear approach, it is possible that if the non-linearity is weak it may not emerge from the short-term noisy data. Hence, to more fully explore the potential of the

GPR framework in cases where wind speed correlations are very likely to be non-linear, a secondary MCP approach was employed.

The approach was motivated by the work of Achberger *et al*. [106] who suggested separate regression of orthogonal wind speed vectors that are then recombined to produce a resultant wind speed and direction at the target site. This approach is referred to as orthogonal regression (OR).

The OR approach involves transforming the time series of wind speed measurements into eastern and northern components based on magnitude and direction using the expressions [106]:

$$u_E = u\sin(\theta)$$

Equation 6.12

$$u_N = u\cos(\theta)$$

Equation 6.13

where u_E and u_N are the observed eastern and northern wind components, u is the magnitude of the observed wind vector and θ is the wind angle (direction from) measured clockwise from north.

In a simple linear regression approach, a separate regression equation is calculated for each wind speed component of the form [106]:

$$\hat{u}_E = \alpha_E + \beta_E u_{E,ref}$$

Equation 6.14

$$\hat{u}_N = \alpha_N + \beta_N u_{N,ref}$$

Equation 6.15

where $\hat{u}_E$ and $\hat{u}_N$ are the predicted eastern and northern wind components at the target site, $u_{E,ref}$ and $u_{N,ref}$ are the concurrent measured values at the reference site and α and β are the regression coefficients. The resultant target site wind speed prediction $\hat{u}_{tar}$ is obtained using:

$$\hat{u}_{tar} = \sqrt{{\hat{u}_E}^2 + {\hat{u}_N}^2}$$

Equation 6.16

This approach is attractive in that despite its simplicity it provides a direct estimate of the target site wind angle through the predicted northerly and easterly wind components. Hence, although not the only available method [115, 156], OR can be used as a means to directly estimate the distribution of wind angles at the target site. This is in contrast to linear regression based on scalar wind speeds where the distributions of wind angles at the reference and target sites are typically assumed to be the same.

Implementation of the OR approach is, however, more challenging than scalar regression. Although not noted in the study by Achberger *et al*. [106], even if the resultant wind speeds are linearly related at the reference and target site, the linear relationships presented in Equation 6.15 and Equation 6.16 between orthogonal components will not always hold. For example, if there is some wind speed dependent veer (changes in wind direction) present at the target site, these relationships may be strongly non-linear. Hence, the OR approach is an ideal candidate for investigating the ability of GPR to model non-linear relationships in noisy wind speed data.

In implementing the OR approach using a GPR framework, Equation 6.15 and Equation 6.16 were modified so that the predicted wind speed components were calculated using non-linear GPR. In this approach, GPR was performed separately on the Easterly and Northerly components before recombination using Equation 6.16. The regression was performed sector-wise using 30º angular sectors and the scatter about the predicted values was modelled using the Gaussian scatter model described previously and applied to each wind speed component independently. Note that it is possible to make some theoretical objections to the use of a simple, constant variance, Gaussian scatter model in the present application, particularly in the presence of variable target site veer where the scatter will be a function of wind speed. However, in practice, the Gaussian model may be a sufficient approximation due to the stochastic nature of the wind speed observations.

The resultant wind speed predictions were obtained using Equation 6.16 and the predicted target site wind angle was extracted from the orthogonal wind speed components using the inverse sine and cosine operations. The OR approach was applied using the same data processing procedures outlined in Chapter 5.

6.2.5 Meteorological measurements and error metrics

For consistency and to allow cross comparison of the prediction approaches, the same wind data and error metrics described in Chapter 5 were used in the current chapter. Thus, the MCP approaches were applied to wind data covering a period of

11 years from a set of 22 reference/target site pairs. A sliding window approach was used to select multiple 3 month training and 10 year test periods in order to increase the robustness of the error statistics. Errors were quantified using the metrics of average absolute percentage error (%Error), mean absolute error (MAE) and mean bias error (MBE) described by Equation 5.2 to Equation 5.4.

6.3 Results and Discussion

6.3.1 Prediction of the scalar wind speed

In this section, results are presented for the application of the GPR approach to prediction of the sector-wise scalar wind speeds and the derived wind resource statistics. This allows direct comparison with the linear approaches studied in Chapter 5.

To demonstrate the application of GPR to MCP, a single reference/target site pair, C6-Rf11, is first considered. As described in Section 6.2, the GPR approach optimises the GP hyperparameters using short-term training data in the training phase. Given the prior assumption of a linear correlation between reference and target site wind speeds, it might be expected that the optimised gradient and y-intercept of the GPR affine function would be similar to the values extracted using LR2. Similarly, while the GPR signal noise, σ_s, is estimated using a likelihood approach, and the equivalent parameter in LR2 is inferred from the variance of the residuals, it might be expected that these estimates would be comparable as they both describe the residual scatter about some mean prediction. Table 6.1 compares these parameters as extracted from the GPR optimisation over the hyperparameters and the standard LR2, least squares regression fit. The parameters are shown for 12 angular sectors using a single 3 month training period. Since these values are from a single site and single training period, they are intended to be illustrative only.

	GPR	LR2	GPR	LR2	GPR	LR2
Sector (deg)	**σ_s (ms^{-1})**	**σ_{res} (ms^{-1})**	**Gradient**		**Y-intercept (ms^{-1})**	
0	1.87	1.97	0.86	0.80	3.09	3.21
30	1.55	1.58	0.66	0.66	3.45	3.46
60	1.53	1.65	0.20	0.41	3.80	3.24
90	2.27	2.28	0.54	0.54	2.74	2.74
120	2.36	3.73	2.13	1.25	-2.23	-0.60
150	2.31	2.65	2.05	2.11	-2.30	-2.84
180	2.09	2.17	0.94	1.04	1.38	0.94
210	1.50	1.50	0.88	0.87	0.84	0.88
240	1.20	1.25	0.75	0.75	1.12	0.81
270	1.43	1.44	0.61	0.57	1.78	1.92
300	1.63	1.61	0.68	0.64	2.64	2.81
330	1.60	1.68	0.89	0.84	2.67	2.60

Table 6.1: Optimised hyperparameters of the noise (or standard deviation of residuals), gradient and y-intercept obtained from the training phase of the GPR, as well as the equivalent parameters extracted from the LR2 linear regression fit. Parameters are shown for 12 angular sectors at the reference/target site pair Rf11/C6 using a single 3 month training period.

As expected, the values of the optimised hyperparameters are generally very similar to those obtained from the linear regression. This indicates that the starting point for GPR, before the prediction phase where the posterior distribution is estimated (Section 6.2.3), is close to the end point of LR2. From this starting point, the GPR approach adds further value by allowing the predicted wind speeds to deviate from the prior assumption of linearity, in cases where there is sufficient evidence for such deviations (Equation 6.9). An interesting case is seen in the sector at 120º where the GPR and LR2 estimates of both the gradient and standard deviation of residuals differ significantly. Analysis of the hourly wind speeds for this sector showed the preferred GPR fit was strongly non-linear due to a cluster of low target site wind speeds. Hence, in this case the optimised gradient hyperparameter was somewhat greater than the LR2 estimate, since the low wind speeds could be accounted for by a non-linear fit in this region. A similar observation was made for the 60º sector demonstrating the greater flexibility of the GPR approach.

Figure 6.1 shows hourly averaged target and reference site wind speeds for a single angular sector (240º) from the same reference/target site pair along with linear and GPR fits to the data. The frequency of observations or predictions at each wind speed is indicated by the colour of the shading. Note this figure includes the same wind speed observations shown in Chapter 5, Figure 5.4 although in the current figure, the GPR predictions have been added. As stated in Chapter 5, these results are intended as an example rather than to be fully representative.

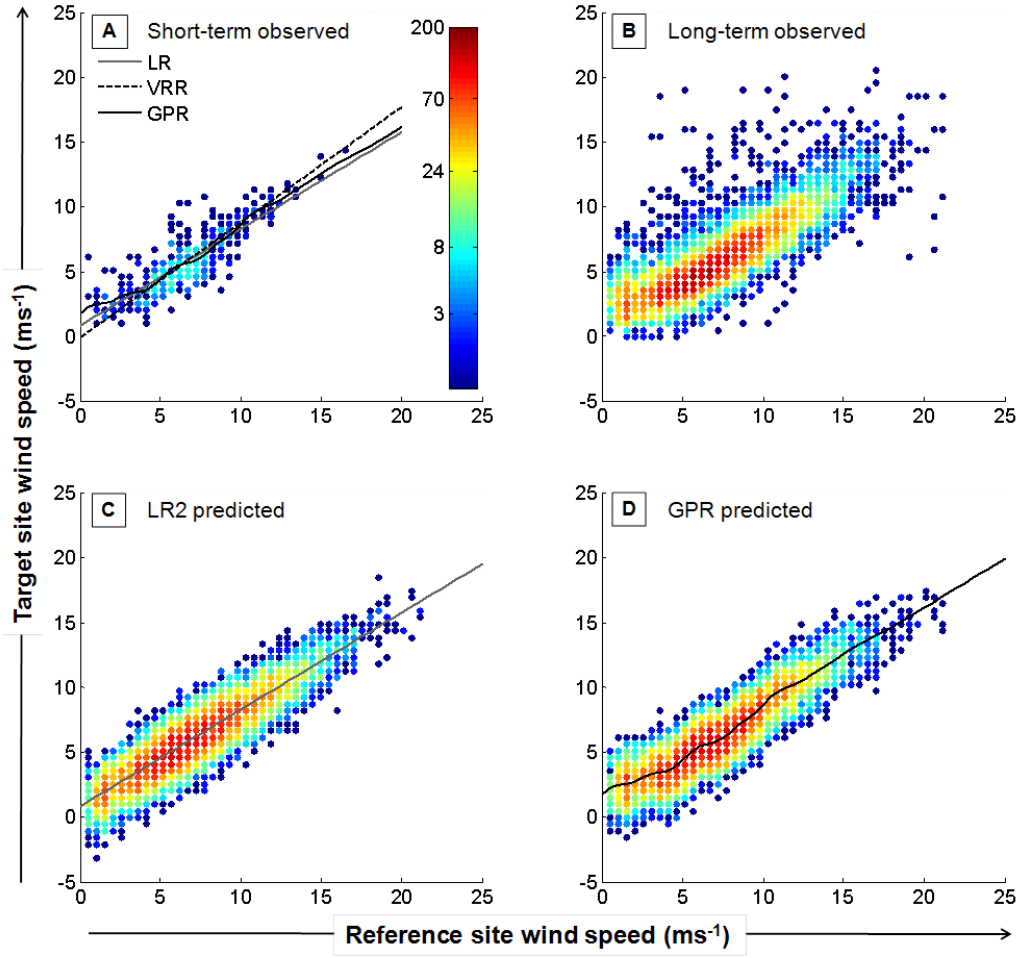

Figure 6.1: Target and reference site wind speeds for a single angular sector from the reference/target site pair C6-Rf11. A) Observations over the 3 month training period along with the linear and GPR fits, B) 10 year observations, C) 10 year predictions using LR2, D) 10 year predictions using GPR. The solid lines represent the mean prediction, the dots show the observed or predicted scatter and the shading represents the frequency.

Figure 6.1A shows the observed wind speeds over the 3 month training period along with the linear and GPR fits. The GPR picks up a small amount of non-linearity in the short-term data, although it is not clear whether this is real or due to the sparsity of the data. Figure 6.1B shows the long-term observations taken from the same angular sector over a period of 10 years and Figure 6.1C and D show the attempt to predict these observations using the MCP approaches of LR2 and GPR, respectively. The solid lines in these panels represent the mean predictions while the scatter shows the predictions that result from applying the Gaussian scatter model described previously. From visual inspection it is not immediately clear which MCP approach best models the data. The slight flattening of the mean regression line at low wind speeds in the GPR approach is perhaps preferable since it tends to reduce the incidences of negative wind speed predictions. However, on average the LR2 and GPR regressions appear to result in similar predictions.

It is of interest that for reference site wind speeds >15 ms^{-1}, the GPR mean prediction becomes very close to linear. Note that during the 3 month training period, there were very few wind speed observations >15 ms^{-1}. Hence, this is an example of the GPR predictions tending towards the prior affine mean function in the absence of training observations, as described in Section 6.2.2.

The width of the distribution of predicted scatter for both LR2 and GPR can be seen to be similar in Figure 6.1C and D. This is a consequence of the similarity between the optimised noise hyperparameter (σ_s) and the estimated standard deviation of residuals (σ_{res}), as shown in Table 6.1 for the 240º sector. It is evident that the Gaussian scatter model reproduces the main density of scatter but the predicted distribution is somewhat narrower than the observed data in Figure 6.1A. This is likely because the short training period does not fully capture the long-term variability. In addition, the Gaussian model does not reproduce the occasional outliers evident in the observed data. These outliers are related to deviations from ideal Gaussian behaviour as discussed in Section 5.3 (Figure 5.5). Due to the stochastic nature of wind flows and the potential for highly localised meteorological phenomena, even highly correlated reference/target sites will exhibit such deviations. Whether a Gaussian scatter model is the most suitable approximation for this behaviour is an open question that will be considered in more detail in Chapter 7.

6.3.1.1 Comparison between GPR and linear regression

Table 6.2 compares the error metrics of %Error, MAE and MBE for the most successful linear MCP approach identified in Chapter 5, namely LR2, as well as the GPR approach introduced in the current chapter. To two significant figures, the error metrics across all parameters are almost identical using both approaches. Hence, even with the greater flexibility of GPR, including the use of a Bayesian framework and removal of the restriction of linearity between the reference and target site wind speeds, the overall error metrics are not improved.

	Method	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	LR2	4.8	14	6.2	7.8
	GPR	4.8	14	6.2	7.7
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	LR2	0.21	11	0.16	0.14
	GPR	0.21	11	0.16	0.14
MBE	LR2	<0.1	-2.8	-0.10	0.13
	GPR	<0.1	-2.8	-0.10	0.13

Table 6.2: Error metrics averaged across 22 target sites and all training/test periods for the MCP approaches of GPR and LR2.

A possible explanation for the similarity between the GPR and LR2 error metrics is that due to the stochastic nature of the wind data, as well as the short duration of the training period, non-linearities either do not exist or do not emerge from the training period, and the hence the GPR reduces to an approximately linear function. To investigate this, the relationship between the long-term GPR predicted target site winds speeds and the long-term input reference site data was investigated by means of the linear correlation coefficient. In effect, this is a measure of how closely the GPR predictions correspond to a linear function of the long-term reference site observations.

The linear correlation coefficient (generally denoted r) is frequently used to investigate the relationship between the target and reference site wind speeds (see Table 5.1 for example). Hence, to avoid confusion in the present case, where the metric was applied to the GPR *predicted* wind speeds, the parameter is denoted as the linearity Lin. For a single 30º sector this may be defined as:

$$Lin_\alpha = \frac{\sum_j (u_{\alpha,j,ref} - \bar{u}_{\alpha,ref})(\hat{u}_{\alpha,j,GPR} - \bar{\hat{u}}_{\alpha,GPR})}{\sqrt{\sum_j (u_{\alpha,j,ref} - \bar{u}_{\alpha,ref})^2}\sqrt{\sum_j (\hat{u}_{\alpha,j,GPR} - \bar{\hat{u}}_{\alpha,GPR})^2}}$$

Equation 6.17

where $u_{\alpha,j,ref}$ and $\hat{u}_{\alpha,j,GPR}$ represent the j^{th} reference-site-observed and target-site-predicted wind speeds respectively and $\bar{u}_{\alpha,ref}$ and $\bar{\hat{u}}_{\alpha,GPR}$ denote the equivalent long-term means. The subscript α denotes wind speeds from a single 30º sector, and the subscript GPR indicates that the predictions are made using the GPR approach.

Since a sector-based MCP approach is used, Lin_α must be calculated for predictions in each of the 12 angular sectors. Each sector contributes differently to the overall long-term wind resource predictions depending on the long-term

frequency of observations within that sector. Hence, linearity in certain sectors will be more significant than others. To reflect this, for any given site, a weighted sum Lin of the 12 sector-wise Lin_α values was calculated. The long-term fractional frequency in each sector was used as the weighting factor W_α, according to:

$$Lin = \sum_{\alpha=1}^{12} Lin_\alpha * W_\alpha$$

Equation 6.18

$$W_\alpha = \frac{freq_{\alpha,ref}}{\sum_{\alpha=1}^{12} freq_{\alpha,ref}}$$

Equation 6.19

where $freq_{\alpha,ref}$ is the observed long-term reference site frequency in sector α. The weighting was performed with respect to the reference site frequencies since these are used to assign the wind speeds to each of the 12 sectors.

In the calculation of Lin_α, the GPR mean function predictions were used before the addition of the scatter term, as represented by the solid line in Figure 6.1D. Note that Lin simply represents the linear correlation between the target site predictions and the reference site observations. For the LR2 approach, before the addition of the scatter term, Lin will be unity since a linear correlation is imposed by linear regression. However, for the GPR approach, Lin can be used as a measure of the linearity of the extracted GPR function with values close to unity indicating that the GPR approach has reduced to a linear function.

Equation 6.18 defines the linearity for a single training/test period. The final estimates of Lin were calculated as the average across all 120 training/test periods. Figure 6.2 shows the site-by-site values for the linearity parameter Lin when using GPR as well as the %Error in $\bar{p}_d$ using both the LR2 and GPR approaches.

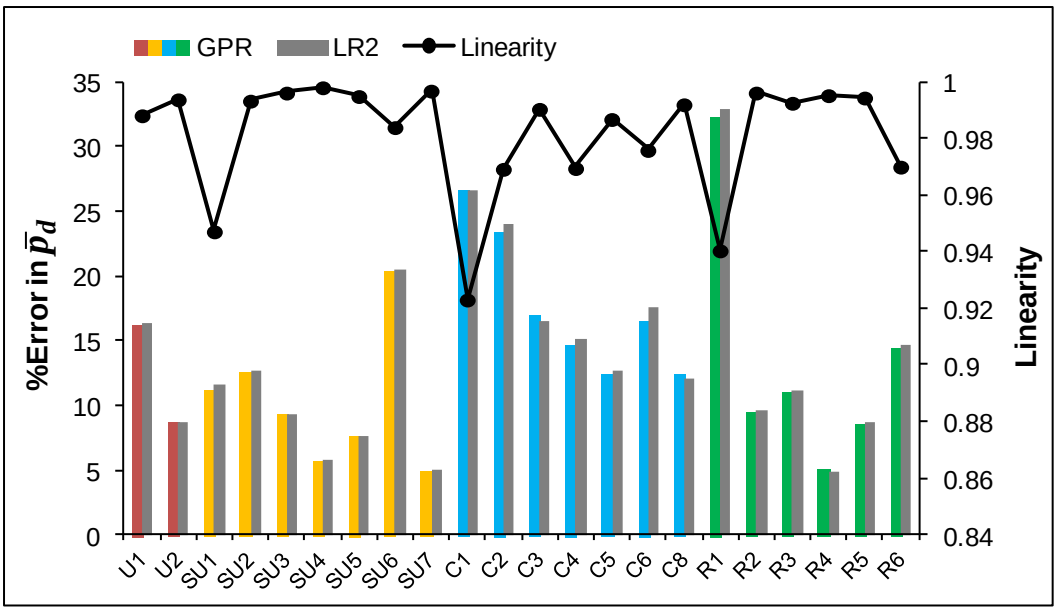

Figure 6.2: %Error in $\bar{p}_d$ for each of the 22 target sites using the GPR and LR2 approaches and a training period of 3 months. The linearity (Lin) of the GPR approach is also shown. Values are averaged across 120 training/test periods.

As reflected in the average statistics shown in Table 6.2, the %Error values are very close using both GPR and LR2. In addition, for all sites, Lin is close to unity with an average value across all sites of 0.98, and values less than 0.97 only occurring in three cases.

For completeness, the above calculations were repeated using a training length of 12 months. However, as with the 3 month training period, the average errors for the LR2 and GPR approaches were almost identical and the linearity parameter was again found to be close to unity. These results indicate that for the current sites, the GPR approach when applied to the scalar wind speeds reduces to a model that is very close to linear and results in almost identical average wind resource predictions as LR2.

6.3.2 Orthogonal regression

As highlighted in Section 6.2.4, a regression based on orthogonal wind vectors (OR) is more likely to feature non-linearities, even if the scalar wind speeds at two sites are linearly related. MCP based on OR thus offers an opportunity to further explore the performance of GPR in a more challenging application. Additionally, an MCP approach based on OR is of interest since a successful application of this techniques can potentially provide more detailed information regarding the distribution of target site wind directions compared to regression of the scalar wind speeds.

Table 6.3 compares the error metrics of %Error, MAE and MBE for both LR2(OR) and GPR(OR) as applied to an orthogonal regression MCP approach. The added value of the non-linear GPR(OR) approach is visible in the average error metrics across all parameters. For example, the %Error in $\bar{u}$ and $\bar{p}_d$ are reduced by 1 and 3 percentage points respectively using GPR(OR) compared to LR2(OR). For longer training periods it was found that this difference becomes much more pronounced with a reduction in the %Error in $\bar{p}_d$ rising to almost 8 percentage points using a 12 month training period. Comparison between the error metrics for scalar regression (LR2) also repeated in Table 6.3, demonstrates that while LR2(OR) leads to larger errors in the predicted wind resource, the GPR(OR) results in very similar, or slightly improved, error metrics compared to LR2. Additionally, the OR approach has the potential advantage of predicting the distribution of target site wind angles as analysed separately in the next section.

	Method	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	LR2	4.8	14	6.2	7.8
	LR2 (OR)	5.9	16	6.4	6.0
	GPR (OR)	4.7	13	5.1	5.3
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	LR2	0.21	11	0.16	0.14
	LR2 (OR)	0.27	13	0.16	0.11
	GPR (OR)	0.21	10	0.13	0.10
MBE	LR2	<0.1	-2.8	-0.10	0.13
	LR2 (OR)	-0.16	-9.1	<0.1	-0.10
	GPR (OR)	<0.1	-0.36	<0.1	<0.1

Table 6.3: Error metrics averaged across 22 target sites and all training/test periods using the orthogonal regression MCP approaches of LR2(OR) and GPR(OR). The results for scalar regression (LR2) are also shown in grey for comparison.

As described in Section 6.3, a linearity parameter Lin was also calculated for the OR approach. Note that OR involves a regression on both the Northerly and Easterly wind vectors resulting in two values of Lin. However these were found to follow similar trends and hence the average of the two was used to assess the overall linearity. Figure 6.3 shows the site-by-site values for Lin as well as the %Error in $\bar{p}_d$ using both the LR2(OR) and GPR(OR) approaches.

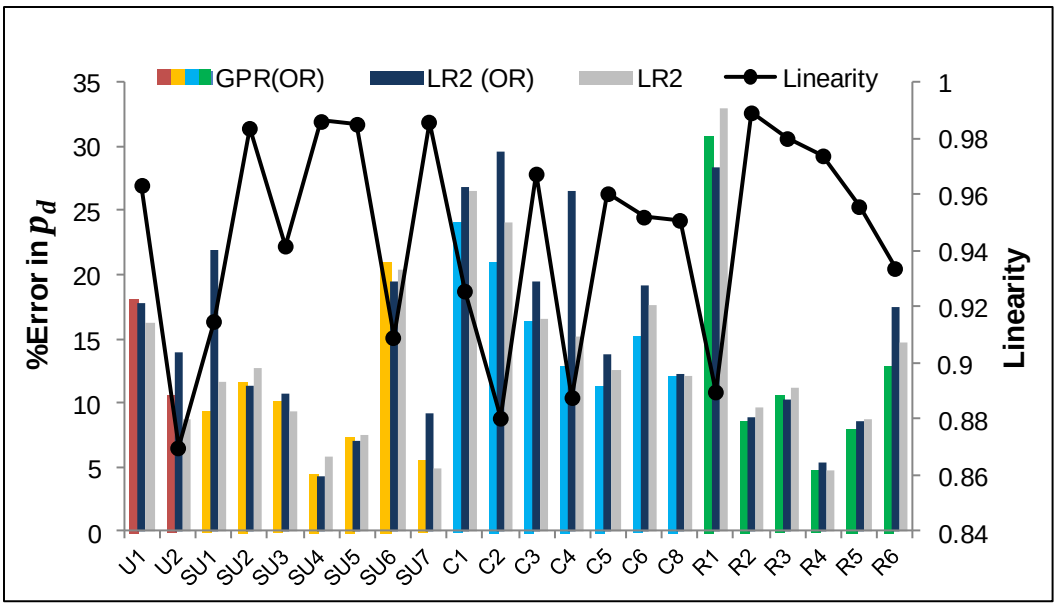

Figure 6.3: %Error in $\bar{p}_d$ for each of the 22 target sites using the GPR(OR) and LR2(OR) approaches and a training period of 3 months. The linearity (Lin) of the GPR(OR) approach is also shown, as well as the %Error for scalar regression, LR2. Values are averaged across 120 training/test periods.

At 14 out of the 22 sites, the GPR(OR) approach results in a smaller %Error in $\bar{p}_d$ compared to LR2(OR), and in many cases these improvements are substantial. In addition, while Lin is still close to unity across the sites, it is noticeably more variable compared to the scalar regression case shown in Figure 6.2, with an average value 0.94 (compared to 0.98 for scalar regression). Due to the short training period, there is no guarantee that non-linearity in the GPR(OR) model represents actual non-linearity in the long-term reference/target site relationship. However, if the GPR(OR) model is successful, on average one would expect to see the largest improvements over LR2(OR) when Lin is lowest. There is some indication of this in Figure 6.3, see for example sites U2, SU1, C2 and C4, however it is not universally the case as shown by sites SU6 and R1 where Lin is low but LR2(OR) performs slightly better than GPR(OR).

Overall, the results show that the GPR(OR) approach performs better than LR2(OR) when applied to orthogonal regression and the resulting errors are similar to those obtained from scalar regression. In addition, the non-linearities of the GPR predictions are increased for orthogonal regression compared to the scalar case. These observations indicate that the orthogonal regression relationships are more likely to be non-linear and GPR is thus better able to model these non-linearities compared to LR2.

6.3.2.1 Target site distribution of wind angles

Since GPR(OR) results in similar error metrics to scalar regression, the OR approach can only really be justified if it provides additional information compared to scalar regression. As outlined previously, an attractive property of OR is that it is capable of predicting a time series of target site wind vectors rather than just the scalar wind speed. These can be used to calculate the long-term distribution of target site wind angles. While this is perhaps of secondary importance in the case of small-scale wind turbines, it can be useful in turbine siting as well as the analysis of local flow effects. In addition, techniques capable of predicting wind direction are of particular interest to the large-scale wind industry where the information is required when considering wind farm layout and wake effects.

For MCP based on scalar regression, the distribution of wind directions at the target site is typically assumed to be the same as the reference site. While this may be a satisfactory assumption in some cases, differences in terrain or local climatology can result in substantial veer between the reference and target sites. To estimate whether the OR approach provides added value compared to scalar regression, the long-term GPR(OR) predicted distributions of target site wind angles were compared with those at the reference and target sites. Figure 6.4 shows this comparison for two sites, the urban site U1 and the rural site R6.

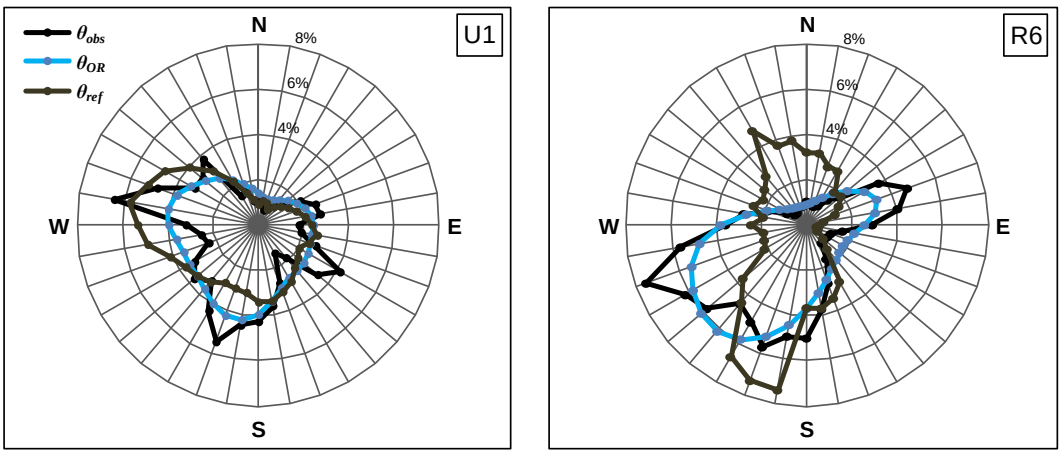

Figure 6.4: Long-term percentage frequency of wind directions for site U1 (left) and R6 (right). The plots show the observed values (θ_{obs}) and predictions based on GPR(OR) (θ_{OR}) and based on the reference site distribution (θ_{ref}).

It can be seen that the GPR(OR) predicted distributions tend to vary more smoothly than the observations at the reference and target sites. For site U1, while both the GPR(OR) predictions and the reference site distribution bear some resemblance to the target site distribution, neither reproduce the complexities observed at the target

site and hence both produce similar errors. For site R6, there is significant veer between the reference and target sites, likely caused by the moderately complex terrain at the two sites. In this case, the reference site distribution is a poor predictor of the target site wind angles. The GPR(OR) predictions perform noticeably better, predicting the reduction in southerly and northerly winds as well as the general form of the target site distribution.

To make quantitative comparisons across sites, a percentage error metric ($\%Error\ in\ \theta_{tar}$) was calculated to represent the angular distribution error. For a particular training/test period, the percentage error in the predicted long-term frequency in each 10º angular sector was first obtained. As with the Lin parameter, these errors were then combined into a single value for a particular site by a weighted sum to account for the relative contribution of each angular sector to the overall error. The weighting factor was obtained from the long-term fractional frequency in each sector as defined below:

$$\%Error\ in\ \theta_{tar} = 100 * \sum_{i=1}^{36} W_i \left[\frac{|freq_{i,obs} - freq_{i,pred}|}{freq_{i,obs}} \right]$$

Equation 6.20

$$W_i = \frac{freq_{i,obs}}{\sum_{i=1}^{36} freq_{i,obs}}$$

Equation 6.21

where $freq_{i,obs}$ and $freq_{i,pred}$ represent the observed and predicted frequencies respectively in the i^{th} angular sector, and W_i is the weighting factor. Note this is in effect equivalent to summing all the individual errors and then normalising this sum as a percentage based on total number of counts across all sectors. This procedure defines the percentage error for a particularly training/test period and the final metric is obtained as the average across all 120 test periods.

Figure 6.5 compares the site-by-site values for the %Error in θ_{tar} based on using either the GPR(OR) approach or the reference site distribution (Rf) to predict the long-term angular distribution at the target site. The linearity parameter values Lin, are also shown.

