Developing a SARIMAX model for monthly wind speed forecasting in the UK

by

Petros Kritharas

BSc (Hons), MSc

A doctoral thesis submitted in partial fulfilment of the requirements
for the award of Doctor of Philosophy of
Loughborough University.
August 2013

This thesis was typeset in XeTeX and its format and style were based on the PhD template from the Electrical Engineering Department of The University of Edinburgh, authored by George Taylor.

Declaration of originality

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements, footnotes or references, and that neither the thesis or the original work contained therein has been submitted to this or any other institution for a degree.

Signature:____________________________

Date:________________________________

Πέτρος Κριθαρᾶς

Petros Kritharas

Abstract

Wind is a fluctuating source of energy and, therefore, it can cause several technical impacts. These can be tackled by forecasting wind speed and thus wind power. The introduction of several statistical models in this field of research has brought to light promising results for improving wind speed predictions. However, there is not converging evidence on which is the optimal method. Over the last three decades, significant research has been carried out in the field of short-term forecasting using statistical models though less work focuses on longer timescales.

The first part of this work concentrated on long-term wind speed variability over the UK. Two subsets have been used for assessing the variability of wind speed in the UK on both temporal and spatial coverage over a period representative of the expected lifespan of a wind farm. Two wind indices are presented with a calculated standard deviation of 4% . This value reveals that such changes in the average UK wind power capacity factor is equal to 7%.

A parallel line of the research reported herein aimed to develop a novel statistical forecasting model for generating monthly mean wind speed predictions. It utilised long-term historic wind speed records from surface stations as well as reanalysis data. The methodology employed a SARIMAX model that incorporated monthly autocorrelation of wind speed and seasonality, and also included exogenous inputs. Four different cases were examined, each of which incorporated different independent variables. The results disclosed a strong association between the independent variables and wind speed showing correlations up to 0.72. Depending on each case, this relationship occurred from 4− up to 12−month lags. The inter comparison revealed an improvement in the forecasting accuracy of the proposed model compared to a similar model that did not take into account exogenous variables. This finding demonstrates the indisputable potential of using a SARIMAX for long-term wind speed forecasting.

Acknowledgements

I would like to thank my supervisor Prof. Simon J. Watson for demonstrating great capacity of probity and academic integrity at all times over the duration of this research. His guidance and continued support have been invaluable.

I must also express my gratitude to the funding bodies of this project; the School of Electronic, Electrical and Systems Engineering (former Department of Electronic and Electrical Engineering) at Loughborough University and the SUPERGEN Wind Energy Technologies Consortium established by the Engineering and Physical Sciences Research Council.

Many thanks to Good Energy Ltd for all the technical information and data contributed to this research.

I also wish to express my sincere thanks to the UK Met Office and the BADC for providing access to the UKCIP gridded dataset and the MIDAS data, respectively. The staff of the Met Office Archive in Exeter should also be acknowledged for providing access to the records related to the meteorological stations used in this research.

Also, I am grateful to the staff and postgraduate students of the Centre for Renewable Energy Systems Technology (CREST) at Loughborough University for creating a stimulating working environment, as well as for their friendship.

Above all, I would like to thank Λέλα, Γιάννη and Μυρτώ for their unwavering support as well as their love and care.

"ἐλαφρὸν ὅστις πημάτων ἔξω πόδα
ἔχει παραινεῖν νουθετεῖν τε τὸν κακῶς
πράσσοντ'· ἐγὼ δὲ ταῦθ' ἅπαντ' ἠπιστάμην.
ἑκὼν ἑκὼν ἥμαρτον, οὐκ ἀρνήσομαι·
θνητοῖς ἀρήγων αὐτὸς ηὑρόμην πόνους."

He who stands free with an untrammelled foot is quick to counsel and exhort a friend in trouble. But all these things I know well. Of my free will, my own free will, I erred, and freely do I here acknowledge it. Freeing mankind myself have durance found.

Προμηθεὺς Δεσμώτης
Αἰσχύλος, 525/524-456/455 Π.Κ.Ε

Prometheus Bound
Aeschylus, 525/524 - 456/455 B.C.E

Contents

List of Figures

List of Tables

Acronyms and abbreviations

agl	Above Ground Level
ADF	Augmented Dickey Fuller test
AIC	Akaike's Information Criterion
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AR	Autoregressive
ARMA	Autoregressive Moving Average model
ARIMA	Autoregressive Integrated Moving Average model
ARIMA-ANN	Autoregressive Integrated Moving Average - Artificial Neural Network model
ARMAX	Autoregressive Moving Average with Exogenous Inputs model
ARX	Autoregressive with Exogenous Inputs model
BADC	British Atmospheric Data Centre
BETTA	British Transmission and Trading Arrangements
BP	Backpropagation
CF	Capacity Factor
CORINE	COoRdinate INformation on the Environment
CLC	CORINE Land Cover
CREST	Centre for Renewable Energy Systems Technology
DF	Dickey Fuller
EEA	European Environment Agency
ECWWF	European Centre for Medium-range Weather Forecasts
EKF	Extended Kalman Filter
ERA-40	ECMWF Reanalysis-40
ERM	Empirical Risk Minimisation
EWEA	European Wind Energy Association
FFT	Fast Fourier Transform