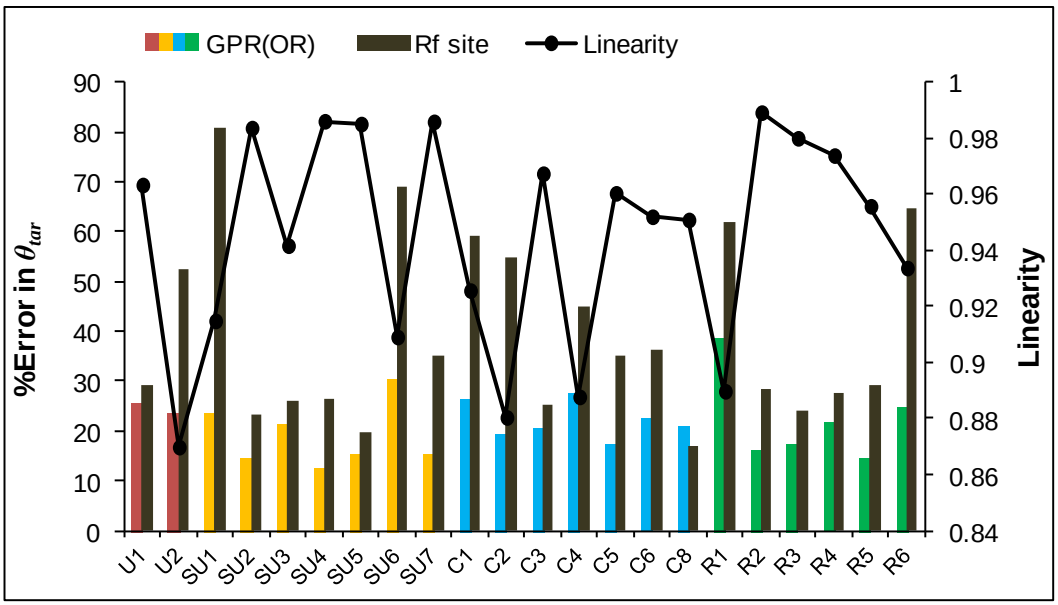

Figure 6.5 %Error in θ_{tar} for each of the 22 target sites using the GPR(OR) approach and the reference site distribution (Rf site) as predictors. The linearity (Lin) of the GPR approach is also shown.

For all sites except site C8, it can be seen that the GPR(OR) approach results in reduced errors compared to using the reference site distribution. The %Error in θ_{tar} when using the reference site distribution can also be used as an indicator of the amount of veer at the target site since larger values indicate greater differences between the reference and target site angular distributions. It is interesting to note that the largest errors (or veer) occur for the smallest values of Lin. As noted in Section 6.2.4, non-linearity may be associated with variable veer between the reference and target sites and hence it is not surprising that the GPR models that deviate most from linear (low Lin) are found at sites with the greatest veer.

Averaged across all 22 sites, the %Error in θ_{tar} was found to be 21% and 40% for predictions based on GPR(OR) and the reference site distribution respectively. Thus, the GPR(OR) approach can be seen to add value through its ability to almost halve the percentage error in θ_{tar} compared to using the reference site distribution. Ultimately, it would be informative to compare this improvement with other methods [115] capable of predicting the distribution of target site wind directions to further quantify its value.

6.4 Conclusions

An MCP approach based on GPR has been applied in a scalar regression context to predict the long-term wind resource at 22 sites based on a short-term training period of 3 months. Despite the greater flexibility of the GPR approach, including

the use of a Bayesian framework and the ability to represent non-linearity, GPR failed to outperform the linear LR2 approach. Further investigation revealed that for the sites considered, the GPR predictions reduce to a model that is close to linear with respect to the reference site wind speeds indicating that non-linearities do not emerge from the short-term, noisy training data. Extension of the training period to 12 m also failed to reveal any significantly different trends.

The study was extended to the more challenging case of orthogonal regression where the correlation between reference/target site wind speeds is more likely to be non-linear. In this case the GPR approach was found to perform better than the linear LR2 approach across all error metrics. In addition, the difference between the GPR and LR2 approaches was found to increase with longer training periods. Further investigation revealed that the GPR predictions were based on models that were more strongly non-linear than in the case of scalar regression. Thus, the improved performance of GPR compared to LR2 is likely related to its ability to extract the non-linear relationships present in orthogonal regression.

An attractive property of the orthogonal regression approach is its potential to predict the distribution of wind angles at the target site rather than relying on the assumption that the reference and target distributions are equivalent. Application of the GPR approach to orthogonal regression was found to almost halve the average percentage error in this distribution compared to using the distribution of wind angles at the reference site as a predictor of the target site distribution. Thus, in cases where the distribution of target site wind angles is required, there is added value in using the GPR approach as applied to orthogonal regression. In terms of the current thesis however, the benefits of orthogonal regression in predicting the target site wind direction are of secondary importance. Hence, while this is an interesting topic for future study, this area is not pursued further in this work.

At this point, it is appropriate to consider in more detail the assumption of Gaussian scatter that has so far been used to predict the distribution of target site wind speeds using both linear and non-linear approaches. In effect, this assumption considers the residual scatter as noise superimposed on some underlying function. This noise is assumed to follow a symmetrical Gaussian distribution of constant variance, centred on the mean prediction. As highlighted in Sections 5.3 and 6.3, real wind speed observations may deviate from these simple assumptions and this raises the question of whether a more appropriate model can be developed. To develop a more sophisticated approach, the underlying, two-dimensional probability distribution that describes correlated wind speeds at the reference/target site pair

must be considered. The following chapter will investigate this underlying distribution in detail.

7 Prediction of Correlated Wind Speeds Using a Bivariate Weibull Measure-Correlate-Predict Approach

7.1 Overview

Thus far, several data-driven MCP techniques, both linear and non-linear have been tested in terms of their ability to predict the long-term wind resource using very short-term onsite measurements. It is clear from Chapter 5 that modelling the residual scatter about the mean prediction (the LR2 approach) is vital in accurately predicting statistics related to the wind speed distribution, including the wind power density. The results from Chapter 6 imply that allowing for the possibility of non-linearity is in fact of secondary importance compared to an accurate representation of the scatter. In light of these results, attention is now turned to how this scatter may best be represented from a theoretical perspective given certain assumptions about the expected form of the wind speed distribution.

In order to fully understand the form of the residual scatter, a bottom-up approach considering the theoretical form of the underlying joint probability distribution between the reference and target site wind speeds is required. With a few exceptions, [110, 121], such approaches have received relatively little attention in the literature. In addition, the MCP techniques implemented in commercial software packages [104, 105, 157] are often restricted to top-down regression or scaling approaches, presumably due to their simplicity and empirical success.

The motivation for considering the underlying joint probability distribution arises from the observation that simple linear regression approaches are based on the assumption of a bivariate Gaussian distribution between two variables [110]. Hence the natural way to model the residual scatter is using a zero mean Gaussian distribution centred on the mean prediction, as discussed in Chapter 5. However, since wind speeds are typically assumed to follow univariate Weibull distributions [41], there is a stronger theoretical justification for describing the correlation between target and reference site wind speeds using a bivariate Weibull (BW) distribution. Such an approach provides a direct mathematical basis for modelling the distribution of wind speeds at the target site given a specific input wind speed at the reference site. In addition, work by Clive [116] showed that given Weibully distributed wind speeds at correlated sites, the relationship between the reference

and target site wind speeds will not be well described by linear regression except in the special case where the Weibull shape factors are equal at both sites. For the 22 reference/target site pairs used in the current study, the ratio of reference/target site Weibull shape factors varied between 0.75 and 1.35, hence, there is cause to examine the justification for the widespread use of linear regression in MCP applications.

Recently, Perea *et al*. [110] used artificially generated wind speed data to investigate the utility of an MCP approach based on a BW probability distribution. Their results indicated that the approach performed better than several established MCP techniques. However, a vital question is whether such a promising approach can be successfully applied to real wind speed observations that will frequently deviate from idealised correlated Weibull distributions, and which may also include correlations that are dependent on the measurement season and wind direction at both sites. Of further interest in the context of small-scale wind energy is the performance of the technique using short measurement periods of less than one year.

The contribution of the work described in this chapter is a detailed investigation of a BW approach to MCP that has not previously been applied to real wind speed measurements at correlated sites. Additionally, real wind speed measurements are supplemented by artificially generated wind data in order to investigate differences between the observed data and idealised BW distributions, as well as the additional challenges associated with applying the approach to real wind observations using a range of onsite measurement periods. Particular attention is given to the success of the technique when applied to very short onsite measurement periods, as is likely for small-scale wind installations, and the technique is compared to the linear MCP approaches considered in Chapter 5.

The main objectives of this study can be summarised as follows:

(i) Application of a theoretical BW framework for MCP and the introduction of a new method for estimating the degree of association between the reference and target sites.

(ii) Comparison of the success of the approach when applied to real wind observations at 22 target sites, as well as artificial wind data based on idealised BW representations of the observed data.

(iii) An investigation of the efficiency with which the BW parameters may be extracted using both real and artificial wind data and the implications for short measurement periods.

(iv) Comparison of the performance of the BW MCP approach with the linear techniques detailed in Chapter 5.

(v) Quantification of the typical prediction errors that may be expected when applying the various MCP approaches to onsite measurement periods of between 1 and 12 months.

This chapter is organised as follows: In Section 7.2 a general bivariate probability approach to MCP is described followed by a specific description of the bivariate Weibull probability approach. In Section 7.3.1 the BW approach is applied to real and artificial bivariate wind speed distributions to investigate the additional complexities that arise when using real wind observations. Finally, the performance of the BW approach is compared to baseline MCP approaches across 22 sites and the effects of measurement length, seasonal variability and goodness-of-fit are investigated in Section 7.3.2.

7.2 Methodology

The majority of MCP approaches are concerned with predicting a long-term historical time series of wind speeds (and possibly wind directions) using short-term concurrent wind measurements at a correlated reference and target site pair. The short-term concurrent wind measurements are used to model the relationship between the two sites, while the long-term historical reference site observations are used as an input to this model to enable prediction of the target site wind speeds.

In the case of simple linear regression, for any input wind speed at the reference site there is a corresponding single-valued wind speed prediction at the target site. If this process is repeated for the full historical time series of reference site wind data, the output is a predicted long-term historical time series at the target site. The assumption is made that statistical parameters extracted from this predicted time series will be representative of the long-term future wind resource. A similar assumption also applies when using a BW probability approach but with the following distinctions. Firstly, the BW approach seeks to directly model the underlying distribution of wind speeds at the target site rather than predicting the historical time series. Secondly, rather than the restriction that a specific reference site wind speed input corresponds to a specific target site wind speed output, the BW approach predicts a distribution of target site wind speeds for every reference site wind speed in the form of a conditional probability distribution. Since wind power is proportional to the cube of the wind speed, these characteristics are

important in achieving accurate predictions of the wind resource. The BW approach will now be described in more detail.

7.2.1 A bivariate probability approach to MCP

As a recap of the information given in Section 3.2.3, the starting point for a bivariate approach to MCP is a description of the underlying bivariate probability density function (pdf) of the reference and target site wind speeds. Given a set of two, correlated, random variables, the bivariate pdf may be described by [110]:

$$f\left(u_{tar}\middle|u_{ref}=u'_{ref}\right)=\frac{f(u'_{ref},u_{tar})}{f(u'_{ref})}$$

Equation 7.1

where in the current application, u_{ref} and u_{tar} represent wind speed observations at the reference and target sites respectively and u'_{ref} is a specific value of u_{ref}, $f(u_{ref},u_{tar})$ is the bivariate pdf and $f(u_{ref})$ represents the univariate pdf at the reference site.

The marginal pdf at the target site, $f(u_{tar})$, is obtained by integrating the product of the conditional pdf (Equation 7.1) and the marginal pdf at the reference site, $f(u_{ref})$, over all reference site wind speeds using [110]:

$$f(u_{tar})=\int f\left(u_{tar}\middle|u_{ref}=u'_{ref}\right)f\left(u_{ref}\right)\mathrm{d}u_{ref}$$

Equation 7.2

The aim of a bivariate probability approach to MCP is to obtain a long-term prediction of the target site marginal pdf of wind speeds $f(u_{tar})$. This distribution can then be used to extract the key statistical parameters of the target site wind resource.

To implement an MCP approach based on an underlying bivariate pdf, we require a prediction of the long-term marginal pdf of wind speeds at the target site, $f_{long}(u_{tar})$, based on a short-term measurement period. Combining Equation 7.1 and Equation 7.2 we have:

$$f_{long}(u_{tar})=\int\frac{f(u_{ref},u_{tar})}{f_{short}(u_{ref})}f_{long}\left(u_{ref}\right)\mathrm{d}u_{ref}$$

Equation 7.3

where the subscripts ‘short’ and ‘long’ refer to the short-term training period and long-term prediction period respectively.

In line with the approach proposed by Perea *et al*. [110], the assumption is made that the short-term measurement period is sufficient to determine the form of the underlying bivariate pdf, $f(u_{ref}, u_{tar})$ using some fitting procedure and that this function does not change with time. To obtain $f_{long}(u_{tar})$ from a short-term measurement campaign we also require an estimate of the long-term reference site wind speed distribution $f_{long}(u_{ref})$. This is obtained by fitting a univariate Weibull distribution to the long-term wind speed observations at the reference site. In practice, the wind speed observations are discrete rather than continuous and the integral in Equation 7.3 is replaced with a summation of the function values at discrete intervals.

7.2.2 Application of a bivariate Weibull probability approach to MCP

While a number of BW constructions are possible [158], in the present application we require a formulation that yields two-parameter, univariate, Weibull marginals and whose likelihood function is analytically tractable. Here the BW previously employed by Johnson *et al*. [159] in relation to strength properties of lumbar, which was later applied to artificial wind data by Perea *et al*. [110] is used. The BW pdf contains five parameters and is described by [159]:

$$f(u_{ref}, u_{tar}) = \frac{k_{ref}}{c_{ref}}\left(\frac{u_{ref}}{c_{ref}}\right)^{\left(\frac{k_{ref}}{D}\right)-1}\frac{k_{tar}}{c_{tar}}\left(\frac{u_{tar}}{c_{tar}}\right)^{\left(\frac{k_{tar}}{D}\right)-1}\left\{\left(\frac{u_{ref}}{c_{ref}}\right)^{\frac{k_{ref}}{D}}+\left(\frac{u_{tar}}{c_{tar}}\right)^{\frac{k_{tar}}{D}}\right\}^{D-2}$$

$$\mathrm{x}\left\{\left[\left(\frac{u_{ref}}{c_{ref}}\right)^{\frac{k_{ref}}{D}}+\left(\frac{u_{tar}}{c_{tar}}\right)^{\frac{k_{tar}}{D}}\right]^{D}+\frac{1}{D}-1\right\}\exp\left\{-\left[\left(\frac{u_{ref}}{c_{ref}}\right)^{\frac{k_{ref}}{D}}+\left(\frac{u_{tar}}{c_{tar}}\right)^{\frac{k_{tar}}{D}}\right]^{D}\right\}$$

Equation 7.4

where k and c are the Weibull shape and scale factors respectively, $0 < D \leq 1$ describes the degree of association between wind speed observations at the two sites and the subscripts ref and tar refer to the reference and target sites respectively. The magnitude of D is inversely related to the degree of correlation between the two sites [158].

Johnson *et al*. [159] showed that the log-likelihood ($\ln L$) function for this distribution is tractable and may be used to fit the BW to concurrent observations of the two

correlated variables using the method of maximum likelihood (MML). The $\ln L$ is given by:

$$\ln L = n \ln\left(\frac{k_{ref}}{c_{ref}}\right) + n \ln\left(\frac{k_{tar}}{c_{tar}}\right) + \left[\left(\frac{k_{ref}}{D} - 1\right)\sum_{i=1}^{n} \ln\left(\frac{u_{ref,i}}{c_{ref}}\right)\right]$$

$$+ \left[\left(\frac{k_{tar}}{D} - 1\right)\sum_{i=1}^{n} \ln\left(\frac{u_{tar,i}}{c_{tar}}\right)\right] + \left\{(D-2)\sum_{i=1}^{n} \ln\left[\left(\frac{u_{ref,i}}{c_{ref}}\right)^{\frac{k_{ref}}{D}} + \left(\frac{u_{tar,i}}{c_{tar}}\right)^{\frac{k_{tar}}{D}}\right]\right\}$$

$$+ \sum_{i=1}^{n} \ln\left\{\left[\left(\frac{u_{ref,i}}{c_{ref}}\right)^{\frac{k_{ref}}{D}} + \left(\frac{u_{tar,i}}{c_{tar}}\right)^{\frac{k_{tar}}{D}}\right]^{D} + \frac{1}{D} - 1\right\} - \sum_{i=1}^{n}\left[\left(\frac{u_{ref,i}}{c_{ref}}\right)^{\frac{k_{ref}}{D}} + \left(\frac{u_{tar,i}}{c_{tar}}\right)^{\frac{k_{tar}}{D}}\right]^{D}$$

Equation 7.5

where n is the total number of wind speed observations, $u_{ref,i}$ and $u_{tar,i}$ represent the i^{th} concurrent wind speed observation at the reference and target sites respectively and $\ln$ is the natural logarithm.

In this work, short-term wind speed observations at the reference and target sites were used to obtain the fitted BW pdf by minimising the negative $\ln L$ (equivalent to maximising $\ln L$) using a multidimensional, non-linear Nelder-Mead search implemented in the MATLAB programming environment, as used previously [159]. Following the method of Johnson *et al*. [159], the minimisation was implemented as follows: (i) starting estimates of $k_{ref}, k_{tar}, c_{ref}$ and c_{tar} were obtained through fitting univariate Weibull distributions to the short-term wind speed observations at the target and reference sites and these were used with an initial value of D = 0.5 to minimise $\ln L$ with respect to D only, (ii) these starting parameters were used for a second minimisation search with respect to all five parameters to obtain the final fitted BW distribution, $f(u_{ref}, u_{tar})$. The predicted long-term target site wind speed distribution $f_{long}(u_{tar})$ was then obtained using Equation 7.3.

As an alternative to the MML described above, a second approach was also implemented. Final estimates of $k_{ref}, k_{tar}, c_{ref}$ and c_{tar} were extracted through univariate Weibull fits to the short-term reference and target site wind observations. The association parameter D was then obtained using the relation between d and the covariance of u_{ref} and u_{tar} given by Lu and Bhattacharyya [158]:

$$cov(u_{ref}, u_{tar}) = c_{ref} c_{tar}\left[\Gamma\left(\frac{D}{k_{ref}} + 1\right)\Gamma\left(\frac{D}{k_{tar}} + 1\right)\Gamma\left(\frac{1}{k_{ref}} + \frac{1}{k_{tar}} + 1\right)\right.$$

$$-\Gamma\left(\frac{1}{k_{ref}}+1\right)\Gamma\left(\frac{1}{k_{tar}}+1\right)\Gamma\left(\frac{D}{k_{ref}}+\frac{D}{k_{tar}}+1\right)\Bigg] \div \Gamma\left(\frac{D}{k_{ref}}+\frac{D}{k_{tar}}+1\right)$$

Equation 7.6

where Γ is the gamma function.

Equation 7.6 was solved numerically to obtain an estimate for D with the restriction $0 < D \leq 1$. This approach allows all five parameters to be obtained without fitting the full two-dimensional distribution. This modified technique is referred to as BW2 in the following discussion. As with the BW approach, $f_{long}(u_{tar})$ was obtained using Equation 7.3.

To determine the statistical parameters of $\bar{u}$, $\bar{p}_d$, σ and k which describe the predicted wind resource, 10^6 random wind speed samples were drawn from the predicted $f_{long}(u_{tar})$. These were used to calculate the error metrics used to assess the success of the approach. Since the angular dependent upwind roughness can affect the scaling between the reference and target site wind speeds [16], the BW approach was implemented using wind data binned into 90º angular sectors with respect to the reference site wind direction, except when investigating the convergence efficiency (Section 7.3.1) where no binning was applied. A sector width of 90º was chosen based on preliminary tests between 30º and 360º. These tests indicated that 90º sectors provided a reasonable balance between obtaining sufficient data in each sector over a range of training periods and accurately capturing the angular variability in the reference/target site correlation. For training periods where there were less than 80 observations within an angular bin, the fitted BW parameters behaved erratically and hence the data from the full range of angles were used. The sector approach results in four predicted wind speed distributions, one for each 90º angular sector. Hence, the final values of $\bar{u}$ and $\bar{p}_d$ were obtained as a weighted sum of the predicted values for each sector, with the weighting defined by the long-term, observed reference site frequency in that sector, using the expression:

$$\bar{u} = \sum_{k=1}^{4} f_k . \bar{u}_k$$

Equation 7.7

where f_k is the long-term fractional frequency of wind directions from the k^{th} angular sector, and $\bar{u}_k$ is the long-term predicted mean wind speed for the k^{th} sector. A similar expression was used to obtain $\bar{p}_d$. This allowed the overall values

of predicted $\bar{u}$ and $\bar{p}_d$ to be compared with the observed values across all sectors. In the case of the predicted distribution parameters σ and k, an estimate was obtained for each sector and these were compared with the observed values on a sector-by-sector basis.

The sector width of 90º was chosen based on the performance of the BW approach for sector widths of between 30º and 360º. Note that using the sector approach the final wind speed distribution will be a linear combination of the predicted Weibull distributions for each sector, a so-called Weibull mixture distribution. This increases the flexibility of the approach and allows for cases where the wind regime is best described by a Weibull mixture distribution (Section 2.3.2), as has been proposed by several authors [50, 51].

7.2.3 Generation of artificial wind speed data

In addition to the long-term observed wind data at multiple sites, which is crucial to investigating the performance of the BW approach, samples of artificial data drawn from specified BW distributions were also used. The purpose of using additional artificial data was (i) to validate the proposed theoretical framework for BW-based MCP (ii) to investigate differences in the fitting efficiency of the BW distribution using real and idealised data, and thereby infer how observed data differs from idealised BW distributions and (iii) to investigate to what extent conclusions based on artificial data may be extrapolated to real observations.

Samples of artificial wind data drawn from specified BW distributions were generated using distribution parameters extracted from BW fits to observed wind data at each reference/target site pair. Hence they can be considered as idealised BW versions of the real wind speed observations covering the same range of shape k, scale c, and association D parameters as the observed data. Where the characteristics of the observed wind data are close to ideal BW distributions, it would be expected that the fitting efficiency and prediction errors for both observed and artificial wind data should be comparable. Similarly, large differences in the fitting efficiency and error metrics between observed and artificial wind data are an indication that the idealised representation may be insufficient. Ultimately, the artificial wind data can be thought of as mimicking the results of a short-term measurement campaign at two correlated sites with an ideal BW distribution, thus providing a first step to validating the methodology.

The wind speed samples were constructed using an approach previously reported by Lu and Bhattacharyya [158] and others [159, 160]. Firstly, correlated, artificial,

random variables that represent n pairs of concurrent wind speeds at the two sites are denoted as ($\mathbf{x} = [x_1, x_2, \ldots x_n]$, $\mathbf{y} = [y_1, y_2, \ldots y_n]$) and these are written in terms of the independent random variables ($\mathbf{v} = [v_1, v_2, \ldots v_n]$, $\mathbf{w} = [w_1, w_2, \ldots w_n]$) for the i^{th} pair using the expressions [158]:

$$x_i = v_i^{D/k_x} w_i^{1/k_x} c_x$$

Equation 7.8

$$y_i = (1 - v_i)^{D/k_y} w_i^{1/k_y} c_y$$

Equation 7.9

where k, c and D are the BW distribution parameters defined previously, $\mathbf{v}$ is a random variable distributed uniformly in the interval [0,1] and $\mathbf{w}$ has an exponential and gamma mixture pdf given by [158]:

$$f(w) = (1 - D + wD)\ \exp(-w),\ \ w > 0$$

Equation 7.10

Using the method of Johnson *et al*. [159], the following procedure was then used to generate random samples from the BW distribution. First, five random variables ($\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4, \mathbf{s}_5$) were generated in the interval [0,1] along with the assignments $\mathbf{v} = \mathbf{s}_1$ and:

$$\mathbf{w} = \begin{cases} -\ln(\mathbf{s}_2) - \ln(\mathbf{s}_3), & \text{if } \mathbf{s}_5 \leq D \\ -\ln(\mathbf{s}_4), & \text{if } \mathbf{s}_5 > D \end{cases}$$

Equation 7.11

After defining the variables ($\mathbf{v}, \mathbf{w}$), artificial wind speed samples ($\mathbf{x}, \mathbf{y}$) were generated with the desired distribution parameters using Equation 7.8 and Equation 7.9. Artificial data sets representing 11 years of hourly wind speed entries were generated for each of the 22 site pairs considered in this study using distribution parameters extracted from BW fits to the observed long-term data records. These were used for comparing the performance of the BW approach using artificial versus real wind data.

7.2.4 Baseline MCP approaches

To assess the utility of the BW approach, its success was compared with two of the linear techniques that were implemented in Chapter 5, namely linear regression with Gaussian scatter (LR2) and the variance ratio method (VR). The techniques are described in more detail in Chapter 3, Section 3.2.3.1 and Section 3.2.2.3.

These techniques were chosen based on their performance in Chapter 5 and the fact that they are widely used in the literature as baseline approaches against which more complex techniques are compared [106, 109, 110]. Note that frequently, new MCP approaches are compared against simple linear regression with no representation of the residual scatter. However, as demonstrated in Chapter 5, this is a somewhat unfair comparison due to the large gains in accuracy that can be achieved through modelling the residual scatter. Hence in this work, the baseline linear regression approach also includes a representation of the residual scatter.

As in Chapter 5, each of the baseline approaches was applied sector-wise to data which were first binned according to the reference site wind direction. Angular sectors of width 30º were used and for angular bins with less than 20 data entries, the regression parameters were obtained by applying a global fit to data from all bins.

7.2.5 Meteorological measurements and error metrics

For consistency and to allow comparison between the different techniques discussed in this work, the MCP approaches were implemented using long-term wind speed and direction data from the same group of 22 target sites and 15 reference sites used in Chapter 5 and Chapter 6. Robust statistics averaged over all years and training seasons were calculated using the sliding window approach detailed in Section 5.2. The error metrics of mean absolute error (MAE), mean bias error (MBE) and absolute percentage error (%Error) were used, as described by Equation 5.2 to Equation 5.4. One difference should be noted here with respect to the distribution parameters of σ and k. As described in Section 7.2.2, when using a sector approach these cannot easily be combined into a single value. Hence, the errors for these parameters were estimated on a sector-by-sector basis by comparing the predicted and observed values for each 90° angular sector separately. The errors were then combined into a single value using a weighted sum, where the weighting factor was obtained from the long-term frequency of observations in each sector.

Clearly, a requirement for the application of a BW approach is that the wind speeds at the target and reference sites should be adequately described by univariate Weibull distributions. This was assessed by calculating the wind power using both the observed data and the fitted univariate Weibull distributions over the entire 11 year data record. The Weibull fits were applied using four 90º angular sectors as described previously. The average and maximum differences in estimated wind power were 1.3% and 4.6% respectively, with a difference of less than 2.0% at the

majority of sites. This, along with a visible inspection of the Weibull fits, indicated that the data were reasonably well represented by Weibull distributions.

7.3 Results and Discussion

7.3.1 Convergence efficiency of the bivariate Weibull parameters using artificial versus observed wind data

The efficiency with which the fitted BW parameters converged with respect to the sample length was compared using observed versus artificial wind data in terms of both the precision and the accuracy of the fitted parameters. To investigate the fitting efficiency, four reference/target site pairs (one from each terrain type) were chosen, along with their associated artificially generated wind data, for detailed investigation. Since similar trends were observed for each of the four site pairs, the results of a single site pair Rf4/R3 located in open, flat terrain are presented here.

The five parameters associated with the fitted BW pdf for the two sites were first determined using MML as described in Section 7.2.2 using the full 11 year data record. The extracted parameters were k_{ref} = 2.04, c_{ref} = 6.01, k_{tar} = 1.96, c_{tar} = 3.98 and D = 0.48. These parameters were used as inputs to create samples of artificial data from the same BW distribution using the method described in Section 7.2.3. To compare the fitting efficiency for the artificial and observed wind data, MML was used to extract the five BW parameters using progressively increasing sample sizes of observed and artificial data. A step size of 24 data points was used, representing one day of hourly averaged wind speeds.

Here, the artificial data are sampled randomly from the specified distribution, hence, for each sample of a particular size, the fitted BW parameters will vary until the sample size is large enough for the parameters to converge. In the case of the observed wind data, a real wind measurement campaign was replicated by choosing samples of consecutive wind data thus introducing additional variability due to seasonally varying atmospheric conditions. The variability in the extracted parameters was investigated using a Monte Carlo approach, whereby for each sample size the fitting procedure was repeated using 200 trials. In the case of the artificial data, the 200 trials were generated randomly from the required distribution, in the case of the observed wind data, the 200 trials corresponded to consecutive observations with random starting points throughout the 11 year data record, thus replicating measurement campaigns initiated at different times throughout the year.

The Monte Carlo approach was used to extract the predicted mean and standard deviation for each distribution parameter and sample size.

Figure 7.1 shows the results of this procedure for the BW target site parameters of k_{tar}, c_{tar} and D. The magnitude of the standard deviation across the 200 trials for each sample size may be used as an indicator of the degree of precision. A large standard deviation indicates that the value of the fitted parameter is dependent on the exact locations of the samples. Hence, a higher fitting efficiency is associated with a more rapid reduction in the standard deviation (or equally a more rapid increase in precision) as the sample size increases. For all three parameters, Figure 7.1 shows that the fitting efficiency, in terms of precision, is considerably greater when using artificial wind data compared to observed wind data. In the case of the observed data, seasonal variations in both the wind speeds and directions are likely to impact the form of the BW distribution leading to the large variations across different trials. Hence, significantly longer data samples may be required to extract precise distribution parameters when using observed wind data compared to artificial data.

In addition to the precision of the fitted parameters, the mean values from the Monte Carlo averaging, are also of interest since they represent the accuracy of the fitting. Figure 7.1 shows that for the artificial samples, the mean parameter values reach the true distribution values with a sample size of just a few days. In the case of the observed wind data, however, there is a large over estimation in the mean value of k_{tar} when using small samples. An increased value of k_{tar} indicates a narrower wind speed probability distribution, likely due to ‘clumping‘ of wind speeds in a relatively narrow range related to seasonal weather patterns. Similarly, the observed wind data results in an over estimation of the mean fitted value of d when using small samples, indicative of poor correlation between the two sites. In contrast, the mean value of c_{tar}, related to the target site mean wind speed, remains close to the true distribution value even for small samples of observed data. This is not surprising since c_{tar} is directly related to the mean wind speed which can be accurately determined from many snapshots of concurrent wind speed observations taken across multiple years (the Monte Carlo approach). For observed sample lengths of around 40 days, the mean fitted parameters are relatively close to the true distribution values. However, the large standard deviation indicates that the extracted parameters lack precision, with large variations possible depending on the measurement season. Similar trends were observed in the fitted parameters of k_{ref} and c_{ref}.

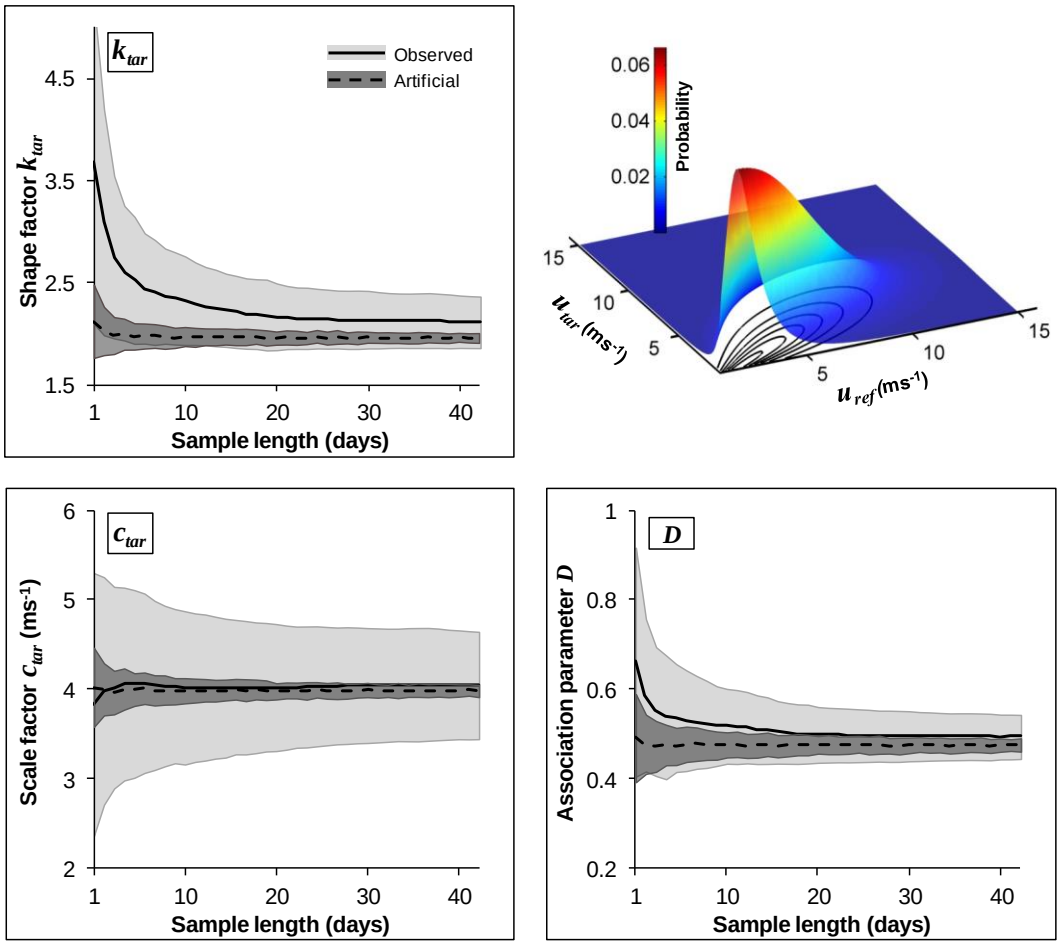

Figure 7.1: Variation in the fitted BW parameters of k_{tar}, c_{tar} and D using artificial (dotted line, dark shading) and consecutively sampled observed (solid line, light shading) wind data from a single reference/target site pair. The lines indicate a mean value averaged across 200 trials, the shading represents +/- one standard deviation from the mean. The inset shows the full BW probability surface.

To investigate if these results were related to seasonal effects, the Monte Carlo procedure was repeated using random, rather than consecutively sampled wind speed observations. Using this approach, concurrent pairs of wind speed observations at the reference and target sites were drawn at random throughout the 11 year data record. This random sampling procedure removes the effect of seasonal weather patterns and mirrors more closely the random sampling of artificial wind data.

Figure 7.2 shows the results of this procedure for the BW parameters of k_{tar} and c_{tar}. The mean and standard deviation of k_{tar} and c_{tar} follow almost identical trends using the artificial and observed wind data with rapid convergence of both the Monte Carlo mean value and the standard deviation. Similar trends were observed for the remaining three BW parameters, indicating that it is the restriction of consecutive sampling, and most likely the associated seasonal weather patterns,

which result in the loss of fitting efficiency when using observed rather than artificial wind data.

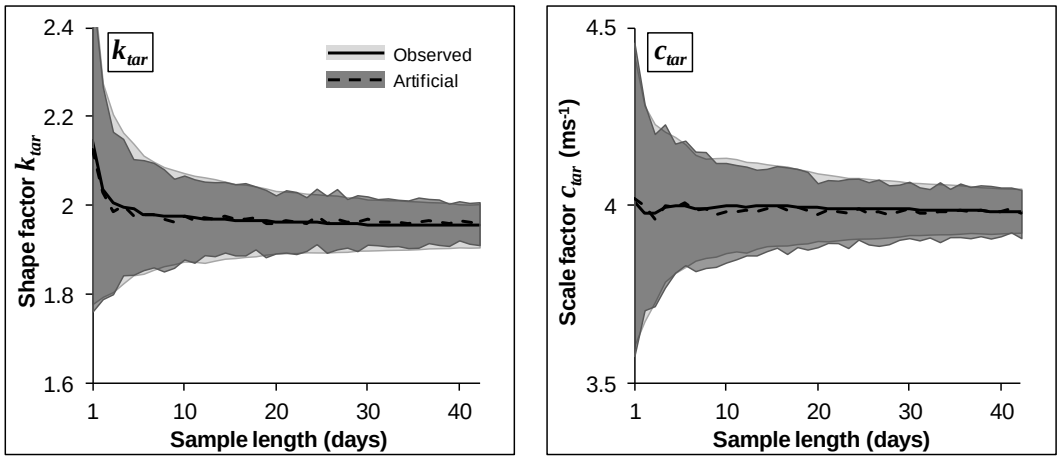

Figure 7.2: Variation in the fitted BW parameters of k_{tar} and c_{tar}, using artificial (dotted line, dark shading) and randomly sampled observed (solid line, light shading) wind data from a single reference/target site pair. The lines indicate a mean value averaged across 200 trials, the shading represents +/- one standard deviation from the mean.

These results highlight some important factors related to the implementation of the BW approach to observed wind data. Firstly, the convergence time is likely to be significantly longer than in the case of artificial data as highlighted by Figure 7.1. This could result in relatively large errors in the estimated parameters when using short data periods. Secondly, assuming these results can be generalised, the values of the parameters k and D may be overestimated on average, when using short data periods. Note that when conducting a measurement campaign, consecutive sampling of wind speeds is the most likely approach due to the time and expense of installing a meteorological mast. However, with the improvement in portable measurement devices such as LiDARs (light detection and ranging), and where multiple sites are to be investigated, a non-consecutive sampling approach such as the 'Round Robin' procedure proposed by Lackner *et al*. [111] is a viable alternative, as discussed in Section 3.2.5. While this approach is not equivalent to random sampling, it may be possible to develop a procedure which captures seasonal variability sufficiently well to enable improved fitting efficiency.

A final observation is noteworthy regarding the two methods (BW and BW2) outlined in Section 7.2.2 for extracting the distribution parameters. Using the alternative approach (BW2), for the four sites considered, it was found that the extracted values of k_{tar}, c_{tar}, k_{ref} and c_{ref} were almost identical (within ~1.5%) using both approaches. However, the BW2 approach resulted in consistently lower estimates of D (by around 10% - 40%) compared to MML. This suggests that

estimates of D based on the covariance are associated with a higher predicted correlation between the reference and target site wind speeds. Interestingly, when applied to the artificial wind data this difference almost vanished indicating that the effect may be due to deviations of the real wind data from idealised bivariate Weibull distributions.

7.3.2 Comparison between the bivariate Weibull and baseline MCP approaches

7.3.2.1 Percentage error metric

To compare the success of the BW and BW2 approaches with the existing MCP methods of LR2 and VR, each approach was applied to observed and artificially generated wind data for the 22 site pairs in order to predict the 10 year wind resource. Figure 7.4 shows the percentage error metrics for $\bar{u}$ and $\bar{p}_d$ using the artificially generated data averaged across all 22 site pairs and all 120 sliding window positions for training lengths of 1-12 months. Note that the 22 artificial wind data sets represent idealised BW versions of the observations at the 22 site pairs since they were generated using Weibull shape and scale parameters extracted from the observed data. Figure 7.4 shows that the BW approach clearly performs better than the regression approaches for all training lengths, in line with the observations of Perea *et al*. [110]. Equivalent trends were also observed for σ and k.

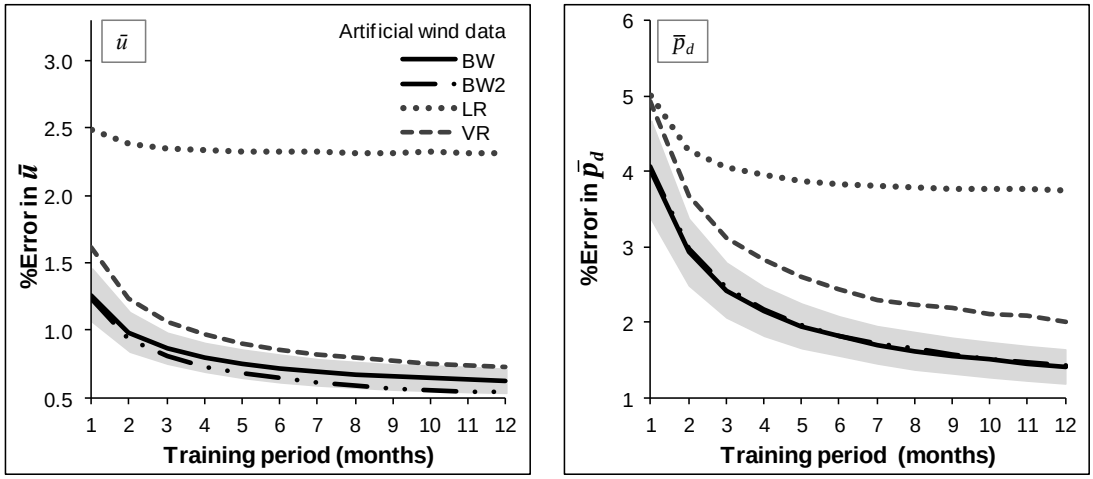

Figure 7.3: Percentage error metrics as a function of training period for the wind resource parameters of $\bar{u}$ and $\bar{p}_d$ using artificially generated wind data. Lines show the mean value averaged across 22 site pairs and 120 starting months. The shaded region represents +/- one standard deviation in the mean for the BW approach.

Figure 7.4 shows the equivalent error metrics for $\bar{u}$, $\bar{p}_d$, σ and k using observed wind data for all 22 site pairs. Note that applying the sliding window approach to

observed wind data ensures that the average error metrics are independent of the season or year in which the short-term measurements were taken, while the standard deviation of the percentage errors (shading Figure 7.4) indicates the magnitude of the intra- and inter-annual variations.

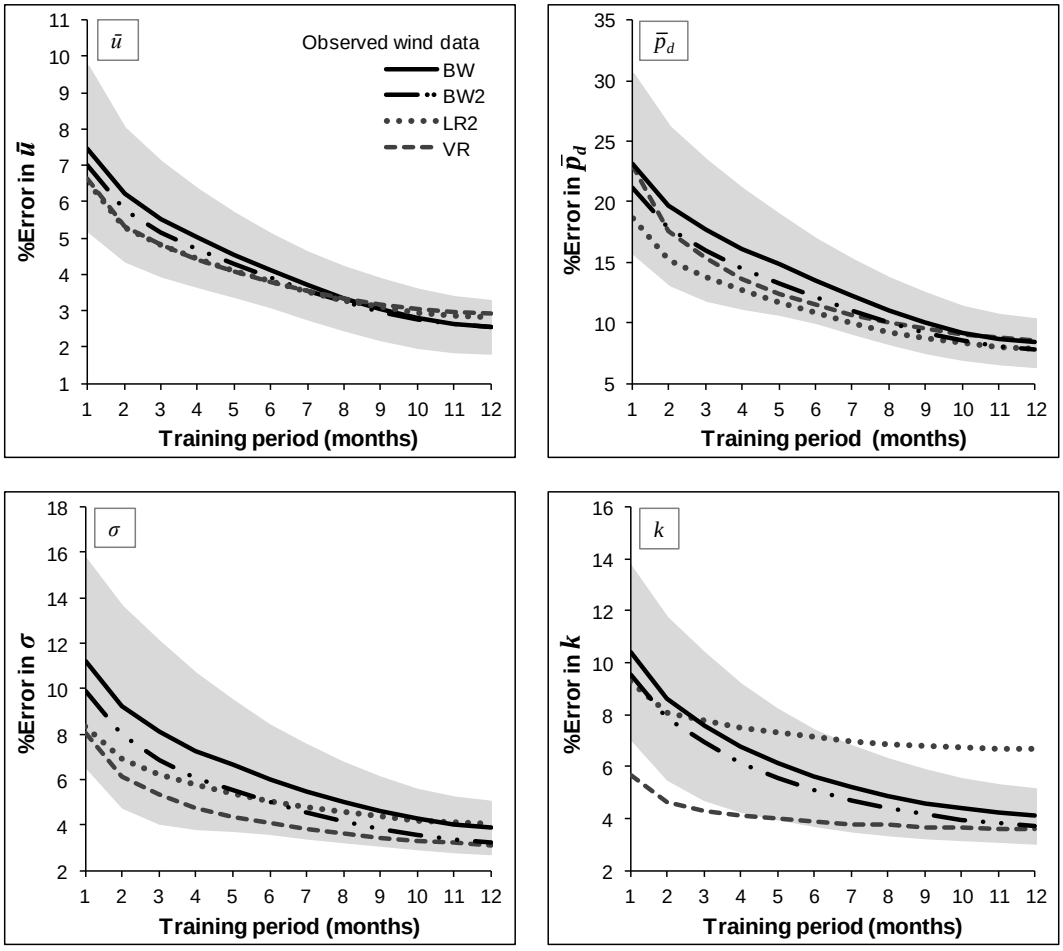

Figure 7.4: Percentage error metrics as a function of training period for the wind resource parameters of $\bar{u}$, $\bar{p}_d$, σ and k using observed wind data. Lines show the mean value averaged across 22 site pairs and 120 starting months. The shaded region represents +/- one standard deviation in the mean for the BW approach.

It is immediately apparent that the error metrics behave quite differently when the MCP approaches are applied to observed wind data. Generally, for short training periods, one or more of the regression approaches results in lower percentage errors than either BW or BW2. Using a full 12 month training period, the BW2 approach performs as well as the best regression approach in terms of the percentage error in $\bar{p}_d$, σ and k and slightly better than the best regression method in terms of $\bar{u}$. It is of interest that for training periods less than 8 months, the relatively simple LR2 method consistently performs as well or better than the other approaches in predicting $\bar{u}$ and $\bar{p}_d$, while for longer training periods all the MCP

approaches tend to converge. For the parameters σ and k which describe the form of the wind speed distribution, the VR approach performs better than the other approaches at short training periods converging with BW2 at longer training periods. For all four parameters, the percentage error metric is notably lower for the BW2 approach compared to BW, particularly at short training periods. Since, as discussed previously, the BW2 approach only differs in the estimation of the D parameter, this suggests that the reference/target site covariance provides a more suitable indicator for this parameter compared to MML.

These results indicate that when using real wind data, the MCP approaches of BW and BW2 do not consistently produce more accurate predictions compared to regression approaches despite their stronger theoretical basis. This is in contrast to results obtained when using artificial wind data (Figure 7.3) and could be due to deviations of the observed wind data from idealised BW distributions, as well as the difficulty in accurately extracting the BW parameters, particularly for short measurement periods. It should be noted that the LR2 approach implemented here includes a Gaussian model of the scatter term ε about the predicted wind speeds, which was shown in Chapter 5 to increase the accuracy of predictions [161]. Without this term, the LR2 method would be considerably less competitive with the BW and BW2 approaches.

The standard deviation in the %Error metrics (shaded region, Figure 7.4) indicates the degree of variability using different training periods, averaged across all sites. As mentioned previously, since overlapping training periods are used (Section 5.2.1) the errors will be correlated and hence +/- two sigma cannot strictly be interpreted as a 95% confidence interval. At training periods of less than one year, the error includes contributions from both seasonal and inter-annual variability in the predictions. While seasonal effects add considerably to the variability in the %Error metric, it is noteworthy that the inter-annual variability at 12 months is still relatively large ($\sigma_{(\%Error\ in\ \bar{u})}$ = +/- 0.7% and $\sigma_{(\%Error\ in\ \bar{p}_d)}$ = +/- 3.2%). This highlights the importance of using multiple training years when testing the performance of new MCP algorithms, even if the training period is a full year. Assessing performance using error metrics calculated from a single training period or training year, as is sometimes reported in literature [121, 125], may lead to erroneous conclusions.

7.3.2.2 Distribution of percentage errors

Figure 7.5 shows the residual percentage error distributions across the 22 sites for a training length of 3 months using the BW and linear approaches averaged over all

120 sliding window positions. Since the errors are averaged across all seasons and and years, the distributions represent the average variability across different sites, as described in Section 5.2.5. The distributions for LR2 and VR were previously shown in Chapter 5 and they are included here for comparison with the BW approaches. A training period of 3 months is shown since measurement periods on these timescales are of interest to small-scale wind installations. Note that the conventional expression for the residual error $\varepsilon_{\%}$ is used, (Equation 5.5), hence a negative residual represents an overestimate and vice-versa.

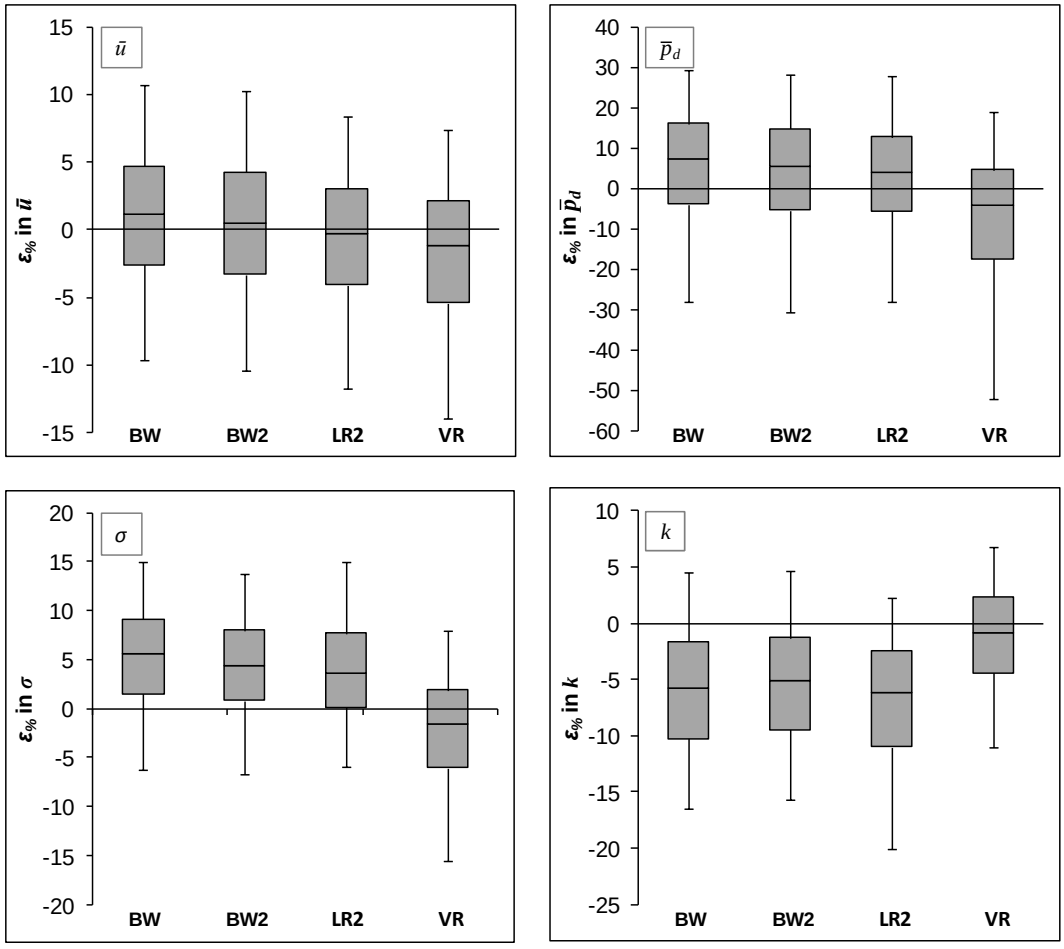

Figure 7.5: Residual percentage error distributions across 22 target sites for mean wind speed $\bar{u}$, mean wind power density $\bar{p}_d$, sample standard deviation σ, and Weibull shape factor k. The error bars represent the 5th and 95th percentiles, the shaded regions encloses the interquartile range.

It is noteworthy that the error distributions are relatively similar for the BW, BW2 and LR2 approaches, with all exhibiting similar bias and spread. For these approaches, the 95th percentiles cover approximately +/-10% in $\bar{u}$ and approximately +/-30% in $\bar{p}_d$. As with the LR2 approach, the BW and BW2 approaches can be seen to underestimate the width of the wind speed distribution

(underestimate σ and overestimate k) resulting in an underestimate of $\bar{p}_d$ when using 3 month training periods.

7.3.2.3 Mean bias errors

In addition to the percentage error, it is also informative to consider the mean bias error (Equation 5.4) which describes the average tendency to overestimate or underestimate a particular parameter. Figure 7.6 shows the MBE metrics averaged across 22 site pairs and all years and seasons using observed data. For a full 12 month training period, the BW approach results in the lowest bias in $\bar{u}$. However, in terms of $\bar{p}_d$, BW2 performs best, closely followed by LR, BW and VR. Note that while BW2 and LR slightly overestimate $\bar{u}$, these approaches also underestimate the width of the wind speed distribution, as indicated by the MBE in σ and k, and these two effects may offset each other resulting in a low net negative bias in $\bar{p}_d$. VR exhibits a very small bias in σ and k and hence the positive bias in $\bar{p}_d$ is a more direct reflection of the positive bias in $\bar{u}$ using this approach. As suggested from the analysis of artificial data in Section 7.3.1, both BW and BW2 tend to overestimate k, especially for short training periods. The behaviour of the bias error across these parameters reveals that the resulting errors in $\bar{p}_d$ are due to a relatively complicated combination of factors, including possible cancellation of errors. Despite these complications, the bias errors are generally small across all MCP approaches for training periods of 12 months, with differences becoming more apparent at shorter training periods.

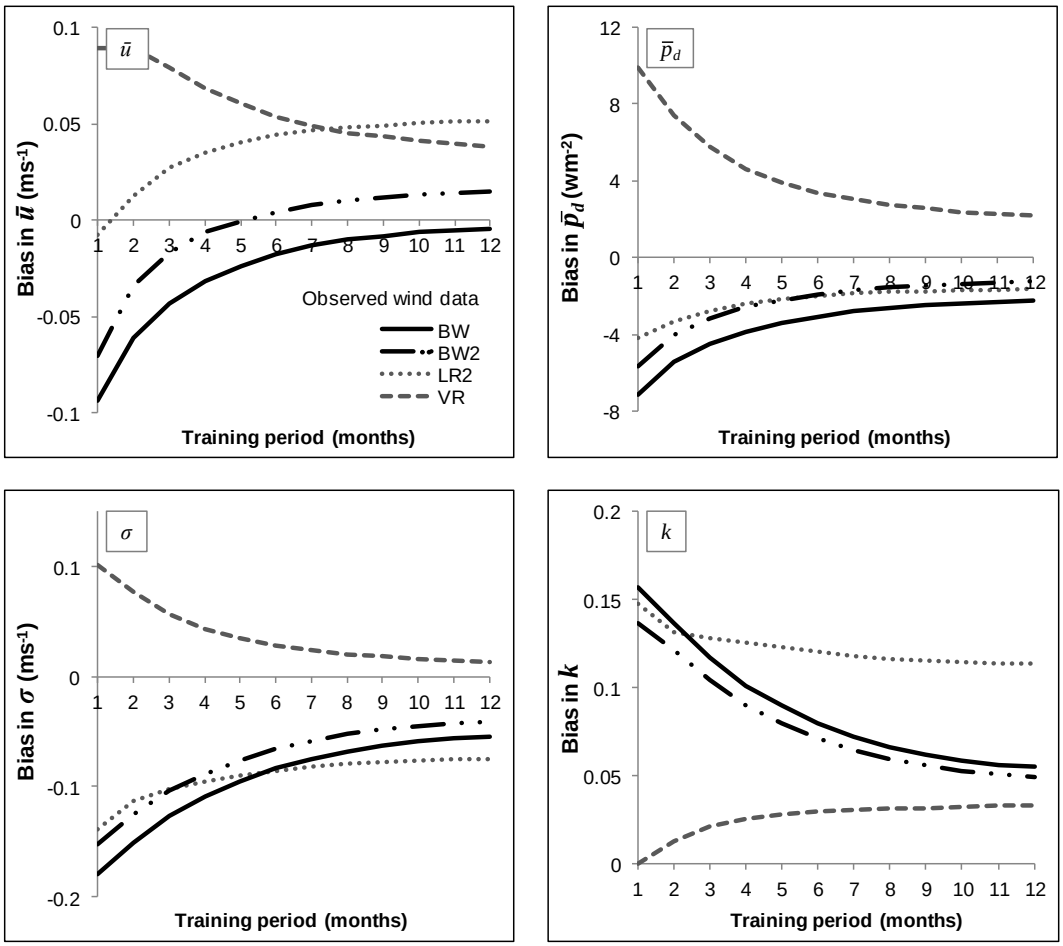

Figure 7.6: MBE metrics as a function of training period for the wind resource parameters of $\bar{u}$, $\bar{p}_d$, σ and k using observed wind data. Lines show the mean value averaged across 22 site pairs and 120 starting months.

The total variability in the bias errors, including intra- and inter-annual as well as site-specific variability was considered for the LR2 approach in Section 5.3.3. A similar analysis was performed here for the BW2 approach. The 2σ values (standard deviation of the bias errors across all sites and years) were calculated as 0.63 ms^{-1} and 42 Wm^{-2} for $\bar{u}$ and $\bar{p}_d$ respectively, using a 3 month training period. These are similar to, although slightly higher than, the values obtained for LR2 of 0.58 ms^{-1} and 36 Wm^{-2} for $\bar{u}$ and $\bar{p}_d$ respectively. As observed in Chapter 5, the error distributions were also found to be leptokurtic for the BW2 approach highlighting the need for caution in the interpretation of these values.

Table 7.1 summarises the metrics of %Error, MAE and MBE for training periods of 3 months and 12 months using the observed wind data. At 12 months, the performance of all four MCP approaches is very similar with BW2 performing very slightly better on average than the remaining approaches. For completeness, the BW approaches were also tested using training data lengths of 18 and 24 months. To implement these longer training periods, the training window (Section 5.2.1) was

extended to 2 years and the test period was reduced from 10 to 9 years. Only small decreases in the error metrics were observed for these training periods, (for example, using BW2 with a 24 month training period, %Error in $\bar{u}$ = 2.4% and in $\bar{p}_d$ = 7.1%), indicating that longer training periods are unlikely to significantly improve the performance of these approaches.

For a shorter training period of 3 months there are clearer differences between MCP approaches with the regression techniques of LR and VR generally resulting in smaller errors than the BW techniques. One possible reason for this is that the BW approach requires a greater number of parameters to be extracted from the training data compared to regression approaches, thus requiring longer training periods. In addition, even if the wind speeds at a site are well described by a Weibull distribution, there is no guarantee the wind speeds observed during a relatively short training period will also follow a Weibull distribution. Overall, average errors are approximately halved by increasing the training period from 3 to 12 months.

3 M	Method	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	BW	5.5	18	8.1	7.6
	BW2	5.2	16	6.9	6.9
	LR2	4.8	14	6.2	7.8
	VR	4.8	15	5.3	4.3
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	BW	0.25	15	0.19	0.15
	BW2	0.23	13	0.17	0.14
	LR2	0.21	11	0.16	0.14
	VR	0.21	11	0.13	<0.1
MBE	BW	<0.1	-4.5	-0.13	0.12
	BW2	<0.1	-3.2	-0.10	0.10
	LR2	<0.1	-2.8	-0.10	0.13
	VR	<0.1	5.7	<0.1	<0.1

12 M	Method	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	BW	2.6	8.4	3.9	4.1
	BW2	2.6	7.8	3.2	3.7
	LR2	2.8	7.9	4.0	6.7
	VR	2.9	8.5	3.1	3.6
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	BW	0.11	6.1	<0.1	<0.1
	BW2	0.11	5.7	<0.1	<0.1
	LR2	0.12	5.8	0.10	0.12
	VR	0.12	6.1	<0.1	<0.1
MBE	BW	<0.1	-2.3	<0.1	<0.1
	BW2	<0.1	-1.2	<0.1	<0.1
	LR2	<0.1	-1.6	<0.1	0.11
	VR	<0.1	2.2	<0.1	<0.1

Table 7.1: Error metrics for the wind resource parameters of $\bar{u}$, $\bar{p}_d$, σ and k using training periods of 3 months (left) and 12 months (right) averaged across 22 target sites and 120 starting months.

7.3.2.4 Seasonal effects

Seasonal effects, using 3 month training periods, were considered in detail for the LR2 and VR approaches in Section 5.3.4. Here, the seasonal variations in the prediction errors using the BW and BW2 approaches are briefly compared with those associated with the LR2 approach. Figure 7.7 shows the variation in the average percentage errors for $\bar{u}$, $\bar{p}_d$, σ, and k using different three month training periods throughout the calendar year. As before, the vertical lines mark training periods corresponding to the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug).

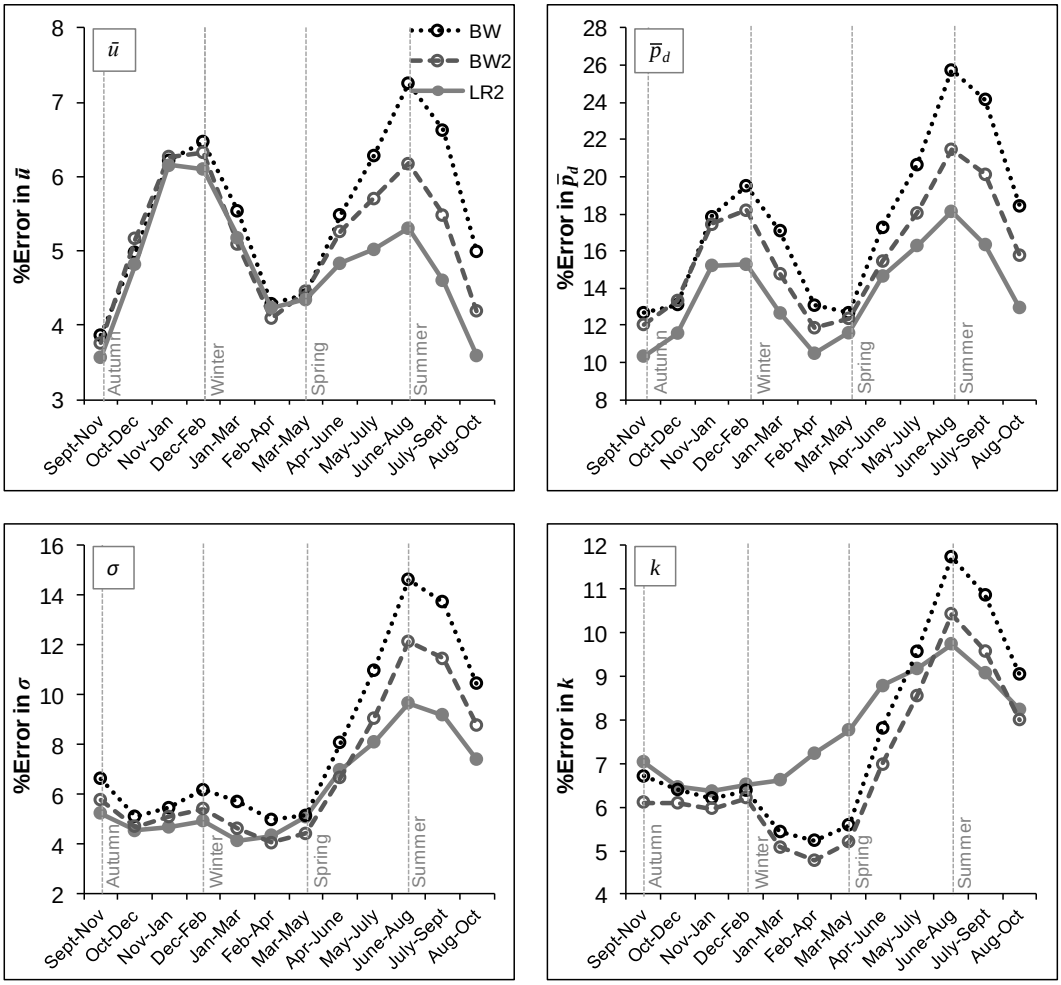

Figure 7.7: Seasonal variation of the percentage errors in mean wind speed $\bar{u}$, mean wind power density $\bar{p}_d$, sample standard deviation σ and Weibull shape factor k averaged across 22 target sites and 10 years using three MCP approaches. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training.