GM	Grey model
GDP	Gross Domestic Product
GWEC	Global Wind Energy Council
IEA	International Energy Agent
KF	Kalman Filter
LLS	Least Square
LLSSVM	Least Square Support Vector Machine
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
MIDAS	Met office Integrated Data Archive System
MLP	Mutilayer Perceptron
MMU	Minimum Mapping Unit
MSE	Mean Square Error
MSL	Mean Sea Level Pressure
NAO	North Atlantic Oscillation
NWP	Numerical Weather Prediction
ODBC	Open Database Connectivity
PNA	Pacific North American
RDBS	Relational Database Management System
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive with Integrated Moving Average
SARIMAX	Seasonal Autoregressive Moving Average with Exogenous Input
SE	Standard Error
SSE	Sum of Squared Errors
SES	Simple Exponential Smoothing
SS	Simple Seasonal
SST	Sea Surface Temperature
SO	System Operator

SRM	Structural Risk Minimisation
SVM	Support Vector Machine
UKCIP	UK Climate Impacts Programme

Nomenclature and Glossary

$^\circ$	A degree of arc
CO_2	Carbon dioxide
$^\circ C$	Degree(s) Celcius
GB	Gigabyte is a unit for measuring digital storage equal to one billion Bytes
GW	Gigawatt is a unit for measuring electric power equal to one billion Watts or 1,000 megawatts (MW)
ha	Hectare is a unit for measuring area equal to 10,000 squared meters (m^2)
km	Kilometer is a unit for measuring length equal to 1,000 meters (m)
kt	Knots is a unit for measuring speed equal to 0.514 ms^{-1}
L	Lagging operator
MHz	Megahertz is a unit for measuring frequency equal to one million Hertz
MW	Megawatt is a unit for measuring electric power equal to one million Watts or 1,000 kilowatts (KW)
MWh	Megawatt hour is a unit for measuring electrical energy equal to one million Watt hours or 1,000 kilowatt hours (kWh)
R^2	Coefficient of determination (is a measure of how well the regression line fits the data)
u	Wind speed in ms^{-1}
u'	Zonal or eastwards wind component in ms^{-1}
u_*	Friction velocity in the surface layer in ms^{-1}
v'	Meridional or northwards wind component in ms^{-1}
z	Height of interest in m
z_o	Surface roughness length in m
∇	Differencing operator
∇_s^D	Seasonal differencing operator

Greek Letters

κ	von Kármán's constant (assumed = 0.4)
ρ	Correlation coefficient (is a measure of how two variables are linearly related defined as the covariance of the variables divided by the product of their σ): $\rho_{X,Y} = \frac{Cov_{X,Y}}{\sigma_X \sigma_X}$
σ	Standard deviation (is a measure of how much dispersed are data from the average value, defined as the square root of the variance): $\sqrt{\sigma^2}$
σ^2	Variance (is a measure of how data are spread out, defined as the average of the squared differences from the average or the mean): $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}$ for the sample and $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$ for the population
$\varphi_p L^p$	AR polynomial of L of order p
$\theta_q L^q$	MA polynomial of L of order q
$\Phi_P L^P$	Seasonal AR polynomial of L^P of order P
$\Theta_Q L^Q$	Seasonal MA polynomial of L^Q of order Q
$\xi_b L$	X polynomial of L of order b

Published Work

The following papers document the results of this research.

- P. Kritharas and S. J. Watson, ”Long Term Forecasting of Wind Speed Using Historical Patterns,” in *European Wind Energy Conference and Exhibition (EWEC)*, (Marseille, France), March 16-19 2009.

- P. Kritharas and S. J. Watson, ”Long term wind speed forecasting based on seasonal trends,” in *Proceedings of the ASME* 3^{rd} *Int. Conf. on Energy Sustainability*, vol. 2, (San Francisco, CA, USA), pp. 897-904, July 19-23 2009.

- P. Kritharas and S. J. Watson, ”A comparison of long-term wind speed forecasting models,” *Journal of Solar Energy Engineering*, vol. 132, no. 4, p. 041008, 2010.

- P. Kritharas, C. M . Murphy, and S. J . Watson, ”Improved long term wind speed prognosis using advanced statistical methods,” in *European Wind Energy Conference and Exhibition (EWEC)* - Poster session, (Warsaw, Poland), April 20-23 2010.

- S. J. Watson and P. Kritharas, ”Long Term Wind Speed Variability in the UK,” in *European Wind Energy Association Conference (EWEA)*, (Copenhagen, Denmark), April 16-19 2012.

- S. J. Watson, P. Kritharas, and G. Hodgson, ”Wind Speed Variability across the UK between 1957 and 2009,” *Journal of Wind Energy*. Article first published online: 8 OCT 2013, DOI: 10.1002/we.1679.