The seasonal variations in the %Error metrics follow very similar trends for the BW and linear LR2 approaches. However, the error peak during the summer months is more exaggerated for BW and BW2 compared to LR2. Since the BW approaches rely more strongly on accurately extracting the distribution parameters, this effect could be associated with the decreased magnitude and variance of the wind speeds experienced during the summer months as discussed in Section 5.3.4.

Figure 7.8 shows the seasonal variation in the MBE for $\bar{u}$ and $\bar{p}_d$. Again, the BW and LR2 approaches show very similar trends with $\bar{u}$ and $\bar{p}_d$ overestimated in the winter and underestimated in summer. The degree of underestimation in these parameters is greatest for the summer months in line with the seasonal trends observed in the %Error metrics.

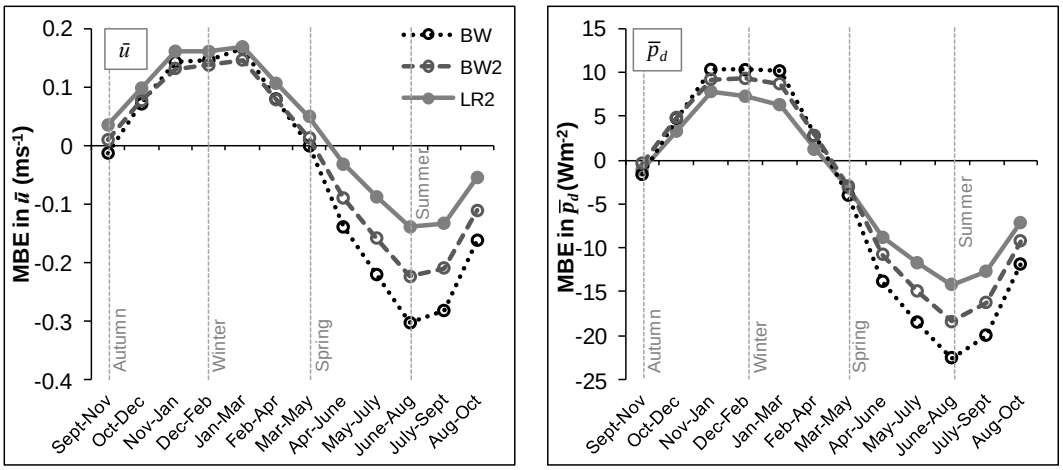

Figure 7.8: Seasonal variation of the mean bias error in mean wind speed $\bar{u}$ and mean wind power density $\bar{p}_d$, averaged across 22 target sites and 10 years using two MCP approaches. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training.

It is of interest that the seasonal trends observed in the error metrics are very similar for the BW and LR2 approaches despite the fact that they utilise very different formulations and assumptions. This indicates that the seasonal factors are more strongly related to the representativeness of the data (i.e. how well the short-measurement period captures the range of meteorological phenomena associated with the sites) rather than factors that are specific to the MCP algorithms used.

7.3.2.5 Errors at individual sites

Thus far, error metrics have been presented as averaged across all 22 target sites. In addition to these average statistics, the performance of the BW and regression approaches at each site is of interest. Figure 7.9 shows the %Error in $\bar{p}_d$ for each site using 3 months training and the best performing bivarate and regression approaches, namely BW2 and LR2. The metrics are averaged across all 120 window positions. The %Error metrics for BW2 and LR appear to follow very similar trends for both approaches, with LR2 generally performing slightly better than BW2. This implies that the performance of the bivariate and regression MCP approaches is more strongly dependent on the characteristics of the site rather than the MCP approaches used. Interestingly, Bass *et al*. [126] reached a similar conclusion when comparing MCP approaches based on ANNs to linear techniques.

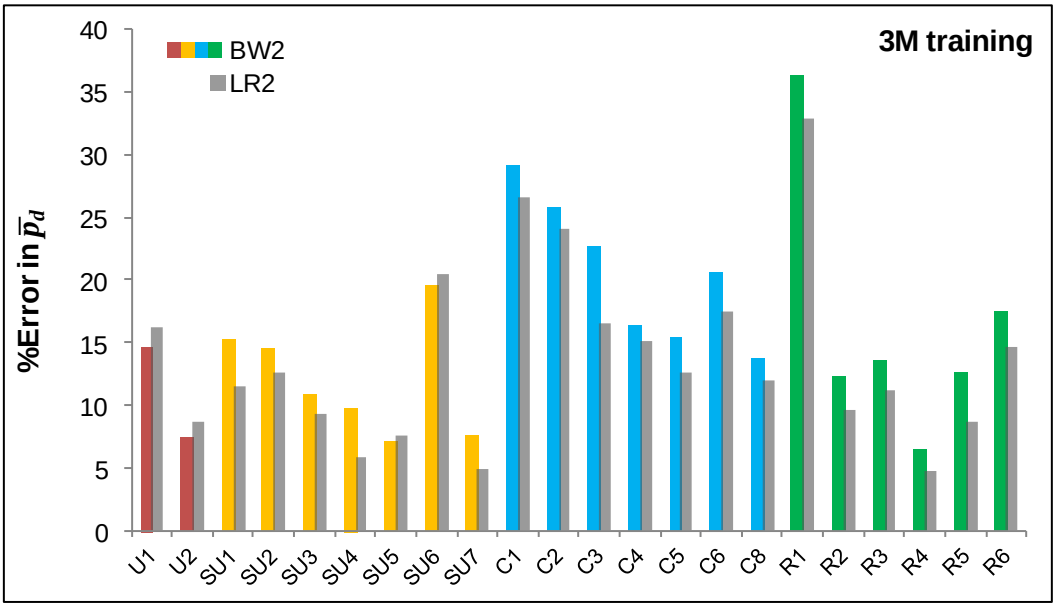

Figure 7.9: %Error in $\bar{p}_d$ for each of the 22 target sites using the BW2 and LR2 approaches and a training period of 3 months.

For completeness, Figure 7.10 shows the equivalent %Error in $\bar{p}_d$ using a full 12 month training period. Again, the %Error metrics for the BW2 and LR2 approaches follow very similar trends. However, for the longer training period the BW2 approach is more competitive with LR2 and the two methods outperform each other at an approximately equal number of sites. There is a small indication that the BW2 approach may be slightly more successful at coastal sites as well as at the two urban sites, however, these differences are relatively small. Similar trends were also observed for the %Error in $\bar{u}$. As mentioned previously, work by Clive [116] showed that wind speeds described by correlated Weibull distributions will exhibit non-linear correlations if the ratio of the Weibull shape factors at the two sites differs. However, for the sites considered in the current study, this metric was not found to be a reliable predictor of whether the BW approaches would outperform linear regression in terms of the %Error in $\bar{p}_d$ or $\bar{u}$.

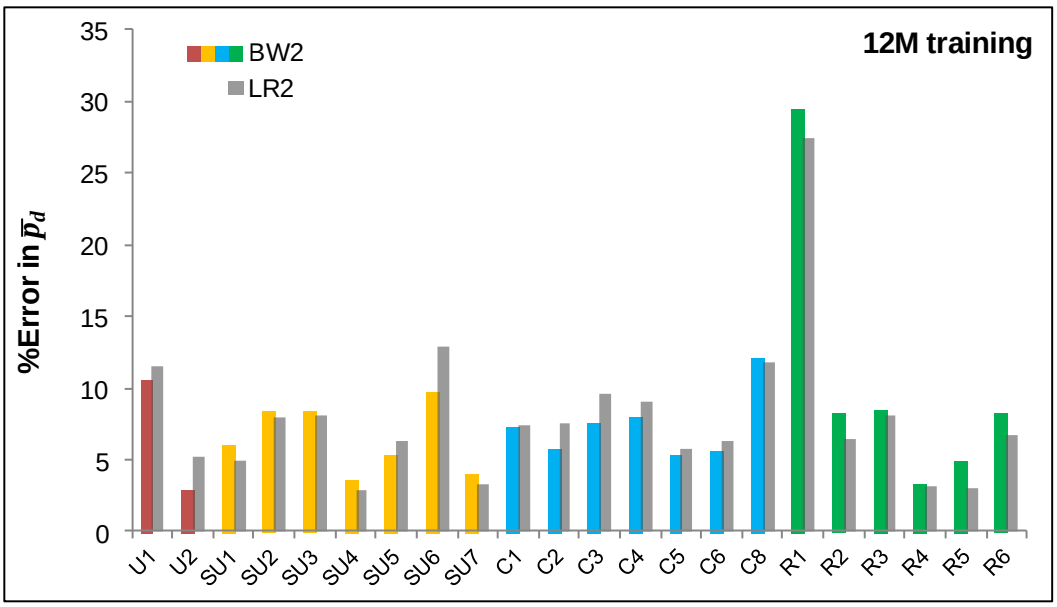

Figure 7.10: %Error in $\bar{p}_d$ for each of the 22 target sites using the BW2 and LR2 approaches and a training period of 12 months.

Since the BW approaches are based on the assumption that wind speeds at the reference and target sites follow univariate Weibull distributions, it is informative to investigate whether the performance of these approaches can be related to the accuracy of this assumption. To achieve this, some measure must be used to determine how 'Weibull-like' the wind speed distributions are at each site.

A number of goodness-of-fit tests are available which allow the null hypothesis (i.e. that the data and the model belong to the same statistical distributions) to be tested directly by means of a test statistic calculated from the observed and assumed distributions. These include:

- Kolmogorov-Smirnov test (estimates the maximum vertical distance between the observed cumulative distribution function (cdf) and the fitted cdf).
- Pearson's Chi-square test (based on the sum of differences between observed and predicted frequencies).
- Anderson-Darling test (related to the difference between the observed and predicted cumulative frequencies).

These tests have been applied to wind speed data by several authors [42, 48, 162] in order to compare different wind speed distributions and numerical methods for estimating distribution parameters. However, as pointed out by Ramirez and Carta, [38, 40] there is a fundamental issue in applying these tests to wind speed data since they assume the observed data are independent. This assumption is

generally not supportable on the timescales (≤1hr) typical of observed wind speed data.

A more direct metric for estimating how well the wind speed distributions conform to univariate Weibull distributions is to simply calculate the values of $\bar{u}$ and $\bar{p}_d$ using both the observed data and Weibull fits to the data. The percentage error in these parameters can be used to assess how well the data are described by univariate Weibull distributions. Using this approach, the percentage error in $\bar{p}_d$ will be more heavily weighted to the success of the fit at higher wind speeds due to the cubic relationship between wind speed and power. Note that the goodness of fit measure can be applied either to the reference or target site observations and either to the full 11 year data record or to each of the training periods individually. All these options were explored and all produced similar results. The goodness of fit using the full 11 year data record at the target sites is presented below. Figure 7.11 compares the %Error in $\bar{p}_d$ from implementation of the BW2 approach with the error in $\bar{p}_d$ from applying a Weibull fit to the target site wind speeds. As before, the Weibull fits were applied sector-wise using 90º angular sectors.

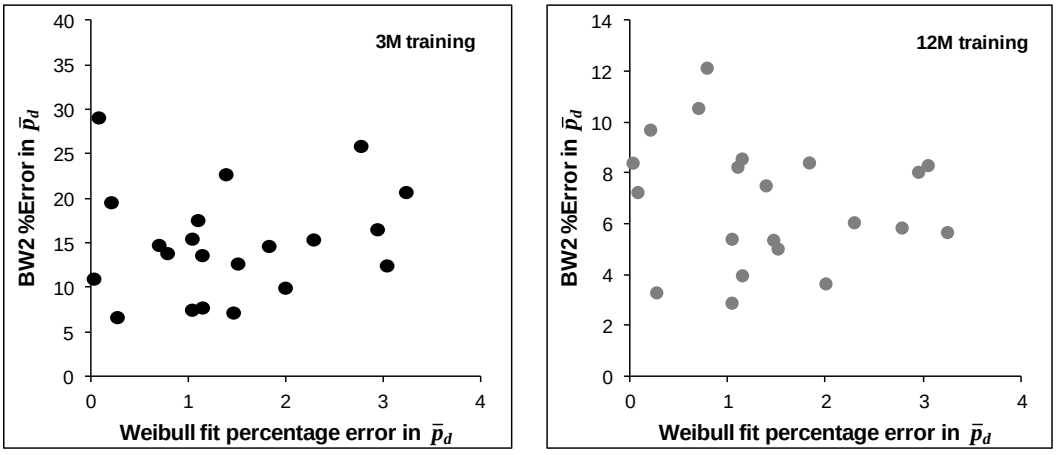

Figure 7.11: %Error in $\bar{p}_d$ for the 22 target sites using the MCP approach of BW2 versus the Weibull fit metric of percentage error in $\bar{p}_d$ using a 3 month (left) and 12 month (right) training period. An outlier has been removed (site R1) to improve clarity.

At training periods of both 3 and 12 months there appears to be no clear relationship between the success of the BW2 technique and the Weibull error. The equivalent relationships were also investigated using the percentage error in $\bar{u}$ as a Weibull fit metric. In all cases, no clear relationship emerged between the BW2 error and the success of the Weibull fit. These results indicate that the performance of the BW approaches is not especially sensitive to the accuracy of the univariate Weibull fits at the reference and target sites. However, as noted in Section 7.2.5,

the wind speed observations at the reference and target sites used in this study were all relatively well described by univariate Weibull distributions. Hence, it is possible that such trends would only emerge for sites with larger deviations from univariate Weibull behaviour. It should also be noted that although a successful BW fit to a reference/target site pair requires that each site be well described by a univariate Weibull distribution, this does not in itself guarantee that the sites will be well described by a correlated BW distribution.

7.4 Conclusions

An MCP approach based on modelling the underlying BW probability distribution of reference and target site wind speeds has been implemented at 22 pairs of UK sites using multiple test periods over an 11 year data record. Building on previous work that applied the technique to artificial wind data, a detailed comparison between the performance of the approach using observed and artificially generated data has been carried out. The results indicate that due to seasonal effects, the data period required for convergence of the extracted BW parameters is likely to be significantly longer when using observed compared to artificially generated wind data and that the Weibull shape factor k and association parameter D may be overestimated on average when using short measurement periods. In addition, estimating D from the covariance of the reference/target site wind speeds was found to result in improved performance across all error metrics compared to estimations based on MML.

The performance of the BW approach was compared quantitatively with two linear MCP methods using observed wind data at the 22 site pairs as well as artificial wind data generated from ideal BW distributions modelled on the same sites. In line with a previous study [110], the BW approach outperformed the linear approaches for all measurement periods when applied to artificial wind data. However, when applied to observed wind data, the regression approaches generally performed better than the BW approaches for short training periods, while all approaches performed similarly for training periods of 12 months. These results suggest that the improved performance of the BW approach when using artificial wind data may not be transferable to real wind observations. This is likely due to the fact that real wind speed observations deviate from idealised BW distributions, as well as the difficulty in accurately estimating the BW parameters from short training periods.

The MCP approaches considered so far (Chapter 5 – Chapter 7) have been based on applying relationships extracted from a short-term training period to long-term

meteorological measurements at a nearby reference site. However, in some cases, the use of nearby meteorological measurements may not be practical due to cost, the absence of a nearby meteorological mast, reliability of the data or differences in the wind climate at the reference and target sites. As alluded to previously, the latter issue is of particular importance at coastal sites, or other regions of complex terrain, where the wind climate may be highly localised and decoupling between reference and target site wind speeds may occur over relatively short distances. In the following chapter, the application of MCP using local operational forecast data in place of nearby reference stations is considered as a means to overcome some of these limitations.

8 Using Output from an Operational Forecast Model as a Tool for Small-Scale Wind Resource Assessment

8.1 Overview

In Chapters 4 - 7 various methods, both semi-empirical and data-driven, for small-scale wind resource assessment have been investigated. In the current chapter, attention is turned to the sources of reference data that may be used to drive these approaches.

For the implementation of data-driven MCP approaches, short-term wind data must first be obtained at the target site. In addition, suitable long-term data must be obtained from a nearby reference site with the same climatology in order to establish a correlation between the two sites and to make a long-term prediction at the target site. Procuring suitable reference data places an additional burden on the site assessment. For example, in some cases, nearby reference sites with similar climatology may simply not be available, data records may not be of sufficient length or quality control procedures may be inadequate. This will be particularly relevant in countries that do not have a long history of meteorological measurements. Similarly, for target sites in complex terrain or in coastal areas, even reference sites that are relatively close may not be suitable for long-term correlations due to highly localised phenomena as discussed in Section 2.4. In addition, the cost of obtaining long-term, high quality reference data may prove to be a further barrier in implementing a data-driven approach.

In the case of the semi-empirical boundary layer scaling model described in Chapter 4, the data requirements are relatively modest. The model is implemented with a single mean wind speed at a reference height. The mean wind speed is assumed to be representative of the regional (1 km) area, and includes contributions from regional variations in topography and average coastal effects. The wind speed databases (NCIC and NOABL) used in Chapter 4 are interpolated from surface wind measurements and are subject to uncertainties associated with the measurements and the interpolation scheme. Since the databases do not include information regarding the wind speed distribution, a generic distribution must be applied leading to further uncertainties in the predicted wind power. While this may not be the most significant source of error (Section 4.3), applying the boundary layer scaling model to a time series of wind speeds rather than a single

mean value is one way in which the error due to uncertainty in the wind speed distribution could potentially be reduced.

Given the issues outlined above, it is relevant to consider alternatives to surface wind observations as sources of reference data for both data-driven and boundary layer scaling approaches to resource assessment. With the ongoing development of numerical weather prediction techniques, global data sets of atmospheric variables including wind speed and direction are becoming widely available. Unsurprisingly, this has led to interest from the wind energy industry seeking convenient, low-cost alternatives to long-term reference data.

In this chapter, output from a state-of-the-art operational forecast model is investigated in terms of its ability to provide reference wind data for use in small-scale wind resource assessment. The data are predominantly investigated in terms of their ability to be used in place of long-term reference site observations in the MCP approaches outlined in Chapters 5-7. However, the data are also considered as an alternative source of reference climatology for the boundary layer scaling approach outlined in Chapter 4. The main objectives of this study are summarised below:

(i) An investigation of the ability of forecast data to represent long-term wind speed trends at the target sites.
(ii) Application of forecast data to MCP approaches in place of long-term reference site observations.
(iii) An investigation of the effect of forecast height on the performance of the MCP approaches.
(iv) Quantitative comparison between the performance of MCP approaches using meteorological observations and forecast data as a long-term reference source.
(v) Quantitative comparison between the performance of the boundary layer scaling approach using forecast data and the NCIC database as the reference climatology.

The chapter is organised as follows: Firstly, alternative sources of reference data are considered including low resolution reanalysis data and higher resolution output from an operational forecast model. Next, the Met Office Unified Model is considered in more detail and factors related to data independence, resolution and time alignment are considered. The ability of the forecast data to represent long-term trends is then investigated through the use of a wind index. Finally, forecast

data are applied to MCP and boundary layer scaling approaches to wind resource assessment and the performance of the data is compared against a baseline.

8.2 Alternative Sources of Reference Data

8.2.1 Reanalysis data

In the wind energy industry, there is increasing interest in the use of global reanalysis data, as well as derived data sets, as long-term reference data for the implementation of MCP. In a recent survey of European wind energy developers, it was noted that while reanalysis data are increasingly being relied upon in wind resource assessments, there is a lack of rigorous validation studies at research level. In addition, the surveyed developers rated the development of downscaling methods to relate large-scale wind databases to onsite conditions as their number one priority [163].

Reanalysis data are produced by numerical models that describe the evolution of atmospheric variables such as temperature, humidity, pressure and wind. The models assimilate large amounts of data including observations from satellites, weather balloons, aircraft, ships, buoys and surface meteorological stations, to produce a three-dimensional grid of modelled atmospheric variables. Reanalysis data are so termed because they represent a second analysis of the data following the primary use in real-time weather forecasts [164]. The second analysis allows incorporation of observations not available to the model when it is run in real-time [165], as well as offering a consistent assimilation and analysis model, which is not necessarily the case for archived real-time forecasts [166].

There are several reasons why reanalysis data are seen as an attractive source of long-term reference data in place of surface wind observations:

- Many data sets are in the public domain and data cover the majority of the globe thus reducing cost and providing reference data even where surface wind observations are sparse.
- The data cover a number of decades providing the opportunity for much longer correlations than may be possible using observed wind data.
- Compared to surface observations, which are typically taken at 10 m above ground level, the wind data are less affected by changes in land-use and local obstructions.

Commercial software tools have started to include direct access to reanalysis data sets for use as reference data within their MCP modules. The latest release of

WindPro (version 2.9, 2013) for example, includes high resolution (3 x 3 km spatial, 1 hour temporal) European reference data produced by a mesoscale model driven using input from a global reanalysis data set [157].

A range of global reanalysis products are currently available with various temporal and spatial resolutions. The Climate Data Guide [167] lists a total of 12 data sets, 10 of which are global. Perhaps the most widely known are NCEP/NCAR [165] (National Centers for Environmental Prediction / National Center for Atmospheric Research), ECMWF [166] (European Centre for Medium-Range Weather Forecasts) and MERRA [168] (Modern Era Retrospective Analysis for Research). Note that the data can differ considerably in both spatial and temporal resolution. For example, NCEP/NCAR has a geographical resolution (lat/long) of 2.5º x 2.5º, (approximately 280 x 170 km for UK latitudes) and temporal resolution of six hours. MERRA on the other hand has a geographical resolution (lat/long) of 1/2º x 2/3º, (approximately 56 x 45 km for UK latitudes) and temporal resolution of one hour. For wind resource assessments, high spatial and temporal resolution are likely to be particularly important.

At present, despite their increased use, few peer reviewed studies have been published regarding the suitability of such data sets for long-term wind resource assessment using MCP. However, there have been several recent conference papers as well as industry reports that have considered this application. One of the more detailed of these studies was carried out by Brower [164] using NCEP/NCAR data. Brower points out that a key reason to be concerned about the use of reanalysis data in MCP is that the type, quality and quantity of the observational data used to drive the models has changed with time and this is likely to produce false trends when extrapolating over several decades. Brower used three approaches to investigate the suitability of NCEP/NCAR in long-term wind resource assessment. Firstly, the annual mean wind speeds were compared with rawinsonde data (weather balloons) and carefully chosen surface observations. It was found that in some cases, spurious trends existed in the reanalysis data that were not reproduced in the observations. Secondly, MCP approaches were applied at 10 sites using both NCEP/NCAR and nearby meteorological masts as long-term reference data. At 8 out of the 10 sites, the mast data resulted in lower errors in the long-term predictions compared to the NCEP/NCAR data. Finally, the internal consistency of the NCEP/NCAR data was investigated using samples of the reanalysis data to predict the long-term average of the same reanalysis data set. At a number of sites, it was found that for samples greater than around 10 years, the error in the long-term prediction increased due to spurious trends and shifts in the

data. For small-scale installations, correlations over a single decade may be sufficient, and hence, spurious trends may be less of an issue as the quality of assimilated data improves.

Jimenez *et al.* [169] compared NCEP/NCAR and MERRA reanalysis data as well as data from three commercial mesoscale models (Vortex, WindTrends and 3TIER) for use in MCP at six locations. Mesoscale models are derived data sets which attempt to increase the spatial resolution of the data by taking account of the local surface properties. The study was relatively simplistic, comparing the square of the linear correlation coefficient between the reference/target sites for each of the reference data sources. In addition, mesoscale data were only available for a limited number of the six sites tested. The results suggested MERRA performed better than the other data sources although more detailed analysis and a larger number of sites are needed to confirm these findings. Liléo and Petrick [170] also compared NCEP/NCAR with MERRA as well as a more recent NCEP release NCEP/CFSR, which has improved spatial and temporal resolution, using observations at 24 meteorological stations in Sweden. Their results also indicated that MERRA performed better in MCP analysis and was less prone to the spurious trends observed in the NCEP data sets. Similar studies have also been carried out by Pinto *et al.* [171] at 20 sites in Portugal.

Taylor *et al.* [172] suggested that many of the issues identified by Brower [164] could be reduced by the use of a mesoscale model driven by a consistent set of observations. Taylor investigated the performance of commercial mesoscale model data *WindTrends*, a reference data set covering North America from 1997-2009. The data set has improved spatial and temporal resolution compared to NCEP/NCAR (20 km and 1 hour) and is driven only by rawinsonde data in an attempt to remove spurious trends due to changes in the assimilated observations. Using a large number of meteorological stations in the USA, the mesoscale data were shown to better reproduce the annual trends in mean wind speed compared to NCEP/NCAR. However, in around 80% of cases, nearby meteorological stations were shown to have stronger linear correlations than the nearest *WindTrends* grid point, implying that although the data set may outperform traditional reanalysis products such as NCEP/NCAR, it may not offer improvements over nearby reference observations.

These studies highlight both the increasing importance of reanalysis data sets as sources of reference data as well as highlighting their limitations. The most recent work indicates that in order to overcome some of these limitations, there is a move

towards the use of high resolution mesoscale models, driven by reliable data inputs. In this sense, the use of operational forecast data which fulfil these criteria may be considered as a natural choice for long-term wind resource assessment.

8.2.2 The Met Office Unified Model

The Met Office Unified Model (UM) is a world leading operational forecast model operated at resolutions of 25 km globally and 4 km (currently transitioning to 1.5 km) within the UK [173]. The UM is an example of a terrain-following, mesoscale model capable of producing, local, site-specific forecasts. This is achieved through progressively higher resolution (12 km, 4 km and 1.5 km) models whose boundary conditions are provided by the global model. Hence, although primarily designed to produce weather forecasts, there is a clear opportunity for using this high resolution data in the prediction of the long-term wind energy resource.

In simple terms, the starting point for numerical weather prediction models such as the UM is the background model field which includes assimilated observations and the previous forecast state. The background field includes estimates of atmospheric variables such as temperature, pressure, humidity and wind speed as well as physical parameterisations of the surface orography and roughness averaged over a grid [174]. The UM is run in small time steps to solve the equations of motion for a rotating fluid, known as the dynamical core, as well as equations governing a whole range of atmospheric processes [173]. The output includes predictions of atmospheric variables used in weather forecasting in addition to the wind speed and direction required for wind resource assessment.

8.2.2.1 The added value of the Unified Model

Given the studies outlined in Section 8.2.1, it might be expected that output from the UM will provide a more accurate representation of the long-term wind resource compared to reanalysis data. This was investigated recently by Wilson and Standen [175] who compared the observed and predicted mean wind speeds at 80 sites using the UM, reanalysis data sets and the Virtual Met Mast (VMM), a commercial product developed by the Met Office as described in the following section. Of the reanalysis data, the higher resolution data sets of MERRA and NCEP/CFSR were found to yield the smallest bias errors while the UM and VMM predictions performed best with the smallest bias and standard deviation of bias. The study concluded that the added value of the UM mesoscale model, as well as the VMM which includes local downscaling, resulted from improved resolution and more accurate representation of the terrain.

Note that the study did not investigate the added value of using UM and VMM output as reference data in an MCP context. However, the results are promising in terms of the ability of mesoscale model data to represent the wind climate more accurately than reanalysis data.

8.2.2.2 The Virtual Met Mast

Given the potential for using the UM output in wind resource prediction, the Met Office are developing a commercial product termed the Virtual Met Mast (VMM) [19]. The VMM aims to provide hourly time series of wind speed and direction for any UK site. The operational details of the VMM are not in the public domain, however the principle is to apply a correction to the UM output, which forecasts the large-scale flow, in order to correct for the local terrain. Note that while the VMM is a relatively new product, previous work has also considered applying corrections to numerical weather prediction models using localised downscaling [176].

A feature of many numerical weather prediction models, including the UM, is that unresolved orography (orography with a length scale below the resolution of the model) is accounted for by parameterisation through an enhanced or effective orographic roughness. This correctly represents the drag (and hence the wind flows) on a synoptic scale but close to the surface (at tens of metres) the artificially enhanced roughness leads to an underestimate of the wind speed. Similarly, speed-up close to the surface over unresolved hills is not accounted for and heights above ground level may not be accurately represented [91]. The main purpose of the VMM is to correct the raw UM output for these effects and hence make realistic predictions of wind speeds close to the surface. This is achieved in two stages. Firstly, the wind speed close to the surface is corrected to remove the effect of the enhanced orographic roughness through the application of a logarithmic wind profile (Equation 2.19) from some reference height down to the surface. The reference height represents the height to which the orographic roughness parameterisation affects the local wind profile, and hence, this varies with terrain, with the highest reference heights over hilly or mountainous terrain [91]. Secondly, the local effect of the previously unresolved orography is accounted for using a linear flow model [89] to represent the orographic effects described in Section 2.4.7. The final VMM wind speed predictions are thus more representative of the local wind compared to the large-scale flow predictions of the UM.

The VMM is optimised for use at heights of 50 m or greater in accordance with the hub heights of medium to large-scale wind turbines. Verification studies have indicated that the VMM consistently underestimates wind speeds at heights below

30 m [20]. Unfortunately, this is the very height regime that is of most interest for small to medium-scale wind turbines (Section 2.1.1). In addition, while the VMM predictions at heights above 30 m have a very low average bias, the study by Wilson and Standen [175] reported an average absolute error in the mean wind speed of over 1.4 ms^{-1} across 80 test sites and this will have significant implications for wind energy estimates. The most recent implementation of the VMM includes the option for the assimilation of short-term onsite measurements to provide a correction to the modelled wind speeds [177, 178]. Note that this is conceptually very similar to the use of short-term onsite measurements in an MCP approach as described in Chapter 5 and is likely to significantly improve the reliability of the wind resource estimates.

Overall, the VMM is a promising wind resource assessment tool particularly in a scoping context for large to medium scale wind projects. Ongoing development is also likely to improve its accuracy and applicability. However, at present, there is a case for supplementing the VMM approach with alternative methods that are more suited to the small-scale wind industry and capable of greater accuracy.

8.2.2.3 The Unified Model as an MCP reference data source

An alternative method for utilizing the output wind data from the UM in wind resource assessment is as a long-term reference source for MCP. Note that in this case the requirements placed on the UM output are somewhat less strict. In order to be a suitable reference source, we only require that the data represent the same climatology as the target site. The data do not need to represent the target site in an absolute sense since consistent biases will be corrected by the MCP process. Of greater importance is that the reference and target site data feature the same long-term trends or perturbations from the long-term averages.

Note that this is the same requirement that is placed on nearby meteorological stations when choosing reference sites. However, in the case of meteorological stations, a compromise is required between data availability, proximity to the target site and terrain similarity. For example, the reference sites used in the MCP analysis detailed in Chapters 5 - 7 are located at distances of 12 – 133 km from the target sites. An advantage of using UM output as a reference source is that the data are available in a relatively high resolution grid and hence data can be interpolated from the nearest forecast nodes. For the 4 km UK forecast data, the nearest node will typically be much closer than the nearest available surface measurements and hence the data could potentially be more representative of the long-term resource at the target sites. Even for the global forecast model, the resolution of 25 km is

significantly higher than reanalysis data sets opening up the possibility of a high quality global reference data source for MCP.

Such an approach is particularly attractive in the case of small-scale wind energy where such a reference data source would allow wind resource assessments to be performed using only onsite data collected over a period of months.

8.3 Methodology

8.3.1 Measure-correlate-predict

The results detailed in Chapters 5 - 7, indicate that the LR2 linear MCP approach (including a Gaussian scatter term) consistently performs well, particularly with respect to predicting the important parameters of $\bar{u}$ and $\bar{p}_d$. Hence LR2 was chosen as the core MCP technique for implementation using forecast data. However, in order to allow for the possibility that the performance of different MCP approaches may be affected by the choice of reference data, error statistics were also calculated for the linear VR method and the best performing bivariate Weibull approach, BW2. As in previous chapters, the sliding window technique described in Section 5.2.1 was also used to calculate reliable error metrics over multiple training and test periods.

In order to put the results in context, the prediction errors were compared with baseline values. The ‘baseline’ referred to throughout the remainder of this chapter refers to the LR2 approach applied using the meteorological observations at the reference sites (Rf) as described in Chapter 5. Hence a direct comparison is possible between the errors obtained using either nearby meteorological stations or forecast data as a long-term reference data source.

8.3.2 Boundary layer scaling

While the current chapter is predominantly concerned with the use of forecast data in an MCP approach, it is also informative to investigate its utility when used in the boundary layer scaling (BS) approach detailed in Chapter 4. In contrast to the NCIC database of mean wind speeds used previously, forecast data includes a complete time series of wind speed and direction. Hence, it is possible to apply the boundary layer scaling approach to produce a complete time series of predicted wind speeds for multiple angular sectors rather than a single-valued mean. This has the potential to improve the wind power estimates through a more accurate description of the wind speed distribution.

To investigate this application, the boundary layer scaling approach detailed in Chapter 4 was implemented using the forecast data in place of the NCIC database. Rather than applying the scaling to a single mean wind speed and using a fixed Weibull shape factor, the scaling was applied to the hourly time series of forecast wind speeds at a height of 200 m. The scaling was used to obtain a new time series of predicted wind speeds, from which the wind resource statistics were calculated directly.

The fixed forecast height of 200 m was used for consistency with the approach applied in Chapter 4 where the NCIC wind speeds were first scaled to a height of 200 m based on the assumption of a fixed height for the internal boundary layer. Additionally, the enhancements detailed in Chapter 4, namely the allocation of an increased regional fetch and the calculation of angularly dependent aerodynamic parameters, were included. As before, the regional aerodynamic parameters were defined for each of four 90º angular sectors. Forecast wind speeds were binned according to these sectors before applying the appropriate scaling for each sector.

8.3.3 Meteorological data

To facilitate comparisons with the results in Chapters 5 - 7, in the present chapter the same set of 22 reference/target site pairs were used (Table 5.1), covering an 11 year period from August 2001 to July 2012. Note that in some cases the same reference site was used for more than one target site hence the total number of unique sites, reference plus target, is 37. In addition to this observed wind data from the Met Office anemometer network, the UK operational forecast model, operated at 4 km resolution, was used as a further source of reference wind data. To avoid confusion with the two sources of reference data, the observed wind data at the reference sites (as detailed in Table 5.1) are referred to as Rf, while the forecast data used as an alternative reference source are referred to as UK4.

The UK4 data covered the same 11 year period outlined above and were sourced directly from the Met Office as part of a collaborative project. The UK4 data consisted of a time series of hourly wind speed and direction predictions at eight heights, (10, 20, 35, 50, 100, 150, 200 and 500 m above ground level) at each of the target sites. The data were interpolated from the forecast nodes to the actual site locations as part of the pre-processing carried out by the Met Office. Since UK4 has only been operational since 2007, the model was hindcast for the period 2001 – 2006 in order to provide a seamless data set covering the full 11 years.

For the hindcast period, boundary conditions were initialised using ECMWF ERA-interim reanalyses, while from 2007 onward, the model was initialised using the

operational forecast model which is run several times per day with observations assimilated. The effect of using different data sources for initialisation pre- and post-2007 is considered in more detail in the following sections.

8.3.3.1 Independence of forecast and observed wind data

As detailed previously, the current study is based on comparing the predicted wind resource with that observed at a range of target sites which form part of the Met Office anemometer network. Data from this network is also assimilated for use in initialising the background field of the forecast models, along with data from a large number of other atmospheric observations. Hence, an important question is whether the forecast data can be considered largely independent of the observations at the target sites. Here the term 'independence' is used to mean that the forecast data are not significantly influenced by the target site observations since any such influence will tend to cause the success of the MCP approaches to be overstated. Note that this issue is only relevant to forecast data post-2007 since the ERA-interim used for initialisation of the pre-2007 hindcast data did not include observations from this network [179]. There are several factors that must be clarified in addressing this issue.

Firstly, the UK4 assimilates observations of a range of atmospheric variables from a large number of sources. These observations are combined in the background model field that is optimised to be meteorologically consistent. Hence, while surface wind speed observations may influence the background field, the forecast data are not forced to fit these observations. Secondly, observations are only used to initialise the forecast at the start of a run, as the forecast evolves it knows nothing of the actual time evolving wind climate at the locations used for initialisation. Thirdly, the UK4 data used in the current study are extracted starting at T+2 hours, where T+0 is the time at which the forecast is initialised. Hence there is a gap of two hours between any assimilated observations and the first forecast point. These factors imply that the forecast wind speeds used in this study are not likely to be strongly influenced by the wind speed observations at any given hour. In addition, several studies [179-181] have shown that assimilation of the 10 m surface wind speeds appears to have very little influence on the forecast output and that other variables related to temperature and humidity are of greater importance.

However, due to the importance of this issue in objectively assessing the performance of forecast data as a reference data source, further tests were also performed as part of the current study. Current expert opinion among the Met Office team responsible for running UK4 is that any influence of the 10 m surface wind

speed observations will have completely dissipated within the first six hours (i.e. by T+5 hours). Hence, MCP algorithms were implemented using the full data set, and after excluding the target site and reference (UK4) data between T+2 and T+5 hours inclusive (note T+0 to T+1 hours is not included in any case) in order to test for differences in the calculated error metrics. The outcome of these tests is presented in Section 8.4.3.

8.3.3.2 Wind speed resolution and time alignment

The UK4 data were provided with a wind speed resolution 0.00001 ms^{-1} and angular resolution of 0.1^0. This represents the precision of the data rather than the accuracy. The meteorological observations were obtained with a resolution of 0.514 ms^{-1} (1 knot) and 10º. For consistency, the UK4 data were rounded to the nearest knot while the angular resolution was left unchanged since a minimum angular bin size of 30º was used in the MCP implementation. The impact of this procedure was tested through implementation of the MCP algorithms with and without rounding and found to have no significant impact on the error metrics.

In addition, the meteorological observations represent hourly values, averaged over the whole preceding hour. Due to Met Office convention, the actual values usually represent the average between HH-70 minutes and HH-10 minutes, where HH is the hourly time stamp. In contrast, the UK4 data are instantaneous values on the hour at HH. Note that according to Taylor's hypothesis [182], the actual averaging time which these values represent depends on the wind speed and length scale or resolution of the model. For example, for wind speeds of 3 – 10 ms^{-1} and a model resolution of 4 km, these values are representative of the preceding 22 – 7 minutes. In any case, due to this convention, the UK4 values are likely be at least partly ahead of the observational data with the same time stamp. In order to correct for this, the UK4 data were shifted by one hour (i.e. the time stamp was set back by one hour) to provide better alignment with the observational data. Again, MCP tests were carried out to investigate the effect of changing the time alignment and this process was found to have very little effect on the overall error metrics.

8.4 Results and Discussion

8.4.1 Representativeness of the reference data

To be suitable for making long-term target site predictions, the meteorological trends at the reference site must be consistent with those at the target site, or equivalently, the reference data must represent a similar wind climate. For example,

if the target site is located in a coastal area and hence subject to mesoscale effects such as sea breezes, the reference data should reflect these same meteorological patterns in order to effectively predict the long-term wind climate.

A useful measure for comparing the meteorological trends at two sites is the wind index (WI). A WI is the mean wind speed calculated over some short averaging period compared to the mean calculated over a longer averaging period. In effect, the WI describes the short-term deviations from a long-term mean. Here the monthly WI is used to compare the monthly mean wind speed to the long-term mean calculated from the entire 11 year data record. Hence, the WI is defined as:

$$WI_m = \frac{\bar{u}_m}{\bar{u}_{11yr}}$$

Equation 8.1

where WI_m and $\bar{u}_m$ are the wind index and mean wind speed for month m respectively and $\bar{u}_{11yr}$ represents the mean wind speed over 11 years.

For two sites with similar wind climatologies, at least on a specified timescale, it would be expected that the WI would follow similar trends. Figure 8.1 compares the monthly WI for the target site observations and the UK4 reference data (height 50 m) at two target sites C5 and R1 between 2001 and 2012. The C5 site is an example of very good agreement between the monthly wind climates represented by the target and reference data. Site C5 is located on the coat of west Wales in relatively flat and open terrain, the climate of which is well predicted by the UK4 data. The R1 site is an example of relatively poor agreement between the monthly wind climates. While the UK4 data approximately follow the long-term trends for this target site, there are relatively large differences between the forecast and observed WI values from month to month. Site R1 is located in a valley within mountainous terrain in northern Scotland. Hence the site is likely subject to localised wind climates that are not well resolved by the UK4 model.

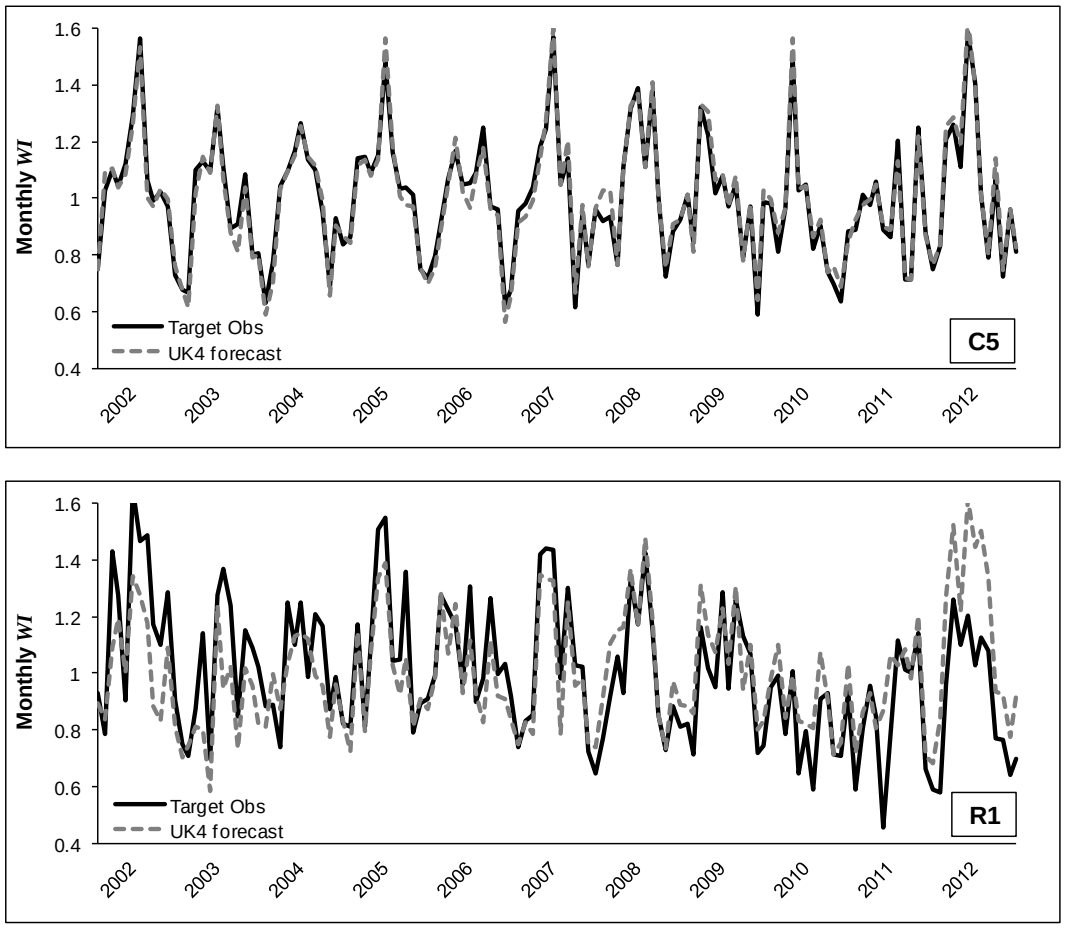

Figure 8.1: Monthly wind indices for the target site observations and UK4 data at sites C5 (top) and R1 (bottom).

As a first step to assessing the suitability of the UK4 forecast data for long-term correlation, WI plots equivalent to those shown in Figure 8.1 were constructed for all target sites. The data were visually checked to ensure no spurious trends or discontinuities were present, with a particular emphasis on the period between the end of 2006 and the start of 2007 for the reasons described in Section 8.3.3. No significant deviations where observed for this period.

The examples presented in Figure 8.1 are relatively extreme cases. For the majority of sites the trends were somewhere between the excellent agreement at site C5 and the poor agreement at site R1. In addition, poor agreement between the forecast and target site data does not necessarily mean that surface wind observations at a nearby reference site would perform better, particularly if the wind climate is highly localised.

Figure 8.2 shows WI values for the target site observations versus the UK4 data for the sites C5 and R1 over the entire 11 year data record. For perfect agreement between the monthly wind climates a linear relationship with zero scatter would be expected. This is close to the observed relationship shown in Figure 8.2 for site C5

while for site R1 there is significantly more scatter and a weaker linear relationship representing the poorer agreement between the two wind climates. Similar plots can also be made using the WI values for the target site observations versus Rf observations to compare the degree of similarity between the target site and Rf data.

In order to quantify the degree of agreement between the WI for two data sets, it is proposed to simply use the linear correlation coefficient (r_{WI}) based on a linear fit to the WI values [98]. For C5 in Figure 8.2, r_{WI} = 0.98 while for site R1, r_{WI} = 0.76 clearly demonstrating the difference in agreement between the two wind climates.

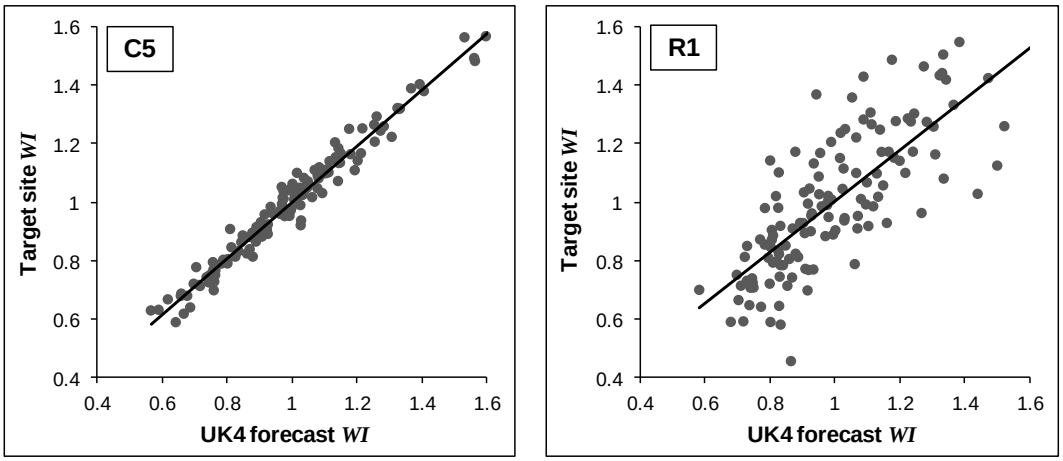

Figure 8.2 Monthly wind indices for the target site observations verses UK4 forecast data at sites C5 (left) and R1 (right). The solid line shows a linear fit to the data.

Note that it is common in the wind industry to use the linear correlation between concurrent wind speed observations at the reference and target sites as a means of comparing reference data sources, either by means of the linear correlation coefficient (r_u) or the coefficient of determination. However, the metric proposed here, r_{WI} based on a linear fit to the WI values, more directly addresses the issue of the representativeness of the long-term trends in the reference data, in line with the issues highlighted by Brower [164]. In addition, r_{WI} may be considered more general in that it imposes the restriction of linearity only on the long-term trends at the two sites rather than the instantaneous wind speeds. A limitation of the r_{WI} metric is that it does not provide information regarding the linear correlation on timescales of less than one month and thus decoupling of the reference/target sites, due to diurnal phenomena for example, will not be identified unless r_u is also considered. Also, while r_{WI} is a useful metric in a research context, in a real-world site assessment the long-term mean wind speed at the target site would not be available for calculation of the target site wind index and hence its application would

be limited by the length of the short-term data. Table 8.1 compares r_{WI} (based on the monthly wind indices) and r_u (based on concurrent hourly average wind speeds) for the UK4 forecast data and the Rf observations

For r_{WI} there is a clear preference for the UK4 data with 15 out of 22 sites achieving higher r_{WI} values compared to the Rf observations and 2 sites showing no preference. This indicates that on average the UK4 data better represents the long-term climate for these sites, most likely because using forecast data at the target site locations removes the effect of spatial separation that is present when using the Rf observations. For the wind speed r_u values, the effect is weaker with exactly half the sites having higher r_u values when using the UK4 data, although the UK4 data achieves a higher overall average r_u.

The two metrics r_{WI} and r_u are most consistent at coastal sites with both metrics showing a preference for UK4 reference data at 5 out of 7 sites. This indicates that for these sites the UK4 data exhibits a stronger linear correlation to the target sites on both monthly and hourly timescales. This is likely because the coastal reference sites (Rf), all of which are located at least 30 km from the target sites, are unable to fully represent the localised target site climates which may exhibit complex seasonal and diurnal variability as described in Section 2.4.5. In contrast, the relatively high resolution UK4 data, extracted at the location of the target sites, appears better able to represent these processes at the 4 km scale. The largest preference for UK4 data, as judged by the r_{WI} and r_u metrics, was observed at reference target site pair Rf7/C1. This site pair are located in the north of Scotland and separated by a distance of 96 km. In addition, while Rf7 is located on the north coast, C1 is located on the east coast. This reference/target site pairing, which was forced due to the lack of nearby reference sites, is clearly non-ideal both because of the large separation and the different coastal orientation at the two sites, which will affect the fetch. Hence, it is not surprising that in this case, the local UK4 forecast provides a more suitable source of reference data. Similar observations were made at a number of coastal sites, C2 for example is located 133 km from the closest Rf site, while C3 is forced to use an Rf site located over 15 km inland. For both sites, the r_{WI} and r_u metrics indicate a strong preference for using UK4 reference data in place of the closest Rf sites. Since scarcity of long-term local reference data may be common in real-world wind resource assessments, UK4 data may prove to be a particularly valuable alternative in such cases.

For the non-coastal sites the results are mixed with the r_{WI} metric averaged across all remaining sites indicating a small preference for UK4 over Rf, while the average

r_u metric indicates no overall preference. This is a reflection of the lower complexity of many of the non-coastal sites, at least on spatial scales resolved by UK4, as well as the greater availability of suitable, local reference sites. For example, for the sites Rf5/SU4, both are located in gentle terrain and the pair are separated by a distance of just 17 km. For this pairing, the r_{WI} and r_u metrics both indicate a preference for Rf reference data over UK4 since the Rf and target sites experience similar wind climates. For the current study, the Rf wind data are collected at the same (or very similar) height to the target site data (10 m above ground level). This is an advantage in the case of the Rf5/SU4 site pair where the sites located in close proximity to each other, with similar fetch and in simple terrain. However, it can be a disadvantage related to Rf data if conditions at one site are disproportionately affected by the local surface and hence highly localised. The results presented here regarding the relative preference for Rf or UK4 data should not be considered absolute since they will depend on the choice of Rf site. However, they demonstrate the advantages of using UK4 data in cases where nearby reference observations are either not available or are located in areas likely to have different climatology to the target site.

	Wind Index r_{WI}		**Hourly wind speed r_u**	
Site	UK4	Rf Observed	UK4	Rf Observed
U1	0.90	0.81	0.81	0.79
U2	0.88	0.89	0.80	0.87
SU1	0.87	0.78	0.79	0.55
SU2	0.92	0.89	0.81	0.82
SU3	0.91	0.90	0.80	0.85
SU4	0.86	0.94	0.77	0.88
SU5	0.96	0.95	0.82	0.92
SU6	0.75	0.73	0.75	0.73
SU7	0.84	0.89	0.76	0.81
C1	0.92	0.64	0.76	0.51
C2	0.95	0.79	0.78	0.66
C3	0.87	0.71	0.74	0.68
C4	0.89	0.92	0.68	0.70
C5	0.98	0.90	0.85	0.79
C6	0.94	0.82	0.76	0.67
C8	0.87	0.94	0.82	0.88
R1	0.76	0.67	0.68	0.53
R2	0.94	0.94	0.81	0.88
R3	0.92	0.89	0.80	0.79
R4	0.94	0.94	0.79	0.85
R5	0.97	0.94	0.85	0.86
R6	0.96	0.91	0.85	0.73
Average	**0.90**	**0.85**	**0.79**	**0.76**

Table 8.1: r_{WI} values between the monthly target site wind indices for the UK4 forecast (50 m height) and Rf observations. r_u values for concurrent hourly wind speeds are also shown. The highlighted cells represent the reference data source with the highest correlation to the target sites.

As described in Section 8.3.3, the UK4 data before 2007 consisted of a hindcast driven by ECMWF reanalysis data. In order to quantify any effect of this procedure on the ability of the forecast to reproduce the climate variability at the target sites, the r_{WI} and r_u metrics were also calculated separately for the periods 2001-2006 and 2007-2012. Averaged across all sites, the r_{WI} metric was found to be the same pre- and post-2007, while the averaged r_u metric was found to be lower for the period 2001-2006 compared to the period 2007-2012, (0.76 compared to 0.81). This indicates that the hindcast pre-2007 UK4 data may be less successful at predicting the variability at the target sites on hourly timescales compared to the post-2007 data, while the monthly averages are unaffected. This effect was not observed in the Rf observations indicating that it is related to the forecast data and not changes in climatology at the target sites post-2007.

The overall conclusions that can be drawn from this analysis are that the UK4 forecast data are in general capable of representing the long-term variability at the target sites as well as, if not better than, the Rf observations. However, the use of hindcast data before 2007 appears to have some impact on the ability of the UK4 forecast to reproduce the target site variability on hourly timescales, although the long-term trends are not affected.

8.4.2 Choice of reference height for UK4 forecast

As described earlier, the UK4 data were obtained at eight heights between 10 m and 500 m. Hence it is necessary to determine which forecast height provides the best reference for applying the MCP algorithms to the targets sites. Note that although 20 out of the 22 target sites are located at 10 m above ground level, the 10 m forecast data does not necessarily best describe the local wind climate due to the effects of unresolved orography noted in Section 8.2.2. The most appropriate forecast level is one of sufficient height that the effects of the orographic roughness parameterisation do not dominate, while remaining within the boundary layer such that the forecast winds remain coupled to the surface observations.

As a first step to identifying this height, the r_{WI} metric between the reference and target site monthly wind indices was calculated for each forecast height and target site. Figure 8.3 shows the average of these values across all target sites as a function of the forecast height. The r_{WI} values exhibit a clear peak for forecast heights of 35-50 m indicating that UK4 data from these heights are most representative of the target site climate variations. A similar trend was also observed for the r_u metric.

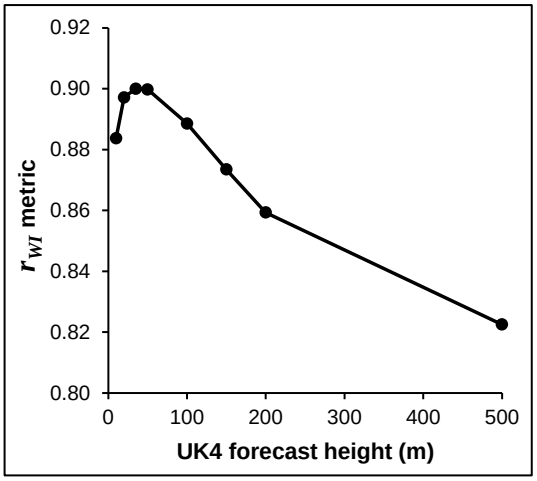

Figure 8.3: r_{WI} metric averaged across all 22 target sites as a function of the UK4 forecast height. The line is a guide to the eye.

To investigate the effect of forecast height further, the LR2 MCP algorithm (linear regression with Gaussian scatter) was applied to predict the 10 year target site wind resource as described in Chapter 5. The UK4 forecast data for each height was used as the reference data source. Figure 8.4 shows the %Error metrics for $\bar{u}$, $\bar{p}_d$, σ and k for each reference height averaged across all 22 target sites. The baseline %Error metrics are also included.

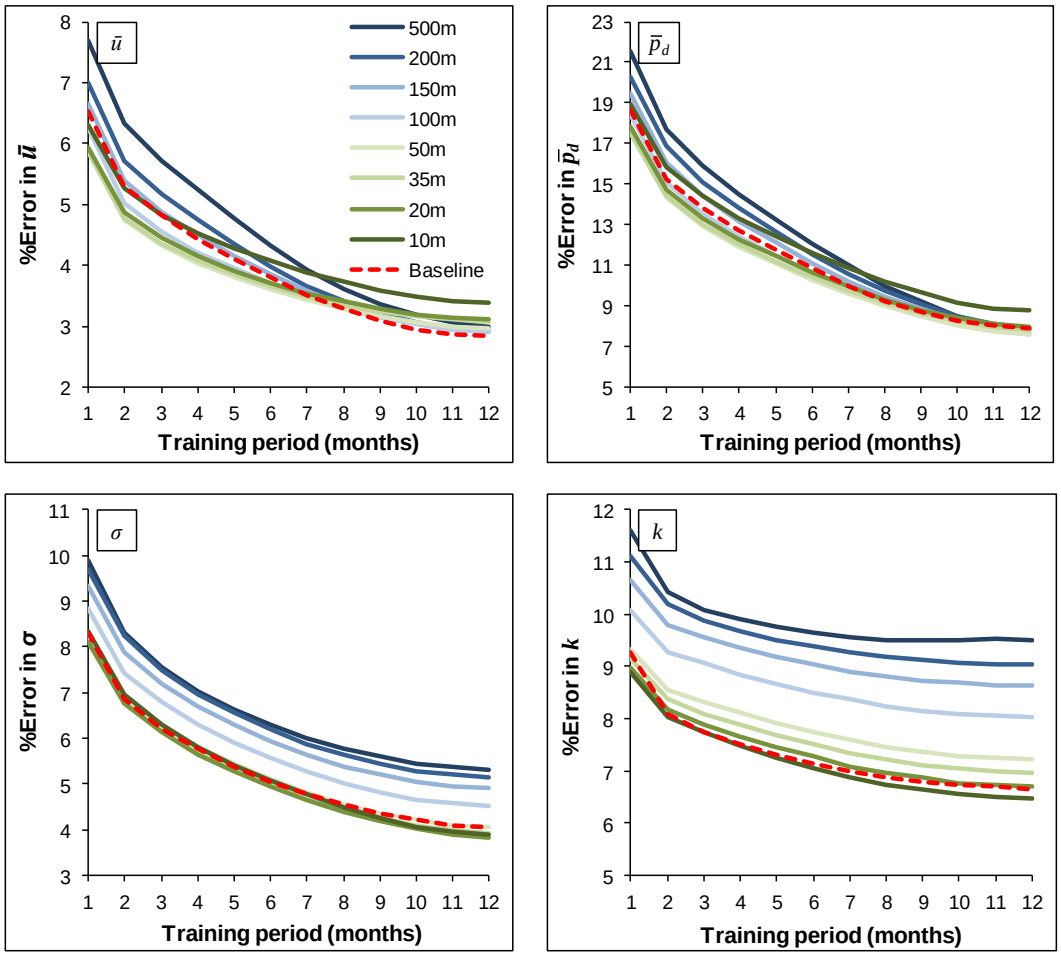

Figure 8.4: %Error metrics for $\bar{u}$, $\bar{p}_d$, σ and k using the LR2 MCP algorithm and UK4 forecast data of different heights as a reference source. Lines show the mean values averaged across 22 site pairs and 120 training periods. The baseline using Rf data as a reference source is also shown.

The %Error in $\bar{u}$ and $\bar{p}_d$ can be seen to generally decrease with decreasing UK4 height before increasing again when using UK4 data below 35 m. Interestingly the effect of measurement height appears to be more pronounced when using shorter training periods, while at 12 months, the errors related to all heights except 10 m tend to converge. For σ and k the trend is more straightforward with a decrease in error with decreasing height, indicating that the width of the wind speed distribution is better predicted when using UK4 data at 10 m. Note that the relationship between the UK4 height and the %Error in σ and k is also likely to be sensitive to the way in which the residual scatter is represented in the MCP algorithm.

These results are summarized in Figure 8.5 which shows the %Error metrics as a function of UK4 forecast height for training periods of 3 and 12 months. The values have been normalised by the baseline, hence a value of less than one indicates that the using the UK4 data as a reference source resulted in smaller errors than using the Rf observations. For $\bar{u}$ and $\bar{p}_d$, the %Error metric follows the inverse of

the trends observed in r_{WI} (Figure 8.3), namely, the highest r_{WI} and lowest %Error are observed for UK4 heights close to 50 m. These trends can also be seen to be stronger when using the shorter training period of three months. In addition, the %Error in $\bar{u}$ and $\bar{p}_d$ is slightly lower using the best UK4 data as a reference source compared to the Rf observations (i.e. < 1), particularly for short training periods. The %Error in σ and k are generally higher when using the UK4 data however, except when using a forecast height of 10 m.

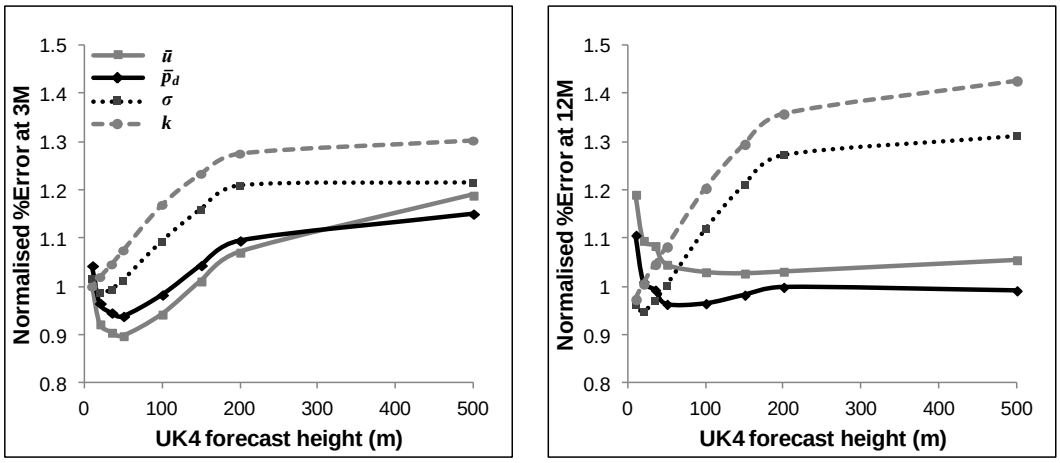

Figure 8.5 %Error metrics normalised by the baseline as a function of the UK4 forecast height for 3 month (left) and 12 month (right) training periods, averaged across 22 site pairs. Lines are a guide to the eye.

In summary, for the current target sites, the most appropriate UK4 height appears to be 50 m with reference data at this height capable of outperforming the Rf observations in terms of the %Error in $\bar{u}$ and $\bar{p}_d$. This reference height is used for the MCP analysis detailed in the following sections.

8.4.3 Independence test

As outlined in Section 8.3.3, while there is good reason to assume that the UK4 forecast data are independent of the target site observations, it is important to confirm that the error metrics calculated using the UK4 data as a reference source are not overly optimistic due to the effect of the data assimilated to the forecast model.

Based on the arguments given in Section 8.3.3, only UK4 data during the four hour period from T+2 to T+5 can be influenced by the assimilations at T+0. In real-time, this four hour period corresponds to 05.00 to 08.00 hours. Note also that only UK4 data from 2007 onwards assimilated surface wind speeds. In order to test for independence, further analysis was undertaken using the five year data period August 2007 – July 2012. From the full five year data set of UK4 forecasts and

target site observations, a second data record was created, termed ‘restricted data’. For the restricted data, all entries between 05.00 and 08.00 hours inclusive were removed. The LR2 MCP algorithm was then applied using the sliding window approach to predict the four year wind resource (one year is reserved for the sliding window) at the target sites using both the restricted and the full five year data sets. Any improvement in the error metrics calculated using the full data compared to the restricted data may be used as an indicator that the forecast data are not independent.

Figure 8.6 compares the %Error metric for $\bar{u}$ and $\bar{p}_d$ using the full and restricted data. It can be seen that both data sets result in very similar error metrics with no indication of any improvement when using the full data compared to the restricted. Very similar results were obtained for the parameters of σ and k as well as the remaining error metrics of MAE and MBE. Hence, it can be concluded that the improved performance of the UK4 data compared to Rf observations is not simply due to a lack of independent observations. This is also in line with previous studies [179-181] that have indicated that the assimilation of 10 m wind data has minimal impact on the model predictions beyond the assimilation time. The remaining analysis described in this chapter is thus based on using the full 11 year data record with no data excluded.

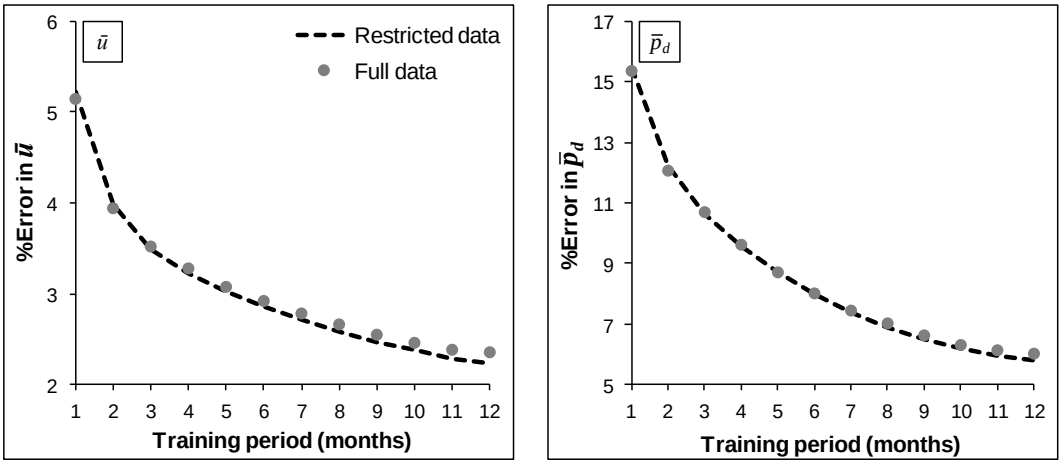

Figure 8.6: %Error metrics for $\bar{u}$ and $\bar{p}_d$, averaged across 22 site pairs using the full and restricted data sets over the period August 2007 – July 2012. The UK4 forecast at 50 m was used as reference data for the LR2 algorithm.

8.4.4 Performance of the MCP algorithms using UK4 reference data

Having established an appropriate height for the UK4 reference data, as well as confirming that the reference data are not unduly influenced by the target site

observations, it is now appropriate to investigate the performance of the MCP algorithms.

Figure 8.7 shows $\bar{u}_{pred}$ at each target site using the LR2 MCP algorithm with a training period of 3 months. The UK4, 50 m forecast was used as the reference data. As previously (Chapter 5, Figure 5.7), the predictions are averaged over 120 test periods, hence seasonal variations are smoothed out. The error bars, which represent +/- 2σ across the 120 training periods, are thus of more relevance to the prediction accuracy than the mean values since they give an estimate of the likely variability in the predictions across different seasons and years.

Figure 8.7 also shows $\bar{u}_{pred}$ based on the direct output of the VMM described in Section 8.2.2. VMM predicted time series for different heights were obtained from the UK Met Office for the same 11 year period as the target site observations and these data were used to calculate $\bar{u}_{pred}$ directly. For the majority of sites the VMM predictions at 10 m were used. However, for sites U1 (20.6 m above ground level) and U2 (22.5 m above ground level), the VMM 20 m predictions were extracted and an additional height correction was applied using a logarithmic wind profile as described in Chapter 4. In addition, no VMM prediction was available for site R1.

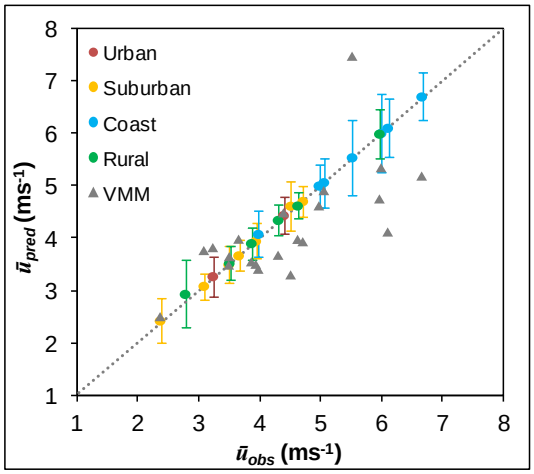

Figure 8.7: Predicted and observed long-term mean wind speeds at 22 target sites using the LR2 MCP approach and the VMM. Error bars represent +/-2σ across the 120 training periods. The dotted line shows a one-to-one relationship.

It is clear from Figure 8.7 that the MCP approach successfully predicts $\bar{u}$ at all sites with very little bias using only 3 months measurements. These results are very similar to those obtained using Rf observations as reference data, (Chapter 5, Figure 5.7). However, the difference in this case is that no long-term reference site observations are required since the UK4 forecast provides the reference data. In contrast, the VMM tends to consistently underestimate the 10 m mean wind speeds

at the majority of sites and in many cases the VMM predictions are outside the LR2 error bars. It is somewhat unfair to directly compare the MCP and VMM predictions at 10 m since firstly the VMM is optimised to work at hub heights of around 50 m and secondly, the VMM implementation shown here does not incorporate any onsite wind data. However, the results do serve as an illustration that despite significant advances such as the VMM, further developments are required to provide accurate wind resource predictions for small-scale turbines located close to the surface.

It is of interest to investigate if the conclusions reached in Chapters 5 - 7 regarding the relative success of different MCP algorithms still hold when using the UK4 data rather than the Rf observations as a reference source. As described previously, the performance of the algorithms is primarily assessed using the error metrics of %Error, MAE and MBE averaged across all target sites using multiple training periods. Figure 8.8 shows the %Error metrics for $\bar{u}$, $\bar{p}_d$, σ and k using the linear MCP algorithms of LR2 and VR as well as the bivariate Weibull approach BW2. The baseline using the LR2 approach with Rf observations is also shown.

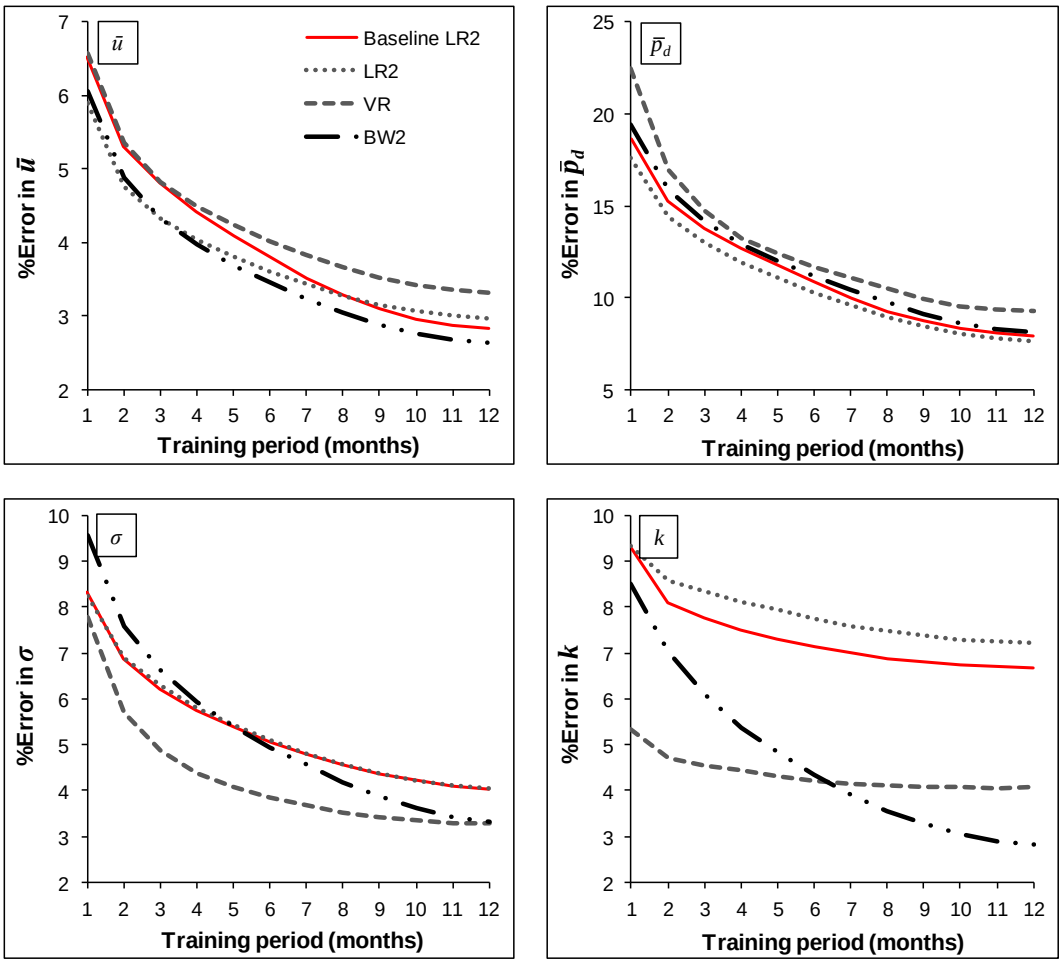

Figure 8.8: %Error in $\bar{u}$, $\bar{p}_d$, σ and k for different MCP algorithms using UK4 reference data at 50 m. Lines show the mean values averaged across 22 site pairs and 120 training periods. The baseline using Rf data as a reference source is also shown.

For $\bar{u}$ and $\bar{p}_d$ the %Error trends are comparable to those shown in Chapter 7, Figure 7.4 where Rf observations were used as a reference source. However, the LR2 and BW2 algorithms perform slightly better using the UK4 data, while the VR algorithm performs slightly worse. For σ and k the trends are also very similar to those shown in Figure 7.4 although the %Error in k tends to fall more steeply for BW2 when using the UK4 data. Very similar trends were also observed for the MAE and MBE metrics as a function of training length. Overall these results indicate that the relative success of the three MCP algorithms, LR2, VR and BW2 is not greatly affected by use of UK4 data in place of Rf observations.

Table 8.2 summarises the metrics of %Error, MAE and MBE for training periods of 3 and 12 months using the UK4 data and all three MCP algorithms. The baseline is also included for comparison.

3 M	Method	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	Baseline	4.8	14	6.2	7.8
	BW2	4.3	14	6.6	6.1
	LR2	4.3	13	6.3	8.3
	VR	4.8	15	4.9	4.6
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	Baseline	0.21	11	0.16	0.14
	BW2	0.18	10	0.12	0.15
	LR2	0.18	9.3	0.15	0.15
	VR	0.20	10	0.12	<0.1
MBE	Baseline	<0.1	-2.8	-0.10	0.13
	BW2	<0.1	-3.4	<0.1	<0.1
	LR2	<0.1	-3.1	-0.11	0.14
	VR	0.11	5.7	<0.1	<0.1

12 M	Method	$\bar{u}$	$\bar{p}_d$	σ	k
%Error	Baseline	2.8	7.9	4.0	6.7
	BW2	2.6	8.1	3.3	2.8
	LR2	3.0	7.6	4.1	7.2
	VR	3.3	9.2	3.3	4.1
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k
MAE	Baseline	0.12	5.8	0.10	0.12
	BW2	0.11	5.5	<0.1	<0.1
	LR2	0.12	5.2	<0.1	0.13
	VR	0.14	6.1	<0.1	<0.1
MBE	Baseline	<0.1	-1.6	<0.1	0.11
	BW2	<0.1	-1.2	<0.1	<0.1
	LR2	<0.1	-2.2	<0.1	0.13
	VR	<0.1	1.5	<0.01	<0.1

Table 8.2: Error metrics for the wind resource parameters of $\bar{u}$, $\bar{p}_d$, σ and k using UK4 reference data and training periods of 3 months (left) and 12 months (right). The baseline using Rf data as a reference source is also shown. Values are averaged across 22 site pairs and 120 training periods

Overall, the errors across all metrics are very similar to those observed in Chapter 7, Table 7.1, demonstrating that the UK4 forecast performs similarly to nearby meteorological observations as a source of reference data. There is also an indication that the UK4 data may lead to slightly reduced errors when using the shorter training period of 3 months.

8.4.5 Seasonal effects

In order to investigate if the seasonal variations in the error metrics are affected by the reference data source, the LR2 MCP algorithm was applied using a training period of three months and both UK4 and Rf observations (the baseline) as reference data. The %Error metrics were averaged for fixed seasons across all years of the data record. Note that this is the same procedure detailed in Chapter 5, Section 5.3.4. The results of this analysis are shown in Figure 8.9.

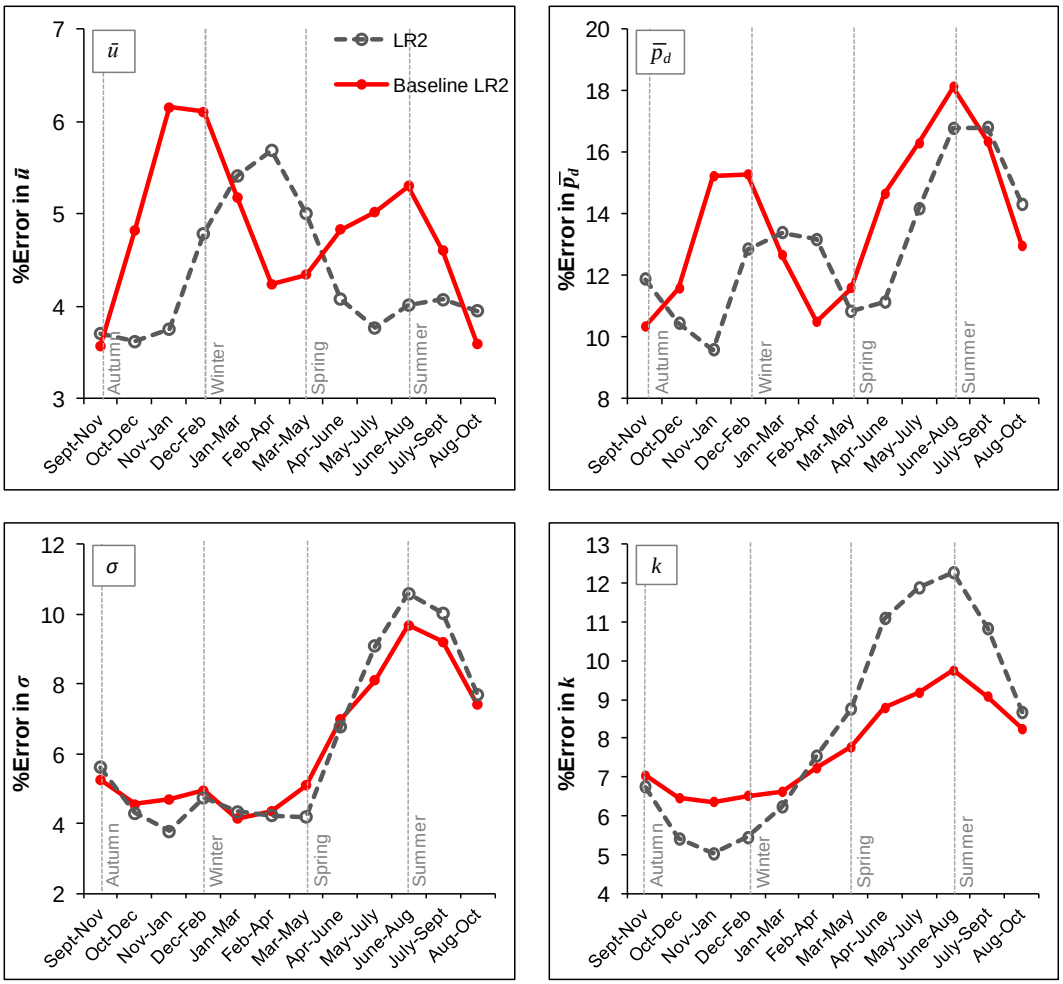

Figure 8.9: Seasonal variation of the %Error in $\bar{u}$, $\bar{p}_d$, σ and k averaged across 22 target sites using the LR2 MCP approach. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training. Values are averaged across 22 site pairs. The lines are a guide to the eye.

Overall, the %Error in $\bar{u}$ and $\bar{p}_d$ is slightly lower for the UK4 data compared to the baseline, in agreement with the results in Table 8.2. For the UK4 data, the %Error in $\bar{u}$ only has a single peak in late winter/early spring and this is shifted with respect to the two peaks in the baseline, this effect is considered in more detail later. The UK4 %Error in $\bar{p}_d$ has two peaks, one which appears to be related to the error in the wind speed and a second which appears to be related to error in σ and k. Note that the influence of σ and k on the peak $\bar{p}_d$ error in summer was also postulated in Chapter 5. The %Error in σ and k follow very similar trends to the baseline, although the magnitude of the error in k is larger using UK4 data during summer. Although not presented here, the seasonal errors for the VR and BW2 methods using UK4 data were also computed and found to exhibit similar trends. The MBE in $\bar{u}$ and $\bar{p}_d$ when using UK4 data were also considered and found to exhibit similar

trends to the baseline in that winter and summer training periods tended to result in overestimates and underestimates respectively.

Based on the above analysis, the main consistent differences between the seasonal errors using UK4 data compared to the baseline are an overall reduction in the seasonally varying error in $\bar{u}$ and $\bar{p}_d$ and a shift in the peak error in $\bar{u}$ from winter to late winter/early spring. To investigate these differences in more detail, two factors were considered. Firstly, the relative impact of coastal sites was investigated since the results presented in Chapter 5, Figure 5.13, indicated that the seasonal error variation may be disproportionately affected by these sites. Secondly, the impact of using hindcast data for training periods pre-2007 was investigated based on the observation of lower r_u values for the UK4 data during this period. The results are shown in Figure 8.10.

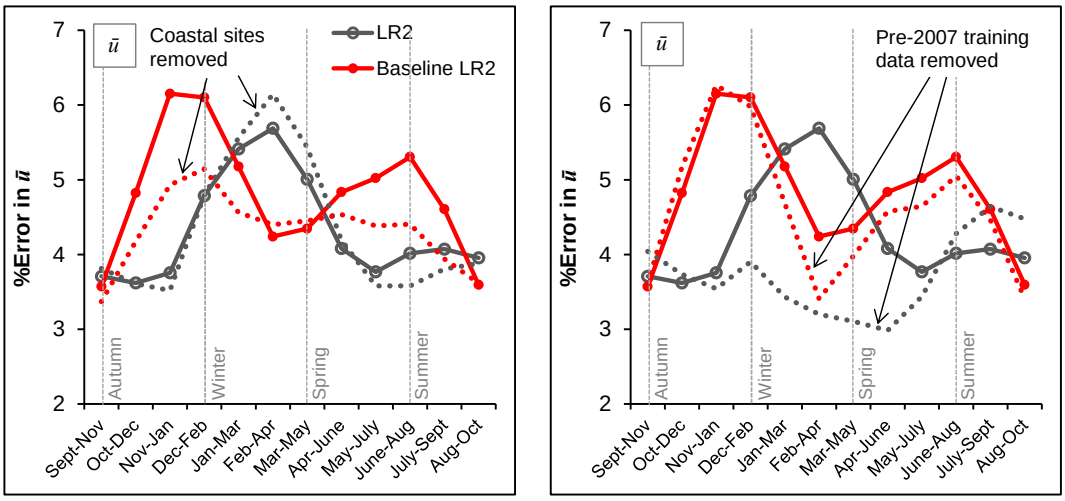

Figure 8.10: Seasonal variation in the %Error in $\bar{u}$ using Rf (baseline) and UK4 reference data. Left: average across all sites (solid lines) and after removal of coastal sites (dotted lines) using training periods across the entire data record. Right: average across all sites after removal of training periods within the pre-2007 hindcast data period (dotted lines). The lines are a guide to the eye.

Figure 8.10 (left) shows that the exclusion of coastal sites results in very little change in the %Error when using UK4 reference data, while for the baseline, which uses Rf observations as a reference source, both error peaks are substantially reduced. This indicates that coastal sites, which were shown in Section 5.3.4 to dominate the seasonally dependent errors, have a less significant impact on the seasonal errors when using UK4 data as a reference source, and hence, UK4 data may be particularly valuable at such sites.

Figure 8.10 (right) shows the effect of removing training periods covering the pre-2007 hindcast data. Note that the hindcast data are still used in the prediction

phase of the MCP implementation in order to predict the wind resource over a full 10 year period but average error statistics are only included for training periods after 2007. For consistency, the Rf training data pre-2007 was also excluded when calculating the baseline predictions. Exclusion of the pre-2007 data has little effect on the trends observed for the baseline. For UK4 data there is a strong effect however, since the peak error in Feb-April disappears and there is an overall reduction in seasonal error variability when excluding the hindcast data. This implies that the use of ERA-interim to initialise the UK4 may adversely affect the wind speed predictions during late winter into early spring. For the hindcast period, the UK4 model was also reinitialised every 48 hours, compared to 24 hours post-2007, and this could be a further source of reduced performance. In a practical sense, these results should not significantly impact on the use of hindcast UK4 data for MCP since the training period will almost always be current, rather than historical. In addition, exclusion of the hindcast training data was found to only have a modest (<1%) impact on the overall error statistics when averaged across all seasons. However, when considering seasonal error variability, the trends observed post-2007 are more likely to be representative of those that will occur using current training periods. Specifically, in contrast to Rf reference data, UK4 data appears to result in no strong preference for autumn, winter or spring training periods.

The effect of excluding coastal sites and the use of hindcast data on the seasonal errors in the distribution parameters of σ and k was also investigated but this analysis revealed no significant effect related to these parameters.

8.4.6 Errors at individual sites

To explore the performance of the UK4 data at individual sites, the average %Error in $\bar{p}_d$ across all seasons using a three month training period was calculated for each site using UK4 and Rf reference data. The results are shown in Figure 8.11.

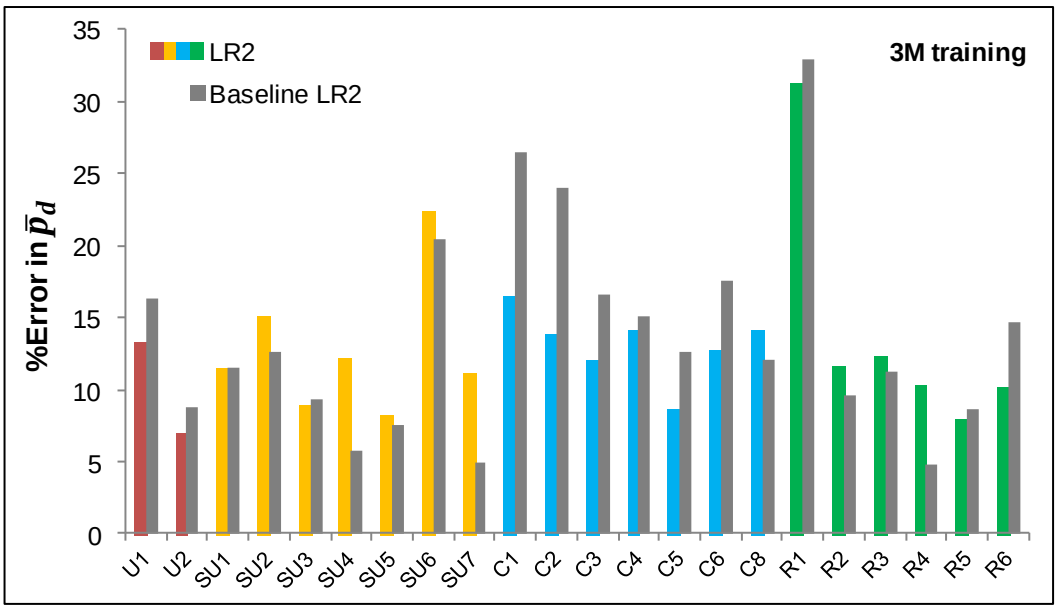

Figure 8.11: %Error in $\bar{p}_d$ for each of the 22 target sites using a training period of 3 months and UK4 and Rf (baseline) reference data. Values are averaged across the 10 year data record.

Comparison with Table 8.1 shows that the r_{WI} and r_u metrics are reasonable predictors for the best reference data source at sites exhibiting relatively large error differences, although they are less successful when the error differences are small. For the urban, sub-urban and rural sites, the results are variable with the Rf data (baseline) resulting in reduced errors at 8 sites, compared to 6 sites for the UK4 reference data. This indicates that while UK4 reference data performs adequately overall, it may not outperform Rf at specific sites. This is likely because UK4 data are not of sufficiently high resolution to represent local effects caused by complex terrain or the built environment and hence cannot add significant value in these terrains. However, for coastal sites there appears to be a clear preference for UK4 reference data with reduced errors at 6 out of 7 sites.

As described previously, coastal sites may be particular sensitive to (i) the separation distance between the reference and target sites (Section 2.4.5) and (ii) seasonally varying weather conditions (Section 5.3.4). The use of UK4 data offers advantages with respect to both of these effects. Firstly, the use of forecast data at the location of the target site avoids the effects of large reference/target site separations. Secondly, as shown in Figure 8.10, forecast data (excluding the hindcast period) are less subject to seasonal variability in the prediction errors compared to Rf data. A possible explanation for the reduced seasonal variability is that the UK4 data better represents the local stability conditions and fetch effects at coastal locations compared to Rf sites which may be located tens of kilometres away. To investigate this further, the diurnal variation in r_u was compared for the

coastal sites using UK4 and Rf reference data. If, as suggested, the UK4 data better represent variable stability conditions, this should result in reduced diurnal variability in r_u when using the forecast data. To allow multiple sites to be compared, the diurnal r_u values were first normalised by the average value for each site shown in Table 8.1. The normalised values, averaged across the 11 year data record, are presented in Figure 8.12.

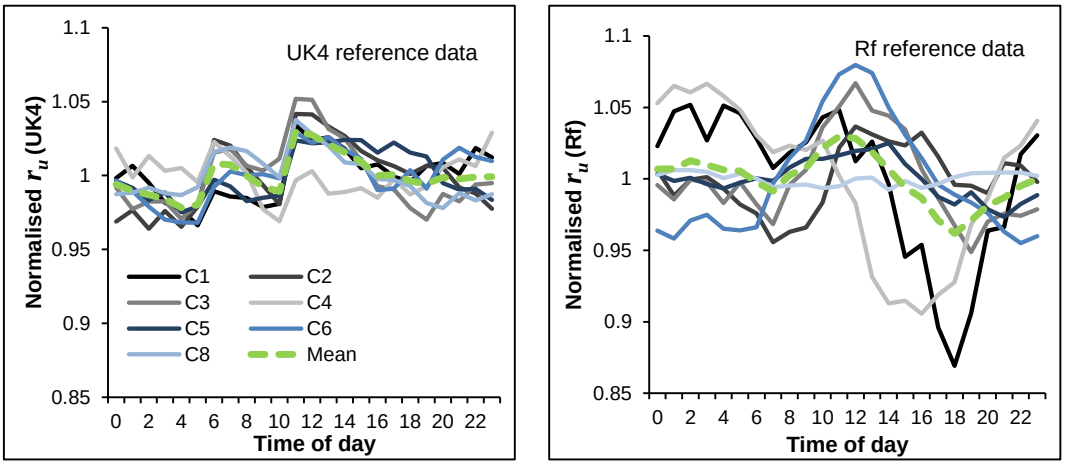

Figure 8.12: Diurnal variability in the normalised linear correlation coefficient r_u between the reference and target site hourly wind speeds using UK4 (left) and Rf (right) reference data at coastal sites. Values are averaged over the 11 year data record.

For both UK4 and Rf, on average the highest correlations occur around midday, presumably due to the higher prevalence of neutral stability conditions around this time. However, it is clear from Figure 8.12 that when using UK4 reference data, the diurnal variation in r_u is markedly reduced for many sites. For example, at site C1, which was noted in Section 8.4.1 for the large reference/target site separation distance and differences in coastal orientation, the significant diurnal fluctuations, presumably due to the non-ideal location of the Rf site, are clearly reduced when using UK4. Similar observations can be made at a many of the coastal sites. This implies that the UK4 data are better able to represent the diurnal changes in atmospheric stability at such sites, and by extension, better able to represent seasonal variability related to these effects. Similar analysis of the remaining sites revealed that the impact of using UK4 data on the diurnal variability in r_u is much weaker at non-coastal sites. This is likely because Rf data are more capable of representing the stability conditions at sites not subject to complex coastal climates or because the effect of stability is less crucial to the reference/target site correlation in non-coastal areas.

8.4.7 Extension to 37 test sites

So far, error statistics have been presented based on the 22 target sites that have been used throughout Chapters 5-7 in assessing the performance of the MCP approaches. However, for the current study, since UK4 data are used in place of Rf observations, there is an opportunity to expand the test sites to include all target and reference sites listed in Chapter 5, Table 5.1. Using this approach, both the target and reference sites are treated as locations where we wish to predict the wind resource and the reference data are obtained solely from the UK4 forecast. The additional reference sites are located mostly in rural or coastal areas. To test the robustness of the conclusions presented thus far, the LR2 algorithm was applied to all 37 sites using the UK4 reference data. The resulting error metrics for both data sets (22 target sites and 37 reference plus target sites) are compared in Table 8.3 for training periods of 3 and 12 months.

	No. Sites	$\bar{u}$	$\bar{p}_d$	σ	k	
%Error	22	4.3	13	6.3	8.3	} 3M
	37	4.1	12	6.1	8.1	
	22	3.0	7.6	4.1	7.2	} 12M
	37	2.8	7.3	4.0	7.0	
		$\bar{u}$ (ms^{-1})	$\bar{p}_d$ (wm^{-2})	σ (ms^{-1})	k	
MAE	22	0.18	9.3	0.15	0.15	
	37	0.18	10	0.15	0.15	
	22	0.12	5.2	<0.1	0.13	
	37	0.12	5.7	0.10	0.13	
MBE	22	<0.1	-3.1	-0.11	0.14	
	37	<0.1	-3.5	-0.11	0.14	
	22	<0.1	-2.2	<0.1	0.13	
	37	<0.1	-2.6	<0.1	0.13	

Table 8.3: Error metrics for the wind resource parameters of $\bar{u}$, $\bar{p}_d$, σ and k for training periods of 3 months (grey shading) and 12 months (no shading) using UK4 reference data. The average metrics are shown for the 22 target sites and the combined 37 target plus reference sites.

It can be seem that the average error metrics remain broadly unchanged even after incorporating the additional 15 sites for both 3 and 12 month training periods. This increases the confidence that the results presented in this chapter show general trends that are broadly representative of UK sites.

8.4.8 Boundary layer scaling using UK4 data

The results considered thus far have all been related to the application of UK4 data in an MCP approach. However, as detailed in Section 8.3.2, the UK4 data at a forecast height of 200 m was also used as a source of reference climatology to implement the boundary layer scaling (BS) approach presented in Chapter 4. Using

this approach it was possible to compare the effect of using different reference climatologies (UK4 and NCIC) on the wind resource predictions at the 22 target sites. Figure 8.13 compares the predicted $\bar{u}$ at the target sites using the BS approach applied to both UK4 and NCIC reference climatology. The predictions obtained using the VMM approach, obtained directly from the Met Office, are also shown for comparison.

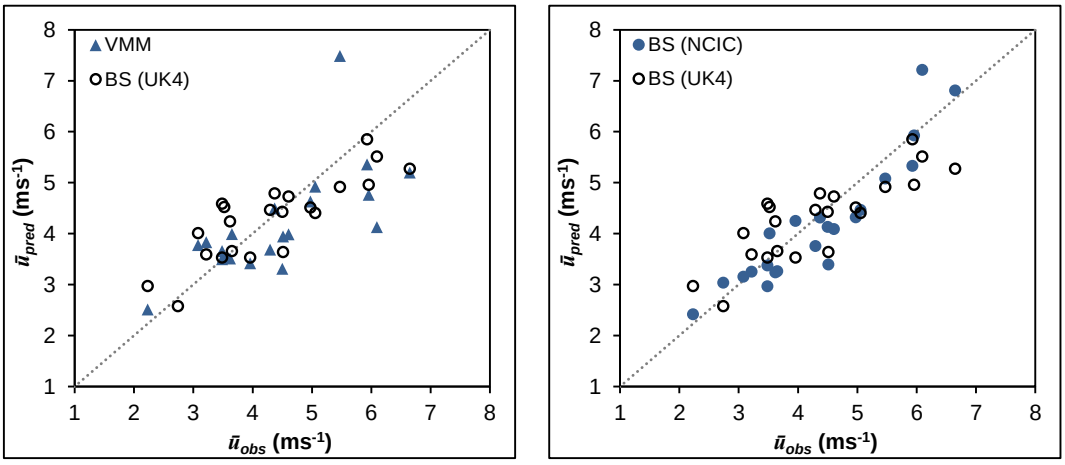

Figure 8.13: Predicted versus observed mean wind speeds using a boundary layer scaling (BS) approach. Left: Comparison between BS (UK4) and VMM predictions, Right: Comparison between BS (UK4) and BS (NCIC). The line represents a one-to-one relationship.

Figure 8.13 indicates that when using UK4 data, the BS approach overestimates low wind speeds and underestimates high wind speeds. Interestingly a similar pattern is seen for the VMM predictions, although it should again be noted that the VMM is not optimised to work at such low hub heights. In addition, the principle of operation of the VMM is to apply corrections for the effective orographic roughness, which otherwise leads to underestimates of the wind speed close to ground level, and corrections for local orography using a linear flow model. In contrast, the principle of operation of the BS approach is to simply downscale wind speeds between the top of the boundary layer and the surface in order to account for regional and local roughness. Figure 8.13 also indicates that on average the NCIC climatology results in more accurate predictions of $\bar{u}$ compared to the UK4 climatology, as indicated by the reduced scatter and closer proximity to the line that represents a one-to-one relationship.

Calculation of the average error statistics revealed that the UK4 climatology resulted in an average %Error across all sites of 13% and 37% in $\bar{u}$ and $\bar{p}_d$ respectively, while for the NCIC climatology the values were 9.5% and 26%. In addition, the downscaled UK4 data failed to result in improved predictions of the Weibull shape factor k, compared to the previous assumption (Chapter 4) of k =

1.89. Hence in its present form it appears that the BS approach is more effective when applied to the NCIC climatology compared to UK4. There are a number of possible reasons for these observations. Firstly, the UK4 data have a spatial resolution of 4 km compared to 1 km for the NCIC data. Hence, the local effects of orography (Section 2.4.7), are likely to be better represented in the NCIC data. However, application of the recently developed Met Office UKV forecast model with a spatial resolution of 1.5 km may go some way to addressing this issue.

Secondly, the NCIC data have a simple interpretation as the mean wind speed at a height of 10 m above open, level terrain, thanks to the way the data set has been derived. In contrast, interpretation of the UK4 data is complicated due to parameterisation of the orographic roughness. The height to which this impacts on the predicted wind speeds will vary with terrain complexity [91], and hence, downscaling from a single reference height of 200 m will introduce errors in the final wind speed predictions. A more sophisticated application should seek to use a variable reference height obtained directly from the forecast model. In addition, it would theoretically be possible to output parameters related to stability from the forecast data which could be used to apply a stability correction to the neutral log law which is used for downscaling in the BS approach.

The factors noted above, as well as the limitations identified in Chapter 4, offer several routes for further investigation before discounting the use of forecast data in a simple boundary layer scaling approach. However, it should be borne in mind that as the complexity of the BS approach is increased, it will become harder to justify its use in place of more sophisticated methods such as the VMM.

8.5 Conclusions

Output from a state-of-the-art UK forecast model (UK4) has been investigated in terms of its suitability for use as long-term reference data for the implementation of MCP approaches to wind resource assessment. In the preparatory stages, precautions were taken to ensure independence of the UK4 and observed data as well as to investigate factors related to time alignment. Next the ability of the UK4 data to represent long-term trends in the target site wind data was investigated through the calculation of a monthly wind index. It was found that the linear correlation coefficient between the monthly wind indices calculated for the target and reference data can be used as a metric to quantify the representativeness of the reference data. Based on this metric, the UK4 data were shown to be capable

of representing the long-term trends in the target site wind speeds, on average, slightly better than observations at nearby meteorological stations.

Further investigations were performed to identify the most appropriate forecast height for use in the MCP prediction algorithms. Based on a range of error metrics, UK4 data at a height of 50 m was shown to be most appropriate. Using this information, MCP algorithms were implemented to calculate the 10 year wind resource at 22 target sites (later extended to 37) using both the UK4 data and nearby meteorological observations as a reference sources. The UK4 data were shown to perform as well as, or slightly better than, the meteorological observations when used as a long-term reference source for a range of training periods. At coastal sites there appears to be a systematic improvement in the predicted $\bar{p}_d$ when using UK4 reference data, and further analysis revealed that this is likely due to improved representation of the local stability conditions and coastal wind flows when using these data. Both linear (LR2, VR) and non-linear (BW2) MCP algorithms were tested and the choice of reference data was found to not substantially affect the relative success of the different approaches. These results are highly promising in that they open up the possibility of assessing the wind resource available to small wind turbines using only short-term onsite measurements and without recourse to long-term reference observations. The approach appears to also offer additional benefits in the prediction of the wind resource at coastal sites.

An investigation of the seasonal variation in the error metrics revealed that while the error in the distribution parameters of σ and k follow similar seasonal patterns using both reference data sources, the pattern of seasonal errors in $\bar{u}$ is affected by the use of UK4 data. For training periods excluding the pre-2007 hindcast, the seasonal variability in the predicted $\bar{u}$ is reduced when using UK4 data with only small differences observed for the training periods of autumn, winter or spring. Additionally, coastal sites appear to have a bigger impact on the seasonal variability in the error in $\bar{u}$ when using nearby meteorological stations compared to UK4 reference data, likely due to better representation of localised meteorological conditions when using the UK4 forecast.

Use of the UK4 time series data as a reference climatology in a boundary layer scaling model failed to produce improved wind resource predictions compared to the use of the NCIC database of mean wind speeds. However, the use of forecast data in such a model introduces several complexities that should be considered in

more detail before reaching a final conclusion regarding the suitability of this approach.

9 Overall Summary and Conclusions

There is currently great potential for harnessing renewable power from carefully sited small-scale wind turbines, as concerned stakeholders strive to decarbonise the electricity supply. Within the UK there are encouraging signs of a growing small-scale wind industry thanks to the favourable wind resource and financial incentives for small-scale generators. However, in order to maximise carbon savings, financial benefit and confidence in the technology, it is crucial that systematic approaches are applied to assessing the wind resource at potential turbine sites. Although detailed site assessment procedures are routinely applied in the large-scale wind industry, such approaches are generally not practical for small-scale turbines due to the high costs and long timescales involved. Hence, low-cost approaches capable of rapidly and accurately characterising the available wind resource are urgently required.

While the large-scale wind industry has benefited from decades of research into wind resource assessment, research related to small-scale installations is still in its infancy. Much of the work completed thus far has related to analytical models using scaling and/or fluid flow approaches to describe flows over rough surfaces or to understand detailed building aerodynamics. These studies have produced promising developments but there has been a need for a systematic evaluation of the accuracy of analytical models in predicting the wind resource at sites located in a range of terrains. In addition, data-driven approaches, particular as applied to very short onsite measurement periods, have received very little attention in the context of small-scale wind energy despite their promising potential.

The work described in this thesis has addressed these issues by considering two major routes to wind resource assessment applicable to small-scale wind energy: (i) an analytical model based on the principles of boundary layer meteorology and (ii) data-driven approaches based on short-term measurements correlated to long-term reference data. The overall aim of this has been to develop techniques capable of accurately predicting the available wind resource without recourse to costly long-term measurement campaigns and to better understand the uncertainties related to both analytical and data-driven approaches.

In Chapter 4, a detailed evaluation of the performance of an analytical, boundary layer scaling approach was carried out using long-term wind data from a large number of UK sites located in a variety of terrains. Based on this evaluation, a

number of improvements to the approach were proposed and tested. However, due to remaining uncertainties in the predicted wind resource, it was suggested that the approach is best applied in a scoping context to identify promising sites for further study. In Chapter 5, the feasibility of obtaining improved predictions by applying data-driven, linear MCP techniques using very short-term onsite measurements was investigated. The results indicated that large improvements in accuracy, compared to a boundary layer scaling approach, could be obtained with just 3 months of onsite measurements. The promising results of this study motivated further exploration of novel MCP approaches including Gaussian process regression in Chapter 6 and a bivariate probability approach in Chapter 7. However, despite the attractive theoretical properties of these approaches, they failed to consistently outperform the best linear MCP method when applied to short measurement periods. Finally, in Chapter 8, output from an operational forecast model was investigated as an alternative source of long-term reference data for both analytical and data-driven approaches. The results indicated that such data can be a valuable source of reference data for MCP but further work is needed to investigate its suitability in a boundary layer scaling approach. These results offer small-scale wind energy developers several low-cost choices for the systematic assessment of the wind energy resource, along with an indication of the likely uncertainties using different approaches and measurement periods.

9.1 Findings

9.1.1 Boundary layer scaling

An existing boundary layer scaling approach, first developed by the UK Met Office, was investigated in Chapter 4 in terms of its ability to predict the spatially averaged mean wind speed and the resultant wind power density. The approach is attractive in that it can be implemented using relatively simple parameterisations of the surface roughness along with a reference climatology, such as a wind atlas, without the requirement of additional onsite measurements or specialist local knowledge. These features allow multiple sites to be assessed rapidly and at low cost. However, when implemented at 38 UK sites in a variety of terrains, the approach was shown to result in relatively large errors in wind speed and power density due to uncertainties in the input parameters.

Several modifications to the methodology were developed including: increasing the size of the regional fetch, applying directionally-dependent regional roughness parameters and optimizing the value of the Weibull shape factor and these were

shown to result in improved predictions. However, even with these improvements, the average errors were unlikely to be small enough to allow investment decisions to be made with confidence. A global sensitivity analysis revealed that the main contributors to the prediction uncertainty were the specified local aerodynamic parameters and the choice of Weibull shape factor, highlighting areas for further study. Despite these uncertainties, the methodology was shown to perform well when applied in a scoping context to judge sites against a viability criterion. Such an approach would allow resources to be concentrated at the most promising sites in order carry out more detailed site assessments. In this context, the methodology was shown to perform noticeably better than the currently legislated NOABL-MCS method, which incorrectly excluded many viable sites.

9.1.2 Linear MCP applied to short measurement periods

Due to the uncertainties inherent in boundary layer scaling approaches, in Chapter 5 attention was turned to data-driven MCP techniques. The contribution of this work was a rigorous assessment of MCP approaches applied to very short measurement periods using multiple training and test periods over a large number of sites. Three linear MCP approaches were tested in terms of their ability to predict the 10 year wind resource using just 3 months of onsite measurements and long-term correlation to a reference site. A sliding window method was used to obtain robust error metrics over 120 training/test periods and 22 target sites. Despite the onsite measurement period being significantly shorter than that typical recommended for MCP, the data-driven approach was shown to result in large improvements in the accuracy of predictions compared to the boundary layer scaling approach. Specifically, linear regression with a Gaussian scatter model (LR2) was shown to result in the most accurate predictions of mean wind speed and power density and the dangers of applying linear regression with no representation of the residual scatter were highlighted. Seasonal variability in the reference/target site relationship was shown to impact both the sign and magnitude of the prediction errors with the most accurate predictions at UK sites obtained when onsite measurements were obtained during autumn or early spring. An average improvement in the error of the predicted wind power density of 8 percentage points was observed when using the best seasons for onsite measurements, with much of this improvement related to coastal sites. These results indicate that even very short measurement campaigns can add significant value to the resource assessment procedure, thus justifying the additional time and expense. In addition, where practically possible, measurements should be obtained during seasons where the prediction errors are likely to be minimised.

9.1.3 Gaussian process regression

Given the encouraging performance of linear MCP approaches applied to very short measurement periods, in Chapter 6, a flexible Gaussian process regression (GPR) framework was developed for the implementation of non-linear MCP. To the best of our knowledge, this was the first application of Gaussian process regression in the context of MCP. Since the GPR approach is not restricted to any specific functional form, it offers much greater flexibility than traditional regression approaches and is able to adapt to the specific form of the reference/target site correlation. However, despite its greater flexibility, GPR did not result in improved performance compared to the best performing linear approach. Further investigation revealed that for the 22 sites considered, the GPR framework reduced to a model that was close to linear thus offering little advantage over linear regression. The GPR framework was further extended to the more challenging case of orthogonal regression, where in the presence of veer, correlations are likely to be non-linear. In this case, GPR was shown to successfully model non-linearities resulting in more accurate predictions of the target site wind vectors compared to linear approaches. Based on the predicted wind vectors, GPR was also shown to offer improved predictions of the long-term distribution of target site wind angles, compared to the assumption of equivalent distributions at the reference and target sites. While such information may be of secondary importance for small-scale turbine installations, this is an area worthy of further study, particularly in the context of turbine siting.

9.1.4 Conditional probability approach

In Chapter 7, an MCP approach was developed based on modelling correlated wind speeds at the reference and target sites using a bivariate Weibull probability distribution. As far as we are aware, the bivariate Weibull probability approach has not previously been applied to MCP using real wind observations. The study was motivated by a desire to model the residual scatter, identified as a significant factor in Chapter 5, using a more rigorous theoretical basis. Specifically, assuming wind speeds at the reference and target sites are described by univariate Weibull distributions, the use of a linear model with Gaussian scatter, as implemented in Chapter 5, can be considered as a simplified representation of an underlying bivariate Weibull distribution. A detailed investigation of the fitting efficiency of the bivariate Weibull distribution was undertaken using real wind observations and samples drawn from idealised distributions modelled on the observed data. The results indicated that while the approach performs well when using idealised data, there are significant challenges related to its application to real wind data. In

particular, it was found that the distribution parameters, required for implementation of the MCP approach, took considerably longer to converge when using real wind observations due to seasonal variability. These issues are particularly problematic when attempting to implement the approach using training periods of a few months. Hence, despite the attractive theoretical properties of the bivariate Weibull approach, it was found that it only became competitive with the best linear approaches when using training periods of around 8 months or more. These results further highlight the excellent performance of linear approaches coupled to a Gaussian scatter model despite their inherent simplifications.

9.1.5 Alternative sources of reference data

In order to reduce the reliance on long-term reference observations, Chapter 8 considered the use of output form a state-of-the art UK forecast model (UK4) in implementing the short-term MCP approaches, as well as a modified BS model applied to a complete time series of wind speeds. The linear correlation coefficient between the monthly reference and target site wind indices was proposed as a metric to assess the degree to which the UK4 data represented the long-term climatology at the target sites. Evaluation of this metric indicated that the UK4 data were capable of representing the target site wind climates as well as, or better than, nearby meteorological masts. The UK4 data were then used as a long-term reference source to implement the MCP approaches developed in previous chapters. Based on the performance of the approaches using forecast data at heights between 10 and 500 m, an optimal forecast height of 50 m was identified and used to calculate average error metrics. The UK4 data were shown to be highly competitive with data from nearby meteorological masts as a reference source for the application of MCP. Specifically, there were indications that the UK4 data may better represent highly localised conditions at coastal sites compared to nearby reference observations. In addition, while seasonal variability was observed in the correlations between the reference and target site data, the variability was reduced when using UK4 data for all seasons except summer. The overall success of the UK4 data applied to short-term MCP opens up the possibility of assessing the wind resource at potential small turbine sites using only wind data collected over a period of months at the target sites. In addition, these findings have the potential to provide large improvements when implementing short-term MCP at coastal sites.

The UK4 data were also investigated as an alternative to the NCIC wind atlas in order to implement the boundary layer scaling (BS) approach developed in Chapter 4. Application of the BS approach to a complete time series of wind speeds, such

as UK4, could potentially offer a better representation of the distribution of wind speeds and the associated wind power. However, using a simple implementation of the BS approach, the UK4 data failed to improve on the predictions obtained using the NCIC data. Further study is required in this area to investigate how the additional complexities of using forecast data may best be included in the model.

9.1.6 A cautionary note

The running theme throughout this work has been the development of indirect methods that are capable of predicting the wind resource without the need for long-term onsite measurements. The motivation for this is that since such measurements are time-consuming and generally costly, they may not be practical except in cases where large investments are to be made. The systematic application of the best available indirect methods, applied with an awareness of their inherent uncertainties and the associated risks, is certainly preferable to discounting a proper site assessment entirely. However, it should be borne in mind that indirect techniques will almost certainly result in larger average uncertainties and errors than long-term onsite measurements and in certain cases they can result in large over- or under-estimates of the wind resource. Hence, indirect methods should not be seen as an alternative to a long-term onsite measurement campaign, if such a campaign is feasible. Rather, indirect methods can provide valuable information where direct methods are simply not practical due to their impact on the total investment cost or where decisions must be taken before a long-term measurement campaign has been completed.

9.2 Limitations and Opportunities for Further Work

Throughout this work, a number of limitations have been identified, some of which offer opportunities for further studies. These are discussed in more detail below.

9.2.1 Boundary layer scaling model

In the boundary layer scaling approach implemented in Chapter 4, many of the limitations are related to the restriction that the methodology should be capable of rapid application at multiple sites without detailed, site-specific information.

An obvious issue with such an approach is that the aerodynamic parameters of the surface must be inferred from simple categorisations of regional land cover and subjective estimates of the local site characteristics. The sensitivity analysis performed in Chapter 4 implied that the wind resource predictions are particularly sensitive to the values of the local aerodynamic parameters. For example, within

the built environment, the local parameters will be subject to uncertainties due to variations in the density, height and heterogeneity of the roughness elements that cannot be captured by a simple urban or sub-urban categorisation. Surface parameters on a regional scale, such as the regional displacement height, are a further source of uncertainty. This parameter is chosen to be the highest of the displacement heights identified from the land cover in the regional grid square, an assumption that could lead to overestimates in regions with highly heterogeneous land cover. In addition, since the flow is assumed to be fully adjusted to the local/regional surface below/above the blending height, no allowance is made for boundary layer growth at the edge of a roughness change. This will affect predictions made close to step changes in roughness that may be encountered at coastal sites or close to the boundary between rural and built environments. Small-scale perturbations to the mean flow caused by local obstructions as well as the effects of local orography on scales smaller than the resolution of the reference climatology may result in further errors.

Many of these limitations can be improved upon in the case of a site-specific implementation of the methodology, where detailed account is taken of the site characteristics. However, this would negate some of the advantages of the approach. An area worthy of further exploration is to what extent detailed local information could be taken into account while maintaining the broad applicability of the overall methodology. The work reported by Millward-Hopkins [87], which utilizes detailed building information on the city-scale, is one example of such an approach applied specifically to the built environment.

Additionally, the sensitivity analysis in Chapter 4 indicated that uncertainty in the Weibull shape factor is a further significant source of error. Improved estimates of this parameter require the use of either a wind atlas that includes estimates of the shape factor on a sufficiently local scale, or alternatively, time series wind speed data that could be potentially obtained from a forecast model.

9.2.2 Data-driven approaches

The limits of the MCP approaches developed in Chapters 5-7 have been explored through application to very short training periods; however, there is scope for further optimisation of the approaches when using limited data. For example, the presence of strong seasonal variability in the error metrics reveals that the reference/target site relationships vary depending on the training season. While this variability may be reduced by using UK4 reference data, ultimately, seasonal terms could be included in the MCP models which take account of the training season. In addition,

protocols could be developed which require a threshold angular sector coverage or variance to be achieved before the data are considered sufficiently representative of the long-term correlation between the reference and target sites.

A limitation of the MCP approaches explored in this work is the use of angular sectors of fixed size and position. A more efficient use of data would be to specify sector width and position based on the characteristics of the reference/target site correlation. For example, some criterion could be applied whereby a large enough change in the regression parameters triggered the definition of a new sector or sectors could be intelligently chosen based on the angular variation in roughness at the reference and target sites. Note that more efficient use of data may also impact on seasonal variability in cases where sector coverage is a significant factor (Section 5.3.4). A further limitation of the current study is the use of hourly averages of wind speed and direction. Variability on shorter timescales will have important impacts for wind power production and it would be informative to investigate whether the new MCP techniques (GPR and BW) proposed in this work are better able to capture this variability. However, there are barriers to carrying out such an investigation, not least the lack of availability of long-term reference/target site data with high temporal resolution. The current study was also limited with respect to the total number of sites considered in each terrain type. Hence, the impact of the terrain type on the performance of short-term MCP should be considered in more detail using a larger number of sites.

Several specific areas for further study can be identified from the Gaussian process regression (GPR) approach developed in Chapter 6. While GPR was only implemented with a single input, the reference site wind speed, the approach is capable of modelling non-linear relationships between multiple variables. Hence, it would be informative to investigate the use of inputs of wind speed and direction from multiple reference sites as well as parameters related to atmospheric stability. In addition, the application in Chapter 6 utilized a single prior mean function and covariance function. Hence, there is scope for the application of different prior functions which may be tailored to have specific properties. There is also scope for further exploration of the promising results obtained using orthogonal regression for predicting the distribution of target site wind angles.

A general point, also of relevance to the large-scale wind industry is the large number of MCP approaches that have been proposed in the literature, only a fraction of which appear to be used in the wind industry. As discussed in Chapter 3, these have been tested to widely varying degrees of rigour, making it hard to

objectively assess their performance. It would be beneficial to develop a large standardised data set that could be used to rigorously compare all existing MCP approaches using consistent criteria. Such a study could also be used as a benchmark against which new MCP techniques could be compared.

9.2.3 Output from an operational forecast model

The promising results reported in Chapter 8, using operational forecast data as a reference source for MCP, highlight several areas that could benefit from further study. While the average error metrics appear to be relatively robust (very little difference was observed after increasing the sites from 22-37) it would be informative to extend the approach to a much greater number of sites. Since no nearby reference site observations are required, the approach could be tested extensively throughout the UK using, for example, the MIDAS database of long-term meteorological observations. Such a study would allow a more detailed investigation of the effect of terrain type on performance and highlight situations where the approach is unlikely to be reliable. In particular, the improved performance at coastal sites should be investigated at a larger number of coastal locations to determine if the results presented in Chapter 8 are generally applicable. In addition, comparisons could be made using reanalysis data as an MCP reference source, in order to investigate the added value of the high resolution forecast model. It would also be informative to investigate the MCP performance of the higher resolution (1.5 km) UKV forecast model recently implemented by the Met Office. However, at present, archived UKV data does not cover a sufficiently long period to allow a detailed long-term study.

The application of UK4 data in a boundary layer scaling context is a further area where more detailed study is warranted. Simple downscaling methods capable of producing site-specific predictions at high resolution based on forecast data could have great utility globally, particularly in regions where dense networks of meteorological observations do not exist. Hence, work to optimise such methods as applied to forecast data, taking account of the issues raised in Chapter 4, as well as the complexities of the forecast data, would be particularly useful.

9.2.4 Energy production estimates

The scope of this thesis has been the estimation of the long-term wind resource as characterised by mean wind speed, distribution of wind speeds and the associated wind power density. In order to ensure generality, these were not converted to energy production estimates using specific turbine power curves. However, since

the efficiency of a turbine will vary with wind speed, the final uncertainty in the energy production will not necessarily be identical to the uncertainty in wind power density. This study could thus be extended using a selection of small wind turbine power curves in order to understand the impact of these uncertainties on energy production. In addition, estimates of the uncertainty in energy production would allow the financial implications, with respect to production uncertainties, of different wind resource assessment approaches to be evaluated. Such information would be valuable in weighing the costs and benefits of different assessment approaches.

9.3 Impacts

The work presented in this thesis can be seen as a step towards developing systematic and consistent approaches to small-scale wind resource assessment, while taking account of practical constraints related to costs and timescales.

The boundary layer scaling model investigated in Chapter 4 provides a low-cost and rapid approach that can be used as a scoping tool for testing multiple sites against a viability criterion. The methodology was shown to perform better than the legislated NOABL-MCS approach, and hence, consideration should be given to updating the guidance given in the microgeneration installer standard to take account of these developments. In addition, the results presented in Chapter 4 offer realistic estimates of the uncertainties likely to arise from applying a simple scaling approach and these should be explicitly communicated to investors to allow them to make better informed decisions.

The data-driven MCP approaches developed in Chapters 5-7 can be used to significantly reduce the uncertainty in the wind resource predictions, compared to a boundary layer scaling approach, using measurement periods as short as 3 months. The detailed analysis of the likely prediction errors as a function of training length and season can be used to make realistic appraisals of the added value of onsite measurements as well as the most appropriate training length and season to use. This will allow small-scale wind developers and investors to identify the most appropriate site assessment procedures to apply, given the funds and timescales available. In addition, the observation that linear regression frequently outperforms more sophisticated MCP approaches, providing the residual scatter is modelled, has implications for the development of new MCP algorithms. Specifically, linear regression with no representation of the residual scatter should not be considered as a suitable baseline against which to compare more sophisticated techniques.

The successful application of high resolution, operational forecast data to MCP, as detailed in Chapter 8, opens up the prospect of completing long-term wind resource assessments using only limited measurements obtained at the proposed turbine site. While it is not yet clear whether UK4 forecast data will be made publically available, this could have a significant impact on the ability of these approaches to be applied in regions where long-term reference data are sparse, or in regions with complex climates.

Overall, it is envisaged that the methodologies developed in this thesis could be combined into a systematic approach to small-scale wind resource assessment. As a minimum, a boundary layer scaling methodology should be deployed to test potential sites against a viability criterion. This would allow a first estimate of the available wind resource to be made along with estimates of the uncertainty, as set out in Chapter 4. Based on the results of this screening, and depending on the size of the investment, the most promising sites could be considered for short-term onsite measurements and long-term correlation using the MCP approaches described in Chapters 5-7. This would allow more precise wind power projections to be made. Such an approach would significantly reduce the chance of small-scale turbines being installed in non-viable locations and potentially make investments more attractive by reducing the financial risk without recourse to costly and time consuming site assessment procedures.

9.3.1 Taking responsibility

While the methods developed in this thesis can be considered low-cost, they still require investment related to development of expertise, reference data and, in the case of data-driven approaches, onsite measurements. Hence, there is an important question regarding who should take responsibility for developing and implementing these strategies.

For example, beyond preserving their long-term reputation, there may be little incentive for developers to implement systematic resource assessment strategies when installing small-scale turbines. It would be significantly easier to simply use a website calculator to provide a NOABL wind speed estimate and associated power projection based on a customer's postcode. In a competitive market, developers that apply more rigorous assessment procedures, the cost of which must ultimately be borne by the customer, may lose out to those that cut costs by using suspect methods. This is especially an issue since many potential customers may not have the relevant expertise to understand the difference between a NOABL estimate and one based on more rigorous assessment procedures.

One possible solution to this would be the provision of a free online implementation of the boundary layer scaling approach, hosted by a third party, to allow developers and customers to quickly and easily make a first estimate of the available wind resource. Note this was the original intention of the WYET previously hosted by the Carbon Trust. The enhancements to the original methodology outlined in Chapter 4 could be used to provide an improved implementation as well as realistic estimates of the likely uncertainties. This would improve transparency by allowing both developers and customers to quickly make a preliminary assessment of a site's potential as well as highlighting the uncertainties inherent in such an approach.

Such a tool could be coupled to an MCP solver where short-term measurements could be uploaded and correlated to long-term reference data for cases where increased certainty is required. The error metrics presented in this thesis could be used as a guide to the likely uncertainties using different MCP training lengths and seasons. The MCP algorithms could most simply be implemented using long-term forecast data as a reference source, as described in Chapter 8, since this would avoid issues related to identifying and obtaining suitable reference sites. While such an approach could in principle be automated, it would require investment in appropriate hardware as well as access to a large database of UK forecast data. Hence, the question remains as to who would facilitate the development of such a tool.

9.3.2 The big picture

Overall, even with the trend of increasing turbine size, the energy generated from small-scale wind turbines in the UK is likely to provide only a small part of the total energy that will be required from low carbon sources in the coming decades. However, wind turbines have become a highly visible icon of renewable energy, and hence, a thriving small-scale wind industry has the potential to contribute to increased awareness of the need for new energy choices as well as a shift in cultural norms related to energy use and generation. Clearly, the visibility of poorly located turbines that never spin, as well as stories of disenchanted customers, will be equally powerful in persuading people that wind energy is an expensive folly. Thus, those involved in the small-scale wind industry should work to ensure that the best available wind resource assessment approaches are applied systematically to potential turbine sites to ensure that turbines are installed at appropriate sites and with a full awareness of the uncertainties.

10 Bibliography

1. DECC, *Energy security strategy, URN 12D/349*, 2012, Department of Energy and Climate Change.
2. HM Government, *The Carbon Plan: Delivering our low carbon future*, 2011.
3. EC, *Directive 2009/28/EC, OJ L 140/16, 5.6.2009*, 2009.
4. HM Government, *Climate Change Act 2008*, 2008.
5. DECC, *Microgeneration Strategy*, 2011, Department of Energy and Climate Change,.
6. RenewableUK, *Small and Medium Wind UK Market Report*, 2013.
7. World Wind Energy Association, *Small Wind World Report*, 2013.
8. AWS Scientific Inc. & National Renewable Energy Laboratory (U.S.), *Wind Resource Assessment Handbook - fundamentals for conducting a successful monitoring program*, 1997.
9. DECC. *NOABL windspeed database*. [cited 2013 July]; Available from: http://tools.decc.gov.uk/en/windspeed/default.aspx.
10. Met Office. *UK Wind Map*. [cited 2013 November]; Available from: http://www.metoffice.gov.uk/renewables/wind-map.
11. Energy Saving Trust, *Location, Location, Location. Domestic small-scale wind field trial report*, 2009.
12. Encraft. *Warwick Wind Trials Final Report*. 2009 [cited 2013 August]; Available from: http://www.warwickwindtrials.org.uk/2.html.
13. Microgeneration Certification Scheme. *MIS 3003, Issue 3.2, 17/04/2013*. 2013 [cited 2013 August]; Available from: http://www.microgenerationcertification.org/installers/installers/installer-standards.
14. Heath, M.A., J.D. Walshe, and S.J. Watson, *Estimating the potential yield of small building-mounted wind turbines.* Wind Energy, 2007. **10**(3): p. 271-287.
15. Best, M., A. Brown, P. Clark, D. Hollis, D. Middleton, G. Rooney, D. Thomson, and C. Wilson, *Small-scale wind energy - technical report, UK Met Office*, 2008.
16. Weekes, S.M. and A.S. Tomlin, *Evaluation of a semi-empirical model for predicting the wind energy resource relevant to small-scale wind turbines.* Renewable Energy, 2013. **50**: p. 280-288.

17. Millward-Hopkins, J.T., A.S. Tomlin, L. Ma, D. Ingham, and M. Pourkashanian, *The predictability of above roof wind resource in the urban roughness sublayer.* Wind Energy, 2011: p. 225-243.
18. Carta, J.A., S. Velázquez, and P. Cabrera, *A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site.* Renewable and Sustainable Energy Reviews, 2013. **27**: p. 362-400.
19. Wilson, C., S. Vosper, and J. Standen, *Improved wind resource site-screening and planning.* European Wind Energy Association annual conference, Copenhagen, 2012.
20. UK Meteorological Office, *Virtual Met Mast verification report*, September 2011.
21. James, P.A.B., M.F. Sissons, J. Bradford, L.E. Myers, A.S. Bahaj, A. Anwar, and S. Green, *Implications of the UK field trial of building mounted horizontal axis micro-wind turbines.* Energy Policy, 2010. **38**(10): p. 6130-6144.
22. Sissons, M.F., P.A.B. James, J. Bradford, L.E. Myers, A.S. Bahaj, A. Anwar, and S. Green, *Pole-mounted horizontal axis micro-wind turbines: UK field trial findings and market size assessment.* Energy Policy, 2011. **39**(6): p. 3822-3831.
23. Walker, S.L., *Building mounted wind turbines and their suitability for the urban scale - a review of methods of estimating urban wind resource.* Energy and Buildings, 2011. **43**(8): p. 1852-1862.
24. DECC, *Energy Consumption in the UK - Domestic energy consumption in the UK between 1970 and 2012*, 2013, Department of Energy and Climate Change,.
25. Bortolini, M., M. Gamberi, A. Graziani, R. Manzini, and F. Pilati, *Performance and viability analysis of small wind turbines in the European Union.* Renewable Energy, 2014. **62**: p. 629-639.
26. Gipe, P., *Wind Power*. 2004, James and James Ltd: London.
27. Allen, S.R., G.P. Hammond, and M.C. McManus, *Energy analysis and environmental life cycle assessment of a micro-wind turbine.* Proceedings of the Institution of Mechanical Engineers Part A - Journal of Power and Energy, 2008. **222**(A7): p. 669-684.
28. Celik, A.N., T. Muneer, and P. Clarke, *An investigation into micro wind energy systems for their utilization in urban areas and their life cycle*

assessment. Proceedings of the Institution of Mechanical Engineers Part A - Journal of Power and Energy, 2007. **221**(A8): p. 1107-1117.

29. Rankine, R.K., J.P. Chick, and G.P. Harrison, *Energy and carbon audit of a rooftop wind turbine.* Proceedings of the Institution of Mechanical Engineers Part A - Journal of Power and Energy, 2006. **220**(A7): p. 643-654.
30. Bahaj, A.S., L. Myers, and P.A.B. James, *Urban energy generation: Influence of micro-wind turbine output on electricity consumption in buildings.* Energy and Buildings, 2007. **39**(2): p. 154-165.
31. Walters, R. and P.R. Walsh, *Examining the financial performance of micro-generation wind projects and the subsidy effect of feed-in tariffs for urban locations in the United Kingdom.* Energy Policy, 2011. **39**(9): p. 5167-5181.
32. HM Government. *Feed in Tariffs*. [cited 2013 December]; Available from: https://www.gov.uk/feed-in-tariffs.
33. Manwell, J.F., J.G. McGowan, and A.L. Rogers, *Wind Energy Explained: theory, design and application*. 2002, John Wiley and Sons Ltd: Chichester.
34. Burton, T., D. Sharpe, N. Jenkins, and E. Bossanyi, *Wind Energy Handbook* 2004: John Wiley and Sons.
35. Van der Hoven, I., *Power spectrum of horizontal wind speed in frequency range from 0.0007 to 900 cycles per hour.* Journal of Meteorology, 1957. **14**(2): p. 160-164.
36. Watson, S., *Quantifying the variability of wind energy.* Wiley Interdisciplinary Reviews: Energy and Environment, 2013.
37. Martin-Martínez, S., A. Vigueras-Rodríguez, E. Gómez-Lázaro, A. Molina-García, E. Muljadi, and M. Milligan, *Wind Power Variability and Singular Events*. Advances in Wind Power. 2012.
38. Carta, J.A., P. Ramirez, and S. Velazquez, *A review of wind speed probability distributions used in wind energy analysis - Case studies in the Canary Islands.* Renewable & Sustainable Energy Reviews, 2009. **13**(5): p. 933-955.
39. Weibull, W., *A Statistical Distribution Function of Wide Applicability.* Journal of Applied Mechanics-Transactions of the Asme, 1951. **18**(3): p. 293-297.
40. Ramirez, P. and J.A. Carta, *Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study.* Energy Conversion and Management, 2005. **46**(15-16): p. 2419-2438.

41. Justus, C.G., W.R. Hargraves, and A. Yalcin, *Nationwide assessment of potential output from wind-powered generators.* Journal of Applied Meteorology, 1976. **15**(7): p. 673-678.
42. Chang, T.P., *Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application.* Applied Energy, 2011. **88**(1): p. 272-282.
43. Ramirez, P. and J.A. Carta, *The use of wind probability distributions derived from the maximum entropy principle in the analysis of wind energy. A case study.* Energy Conversion and Management, 2006. **47**(15-16): p. 2564-2577.
44. Al-Fawzan, M.A., *Methods for estimating the parameters of the Weibull distribution.* InterStat, 2000 **Vol 1**.
45. Seguro, J.V. and T.W. Lambert, *Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis.* Journal of Wind Engineering and Industrial Aerodynamics, 2000. **85**(1): p. 75-84.
46. Tuller, S.E. and A.C. Brett, *The characteristics of wind velocity that favor the fitting of a weibull distribution in wind-speed analysis.* Journal of Climate and Applied Meteorology, 1984. **23**(1): p. 124-134.
47. Celik, A.N., A. Makkawi, and T. Muneer, *Critical evaluation of wind speed frequency distribution functions.* Journal of Renewable and Sustainable Energy, 2010. **2**(1).
48. Chang, T.P., *Estimation of wind energy potential using different probability density functions.* Applied Energy, 2011. **88**(5): p. 1848-1856.
49. Technical University of Denmark. *WAsP - the wind atlas, analysis and application program*. 2013 [cited 2013 December]; Available from: http://www.wasp.dk/.
50. Carta, J.A. and P. Ramirez, *Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago.* Energy Conversion and Management, 2007. **48**(1): p. 281-291.
51. Akdag, S.A., H.S. Bagiorgas, and G. Mihalakakou, *Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean.* Applied Energy, 2010. **87**(8): p. 2566-2573.
52. Britter, R.E. and S.R. Hanna, *Flow and dispersion in urban areas.* Annual Review of Fluid Mechanics, 2003. **35**: p. 469-496.
53. Oke, T.R., *Boundary Layer Climates*. 1987, Methuen and Co.: New York.

54. Bergström, H., P.-E. Johansson, and A.-S. Smedman, *A study of wind speed modification and internal boundary layer heights in a coastal region.* Boundary-Layer Meteorology, 1988. **42**(4): p. 313-335.
55. Baldocchi, D. *Wind and Turbulence, Part 2, surface boundary layer theory and principles (lecture series)*. [University of California, Berkley. Unpublished work] 2012 [cited 2013 December]; Available from: http://nature.berkeley.edu/biometlab/espm129/notes/.
56. Millward-Hopkins, J.T., *Predicting the wind resource available to roof-mounted wind turbines in urban areas*. 2013, PhD thesis, Energy Research Institute, University of Leeds.
57. Bradshaw, P., *Possible origin of Prandtl's mixing-length theroy.* Nature, 1975. **249**: p. 135-136.
58. Garratt, J.R., *The Internal Boundary-Layer - A Review.* Boundary-Layer Meteorology, 1990. **50**(1-4): p. 171-203.
59. Rao, K.S., J.C. Wyngaard, and O.R. Coté, *The Structure of the Two-Dimensional Internal Boundary Layer over a Sudden Change of Surface Roughness.* Journal of the Atmospheric Sciences, 1974. **31**(3): p. 738-746.
60. Peterson, E.W., *Modification of mean flow and turbulent energy by a change in surface roughness under conditions of neutral stability.* Quarterly Journal of the Royal Meteorological Society, 1969. **95**(405): p. 561-575.
61. Elliot, W.P., *The growth of the atmospheric internal boundary layer.* Transactions of the American Geophysical Union, 1958. **39**: p. 1048-1054.
62. Bradley, E.F., *A micrometeorological study of velocity profiles and surface drag in the region modified by a change in surface roughness.* Quarterly Journal of the Royal Meteorological Society, 1968. **94**(401): p. 361-379.
63. Goode, K. and S.E. Belcher, *On the parameterisation of the effective roughness length for momentum transfer over heterogeneous terrain.* Boundary-Layer Meteorology, 1999. **93**(1): p. 133-154.
64. Wieringa, J., *An objective exposure correction method for average wind speeds measured at a sheltered location.* Quarterly Journal of the Royal Meteorological Society, 1976. **102**(431): p. 241-253.
65. Mason, P.J., *The formation of areally-averaged roughness lengths.* Quarterly Journal of the Royal Meteorological Society, 1988. **114**(480): p. 399-420.
66. Cheng, H. and I.P. Castro, *Near wall flow over urban-like roughness.* Boundary-Layer Meteorology, 2002. **104**(2): p. 229-259.

67. Schmid, H.P. and T.R. Oke, *A model to estimate the source area contributing to turbulent exchange in the surface-layer over patchy terrain.* Quarterly Journal of the Royal Meteorological Society, 1990. **116**(494): p. 965-988.
68. Bou-Zeid, E., M.B. Parlange, and C. Meneveau, *On the Parameterization of Surface Roughness at Regional Scales.* Journal of the Atmospheric Sciences, 2007. **64**(1): p. 216-227.
69. Barthelmie, R.J. and J.P. Palutikof, *Coastal wind speed modelling for wind energy applications.* Journal of Wind Engineering and Industrial Aerodynamics, 1996. **62**(2–3): p. 213-236.
70. Barthelmie, R.J., *The effects of atmospheric stability on coastal wind climates.* Meteorological Applications, 1999. **6**(1): p. 39-47.
71. Sempreviva, A., S.E. Larsen, N.G. Mortensen, and I. Troen, *Response of neutral boundary layers to changes of roughness.* Boundary-Layer Meteorology, 1990. **50**(1-4): p. 205-225.
72. Van Wijk, A.J.M., A.C.M. Beljaars, A.A.M. Holtslag, and W.C. Turkenburg, *Evaluation of stability corrections in wind speed profiles over the North Sea.* Journal of Wind Engineering and Industrial Aerodynamics, 1990. **33**(3): p. 551-566.
73. McMahon, S. and S. Watson, *A comparison of the correlation between onshore and offshore wind speed data.* European Wind Energy Association annual conference, Vienna, 2013.
74. Barthelmie, R.J., B. Grisogono, and S.C. Pryor, *Observations and simulations of diurnal cycles of near-surface wind speeds over land and sea.* Journal of Geophysical Research: Atmospheres, 1996. **101**(D16): p. 21327-21337.
75. Grimmond, C.S.B. and T.R. Oke, *Aerodynamic properties of urban areas derived from analysis of surface form.* Journal of Applied Meteorology, 1999. **38**: p. 1262-1292.
76. Grimmond, C.S.B. and T.R. Oke, *Turbulent heat fluxes in urban areas: Observations and a local-scale urban meteorological parameterization scheme (LUMPS).* Journal of Applied Meteorology, 2002. **41**(7): p. 792-810.
77. Kastner-Klein, P. and M. Rotach, *Mean flow and turbulence characteristics in an urban roughness sublayer.* Boundary-Layer Meteorology, 2004. **111**: p. 55-84.

78. Oikawa, S. and Y. Meng, *Turbulence Characteristics and Organized Motion in a Suburban Roughness Sublayer.* Boundary-Layer Meteorology, 1995. **74**(3): p. 289-312.
79. Cheng, H., P. Hayden, A.G. Robins, and I.P. Castro, *Flow over cube arrays of different packing densities.* Journal of Wind Engineering and Industrial Aerodynamics, 2007. **95**(8): p. 715-740.
80. Macdonald, R.W., *Modelling The Mean Velocity Profile In The Urban Canopy Layer.* Boundary-Layer Meteorology, 2000. **97**(1): p. 25-45.
81. Grimmond, C.S.B., T.S. King, M. Roth, and T.R. Oke, *Aerodynamic roughness of urban areas derived from wind observations.* Boundary-Layer Meteorology, 1998. **89**(1): p. 1-24.
82. MacDonald, R.W., R.F. Griffiths, and D.J. Hall, *An improved method for the estimation of surface roughness of obstacle arrays.* Atmospheric Environment, 1998. **32**(11): p. 1857-1864.
83. Raupach, M.R., *Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index.* Boundary-Layer Meteorology, 1994. **71**(1-2): p. 211-216.
84. Oke, T.R., *Street design and urban canopy layer climate.* Energy and Buildings, 1988. **11**(1–3): p. 103-113.
85. Millward-Hopkins, J.T., A.S. Tomlin, L. Ma, D. Ingham, and M. Pourkashanian, *Estimating aerodynamic parameters of urban-like surfaces with heterogeneous building heights.* Boundary-Layer Meteorology, 2011. **141**(3): p. 443-465.
86. Millward-Hopkins, J.T., A.S. Tomlin, L. Ma, D.B. Ingham, and M. Pourkashanian, *Aerodynamic parameters of a UK city derived from morphological data.* Boundary-Layer Meteorology, 2013. **146**(3): p. 447-468.
87. Millward-Hopkins, J.T., A.S. Tomlin, L. Ma, D.B. Ingham, and M. Pourkashanian, *Mapping the wind resource over UK cities.* Renewable Energy, 2013. **55**: p. 202-211.
88. Taylor, P.A. and R.J. Lee, *Simple guidelines for estimating wind speed variations due to small scale topographical features.* Climatological Bulletin, 1984. **18**(2): p. 3-32.
89. Jackson, P.S. and J.C.R. Hunt, *Turbulent wind flow over a low hill.* Quarterly Journal of the Royal Meteorological Society, 1975. **101**(430): p. 929-955.
90. Sherman, C.A., *A mass-consistent model for wind fields over complex terrain.* Journal of Applied Meteorology, 1978. **17**(3): p. 312-319.

91. Howard, T. and P. Clark, *Correction and downscaling of NWP wind speed forecasts.* Meteorological Applications, 2007. **14**(2): p. 105-116.
92. Wong, R., S. Webster, and S. Vosper, *High-resolution dynamical downscaling techniques for wind resource assessment.* European Wind Energy Association annual conference, Copenhagen, 2012.
93. The Carbon Trust, *Small-scale wind energy - policy insights and practical guidance*, 2008.
94. Perry, M. and D. Hollis, *The development of a new set of long-term climate averages for the UK.* International Journal of Climatology, 2005. **25**(8): p. 1023-1039.
95. Centre for Ecology and Hydrology. *Land Cover Map 2007 (25m raster, GB)*. [cited 2013 November]; Available from: http://www.ceh.ac.uk.
96. Skibin, D., *On the evaluation of wind power from short wind records - comment.* Journal of Climate and Applied Meteorology, 1984. **23**(10): p. 1477-1479.
97. Liléo, S., *How to measure representativeness.* European Wind Energy Association - wind resource assessment workshop, Dublin, 2013.
98. Watson, S.J., P. Kritharas, and G.J. Hodgson, *Wind speed variability across the UK between 1957 and 2011.* Wind Energy, DOI: 10.1002/we.1679, 2013.
99. Rogers, A.L., J.W. Rogers, and J.F. Manwell, *Comparison of the performance of four measure-correlate-predict algorithms.* Journal of Wind Engineering and Industrial Aerodynamics, 2005. **93**(3): p. 243-264.
100. Putnam, P., *Power from the wind*. 1948, D. Van Nostrand Company: New York.
101. Justus, C.G., K. Mani, and A.S. Mikhail, *Interannual and month-to-month variations of wind speed.* Journal of Applied Meteorology, 1979. **18**(7): p. 913-920.
102. Anderson, M. and J. Bass, *Renewable Energy Systems (RES). A review of MCP techniques*, 2004.
103. Thøgersen, M., M. Motta, T. Sørensen, and P. Nielsen, *Measure-correlate-predict methods: case studies and software implementation.* European Wind Energy Association annual conference, Milan, 2007.
104. Schlez, W., *GH WindFarmer theory manual - Garrad Hassan and Partners Ltd.* 2009.
105. ReSoft. *WindFarm: wind farm anlysis, design and optimisation*. [cited 2013; Available from: www.resoft.co.uk.

106. Achberger, C., M. Ekstrom, and L. Barring, *Estimation of local near-surface wind conditions - a comparison of WASP and regression based techniques.* Meteorological Applications, 2002. **9**(2): p. 211-221.
107. Derrick, A., *Development of the measure-correlate-predict strategy for site assessment.* British Wind Energy Association annual conference, 1992: p. 259-265.
108. Bardsley, W.E. and B.F.J. Manly, *Regression-based estimation of long-term mean and variance of wind-speed at potential aerogenerator sites.* Journal of Climate and Applied Meteorology, 1983. **22**(2): p. 323-327.
109. Rogers, A.L., J.W. Rogers, and J.F. Manwell, *Uncertainties in results of measure-correlate-predict analyses.* European Wind Energy Association annual conference, Athens, 2006.
110. Romo Perea, A., J. Amezcua, and O. Probst, *Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource.* Journal of Renewable and Sustainable Energy, 2011. **3**(2).
111. Lackner, M.A., A.L. Rogers, and J.F. Manwell, *The round robin site assessment method: A new approach to wind energy site assessment.* Renewable Energy, 2008. **33**(9): p. 2019-2026.
112. Jung, S., O. Arda Vanli, and S.D. Kwon, *Wind energy potential assessment considering the uncertainties due to limited data.* Applied Energy, 2013. **102**: p. 1492-1503.
113. Ellison, S.L.R., V.J. Barwick, and T.J.D. Farrant, *Practical statistics for the analytical scientist, a bench guide*. 2009, RSC Publishing.
114. Moritmer, A., *A new correlation/prediction method for potential wind farm sites.* Proceedings of the BWEA, 1994.
115. Woods, J.C. and S.J. Watson, *A new matrix method of predicting long-term wind roses with MCP.* Journal of Wind Engineering and Industrial Aerodynamics, 1997. **66**(2): p. 85-94.
116. Clive, P., *Non-linearity in MCP with Weibull distributed wind speeds.* Wind Engineering, 2008. **32**(3): p. 319-323.
117. Hanson, B., K. Klink, K. Matsuura, S.M. Robeson, and C.J. Willmott, *Vector Correlation - Review, Exposition, and Geographic Application.* Annals of the Association of American Geographers, 1992. **82**(1): p. 103-116.
118. Nielsen, M., L. Landberg, N.G. Mortensen, R.J. Barthelmie, and A. Joensen, *Application of measure-correlate-predict approach for wind resource measurement.* European Wind Energy Association Annual Conference, Copenhagen, 2001.

119. García-Rojo, R., *Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach.* Wind Engineering, 2004. **28**(2): p. 213-224.
120. Casella, L., *A new method based on joint probability functions for long term wind resource estimation.* European Wind Energy Association annual conference, Copenhagen, 2012.
121. Carta, J.A. and S. Velazquez, *A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site.* Energy, 2011. **36**(5): p. 2671-2685.
122. Kalogirou, S.A., *Artificial neural networks in renewable energy systems applications: a review.* Renewable & Sustainable Energy Reviews, 2001. **5**(4): p. 373-401.
123. Bechrakis, D.A., J.P. Deane, and E.J. McKeogh, *Wind resource assessment of an area using short term data correlated to a long term data set.* Solar Energy, 2004. **76**(6): p. 725-732.
124. Lopez, P., R. Velo, and F. Maseda, *Effect of direction on wind speed estimation in complex terrain using neural networks.* Renewable Energy, 2008. **33**(10): p. 2266-2272.
125. Velazquez, S., J.A. Carta, and J.M. Matias, *Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kW h produced by a wind turbine at a candidate site: A case study in the Canary Islands.* Applied Energy, 2011. **88**(11): p. 3869-3881.
126. Bass, J.H., M. Rebbeck, L. Landberg, M. Cabre, and A. Hunter, *An improved measure-correlate-predict algorithm for the prediction of the long term wind climate in regions of complex environment*, 2000, Joule Project JOR3-CT98-0295.
127. Williams, C.K.I. and C.E. Rasmussen, *Gaussian processes for regression*, in *Advances in Neural Information Processing Systems 8: Proceedings of the 1995 Conference*, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, Editors. 1996, M I T Press: Cambridge. p. 514-520.
128. Kirk, P.D.W. and M.P.H. Stumpf, *Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data.* Bioinformatics, 2009. **25**(10): p. 1300-1306.
129. Stegle, O., S.V. Fallert, D.J.C. MacKay, and S. Brage, *Gaussian process robust regression for noisy heart rate data.* IEEE Transactions on Biomedical Engineering, 2008. **55**(9): p. 2143-2151.

130. Jiang, X., B. Dong, L. Xie, and L. Sweeney, *Adaptive Gaussian process for short-term wind speed forecasting.* Proceedings of the 2010 conference on ECAI: 19th European Conference on Artificial Intelligence, 2010: p. 661-666.
131. Mori, H. and E. Kurata, *Application of Gaussian process to wind speed forecasting for wind power generation.* 2008 IEEE International Conference on Sustainable Energy Technologies, 2008: p. 956-959.
132. Kou, P., F. Gao, and X.H. Guan, *Sparse online warped Gaussian process for wind power probabilistic forecasting.* Applied Energy, 2013. **108**: p. 410-428.
133. Rasmussen and Williams, *Gaussian processes for machine learning*. 2006, ISBN-10 0-262-18253-X.
134. Taylor, M., P. Mackiewicz, M. Brower, and M. Markus, *An analysis of wind resource uncertainty in energy production esitmates.* European Wind Energy Association annual conference, London, 2004.
135. Barros, V.R. and E.A. Estevan, *On the evaluation of wind power from short wind records.* Journal of Climate and Applied Meteorology, 1983. **22**(6): p. 1116-1123.
136. Barros, V.R. and E.A. Estevan, *On the evaluation of wind power from short wind records - reply.* Journal of Climate and Applied Meteorology, 1984. **23**(10): p. 1480-1483.
137. Oliver, A. and K. Zarling, *The effect of seasonality on wind speed prediction bias in the plains*, in *American Wind Energy Assocciation,* 2010: Dallas.
138. Thomson, D.: Met Office Fellow and Research Scientist, Personal communication (March 2012).
139. Panofsky, H.A. and A.A. Townsend, *Change of terrain roughness and the wind profile.* Quarterly Journal of the Royal Meteorological Society, 1964. **90**(384): p. 147-155.
140. Met Office, *Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853 - current). NCAS British Atmospheric Data Centre 2012.*
141. CEDA WPS. [cited 2013 November]; Available from: http://ceda-wps2.badc.rl.ac.uk/ui/home.
142. Met Office. *Met Office surface data users guide*. [cited 2013 August]; Available from: http://badc.nerc.ac.uk/data/ukmo-midas/ukmo_guide.html#7.
143. Mertens and S. Mertens, *The energy yield of roof mounted wind turbines.* Wind Engineering, 2003. **27**(6): p. 507-518.

144. Microgeneration Certification Scheme. *MIS 3003, Issue 1.4, 10/01/2009*. 2009 [cited 2013 August]; Available from: http://www.microgenerationcertification.org/installers-manufacturers/2-uncategorised/123-archived-installers-standards.

145. Landmap. *Spatial Discovery. Available through the Cities Revealed agreement (Cities Revealed (C) the Geoinformation Group 2008)*. 2008 [cited 2012 February]; Available from: http://www.landmap.ac.uk.

146. Millward-Hopkins, J.T., A.S. Tomlin, L. Ma, D.B. Ingham, and M. Pourkashanian, *Assessing the potential of urban wind energy in a major UK city using an analytical model.* Renewable Energy 2013. **60**: p. 701-710.

147. Sobol, I.M., *On the distribution of points in a cube and the approximate evaluation of integrals.* U.S.S.R. Computational Mathematics and mathematical physics, 1967. **7**(4): p. 86-112.

148. Ziehn, T. and A.S. Tomlin, *GUI-HDMR - A software tool for global sensitivity analysis of complex models.* Environmental Modelling & Software, 2009. **24**(7): p. 775-785.

149. Drew, D.R., J.F. Barlow, and T.T. Cockerill, *Estimating the potential yield of small wind turbines in urban areas: A case study for Greater London, UK.* Journal of Wind Engineering and Industrial Aerodynamics, 2013. **115**: p. 104-111.

150. RenewableUK, *Generate your own power*, 2010.

151. Carta, J.A., S. Velázquez, and J.M. Matías, *Use of Bayesian networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site.* Energy Conversion and Management, 2011. **52**(2): p. 1137-1149.

152. Bludszuweit, H., J.A. Dominguez-Navarro, and A. Llombart, *Statistical analysis of wind power forecast error.* IEEE Transactions on Power Systems, 2008. **23**(3): p. 983-991.

153. Earl, N., S. Dorling, R. Hewston, and R. von Glasow, *1980-2010 Variability in UK surface wind climate.* Journal of Climate, 2013. **26**(4): p. 1172-1191.

154. MacKay, D.J.C., *Introduction to Gaussian processes*. Neural Networks and Machine Learning, NATO ASI Series, ed. C.M. Bishop. Vol. 168. 1998, Berlin: Springer.

155. Rasmussen, C.E. and H. Nickisch. *Gaussain Processes for machine learning, Matlab version 3.0-2010-07-23*. 2010; Available from: http://www.gaussianprocess.org/gpml/code/matlab/doc/.

156. Velázquez, S., J.A. Carta, and J.M. Matías, *Influence of the input layer signals of ANNs on wind power estimation for a target site: A case study.* Renewable and Sustainable Energy Reviews, 2011. **15**(3): p. 1556-1566.
157. EMD International WindPRO. [cited 2013 April]; Available from: www.emd.dk.
158. Lu, J.C. and G.K. Bhattacharyya, *Some new constructions of bivariate Weibull models.* Annals of the Institute of Statistical Mathematics, 1990. **42**(3): p. 543-559.
159. Johnson, R.A., J.W. Evans, and D.W. Green, *Some bivariate distributions for modeling the strength properties of lumber.* United States Department of Agricutlure, Forest Service, Research Paper FOL-RP-575, 1999.
160. Lee, L., *Multivariate distributions having Weibull properties.* Journal of Multivariate Analysis, 1979. **9**(2): p. 267-277.
161. Weekes, S.M. and A.S. Tomlin, *Data efficient measure-correlate-predict approaches to wind resource assessment for small-scale wind energy.* Renewable Energy 2014. **63**: p. 162-171.
162. Dorvlo, A.S.S., *Estimating wind speed distribution.* Energy Conversion and Management, 2002. **43**(17): p. 2311-2318.
163. J. M. Rodrigo, *State-of-the-art wind resource assessment, WAUDIT Deliverable D-7*, 2010, CENER National renewable energy centre.
164. Brower, M.C., *The use of NCEP/NCAR reanalysis data in MCP.* European Wind Energy Association annual conference, Athens, 2006.
165. Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, *The NCEP/NCAR 40-year reanalysis project.* Bulletin of the American Meteorological Society, 1996. **77**(3): p. 437-471.
166. Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Holm, L. Isaksen, P. Kallberg, M. Koehler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.N. Thepaut, and F. Vitart, *The ERA-Interim reanalysis: configuration and performance of the data assimilation system.* Quarterly Journal of the Royal Meteorological Society, 2011. **137**(656): p. 553-597.

167. University Corporation for Atmospheric Research (UCAR). *Climate Data Guide*. [cited 2013 October]; Available from: https://climatedataguide.ucar.edu/.

168. Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C.R. Redder, R. Reichle, F.R. Robertson, A.G. Ruddick, M. Sienkiewicz, and J. Woollen, *MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications.* Journal of Climate, 2011. **24**(14): p. 3624-3648.

169. Jimenez, B., K. Moennich, and F. Durante, *Comparison between NCEP/NCAR and MERRA reanalysis data for long-term correction in wind energy assessment.* European Wind Energy Association annual conference, Copenhagen, 2012.

170. Liléo, S. and O. Petrick, *Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis.* European Wind Energy Association annual conference, Brussels, 2011.

171. Pinto, C., R. Guedes, P. Pinto, and M. Ferreira, *NCEP/NCAR reanalysis data for the Portuguese mainland.* European Wind Energy Association annual conference, Athens, 2006.

172. Taylor, M., J. Freedman, K. Waight, and M. Brower, *Using simulated wind data from a mesoscale model in MCP*, 2009, AWS Truewind.

173. Davies, T., M.J.P. Cullen, A.J. Malcolm, M.H. Mawson, A. Staniforth, A.A. White, and N. Wood, *A new dynamical core for the Met Office's global and regional modelling of the atmosphere.* Quarterly Journal of the Royal Meteorological Society, 2005. **131**(608): p. 1759-1782.

174. Met Office. *Numerical Models in Meteorology*. [cited 2013 October]; Available from: http://www.metoffice.gov.uk/research/modelling-systems/numerical-models.

175. Wilson, C. and J. Standen, *The added value and validation of mesoscale models compared to atmospheric reanalyses for resource assessment.* European Wind Energy Association annual conference, Vienna, 2013.

176. Landberg, L. and S.J. Watson, *Short-term prediction of local wind conditions.* Boundary-Layer Meteorology, 1994. **70**(1-2): p. 171-195.

177. Hume-Wright, L., P.M. Lee, and A. Skea, *Accounting for uncertainty unquantified in MCP.* European Wind Energy Association annual conference, Vienna, 2013.

178. Met Office. *Virtual Met Mast and Met Mast Plus*. 2012 [cited 2013 October]; Available from: http://www.metoffice.gov.uk/media/pdf/i/e/12_0341_Virtual_Metmast___VMMPLUS-v4_FINAL.pdf.

179. Met Office, *Global assimilation of air temperature, humidity, wind and pressure from surface stations: practice and performance.* Forecasting Research Technical Report No: 582, 2013.

180. Macpherson, B., B.J. Wright, W.H. Hand, and A.J. Maycock, *The impact of MOPS moisture data in the U.K. Meteorological Office mesoscale data assimilation scheme.* Monthly Weather Review, 1996. **124**(8): p. 1746-1766.

181. Auría, A.C. and B.N. Fernández-Victorio, *Usage of SYNOP 10 meters wind observations in HIRLAM.* HIRLAM Newsletter, 2009. **50**(03): p. 8-17.

182. Taylor, G.I., *The spectrum of turbulence.* Proceedings of the Royal Society of London A, 1938. **164**.