Chapter 1
Introduction

THE following sections act as an introduction to the research topic of this study. By the end of this chapter several answers will have been given so that the readers will be able to appreciate the importance of this particular research. Among others, an example of the answers aimed to be provided is as follows:

- Is the use of renewable energy the answer to the damaging effects of global warming?
- What is the contribution of wind energy in the overall energy mix and in the economy?
- What are the characteristics and constraints of the British Electricity Market that increasingly necessitate our focus on wind power forecasting?
- What is the importance of wind power forecasting and current research trends?
- Why long-term wind speed/power forecasting?

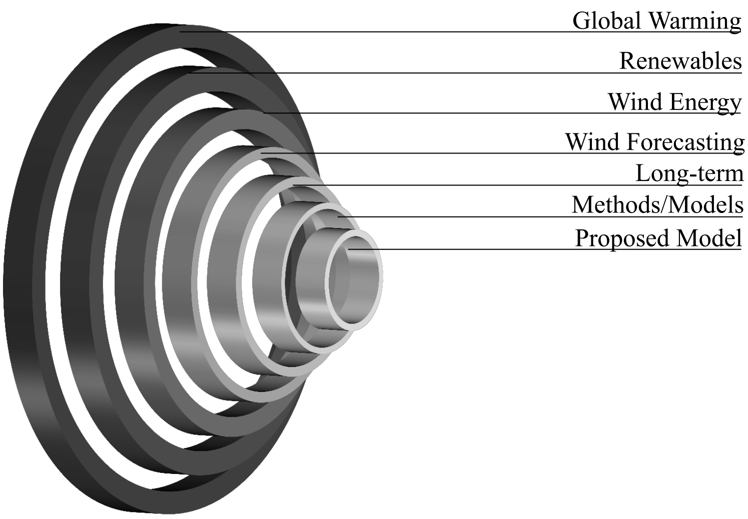


Figure 1.1: *The concept of the concentric circle approach*

This endeavour was aided by forming an approach based on identifying the fundamental research boundaries. In a deeper level, the perspective of this study is best described by descending the subject matter. To illustrate this, the concentric circles in Figure (1.1) serve this purpose. The circles represent the different layers each of which deepens in and sets the limits of the different steps taken towards the conduct of this work. The peripheral circle represents the generic research framework where the initial context of global warming is explained. As the circles get closer to the nucleus, a microscopic approach clearly formulates the stages of this work. Firstly, the research context includes the issues that the study is concerned with, the reasons of being important and original, and, how it builds on previous academic work. The problem statement and the research questions then follow progressing to the stage of the methods and approaches considered to be the optimum for addressing the research questions.

1.1 Global Warming: Facts, Effects, and Remedy

The electricity supply risk due to scarcity of conventional energy supplies [1] along with the fear that global warming will preserve its incremental trend [2] are major issues ecumenically recognised. In particular, the uprising trend in global temperature has been addressed long before terms such as global warming and climate change were yet an issue. Hansen et al. [3] concluded that a rise of about 0.2 °C in global temperature between 1960s and 1980s is associated with anthropogenic carbon dioxide (CO_2) emissions. They also predicted that under the worst case scenario, where no action was taken to cut down the CO_2 emissions, temperature would continue to rise. It is of paramount importance to note that at the time that the study was published, the cooling of Northern Hemisphere had led scientists to the common misconception that global temperature declined.

Global warming affects not only ocean rising or polar ice melting but also wind speed. According to a research published, there is an inverse correlation between temperature and wind speed [4]. In the same study, Ren states that an increase of the magnitude of 2 to 4 °C would result in a weaker atmospheric circulation over the majority of higher latitude regions, which in turn could result in a 4 to 12% decrease in wind speed. There is a need for the public and governments to realise even further this dynamic relationship and the consequences that such an increase in temperature could trigger. Evidence has recently been brought to light that there is a 90%

probability that temperatures will have been risen 3.5 to 7.4 °C by the end of this century [5,6] due to global warming's effects. Following the conclusions drawn from the aforementioned studies, one of the co-authors, Prinn made the following statement: *"there is significantly more risk than we previously estimated. This increases the urgency for significant policy action."* [7].

The awareness of these issues and their consequences have been evolved into a pragmatic and rational series of actions that need to be taken. Thus, various directives and legislations have been published [8–10]. These policies aim to tackle emissions, reduce the use of fossil fuels, introduce renewables and hence mitigate climate change. The efficacy of these policies is supported by a recent report which manifests that the countries that ratified the 1997 Kyoto Protocol managed to cut down their emissions to about 8% as compared to the levels measured in 1990 [11].

1.2 Wind Energy Statistics at a Glance

Comparing all forms of renewable energy in terms of generation capacity, wind is the one that has been growing at the fastest pace in recent years. Therefore, it has become the flagship of the aforementioned policies. To put some numbers to those facts, Europe by the end of 2012 counted 11.4% of the total energy capacity coming from wind as compared to 2.2% which was recorded in 2000 [12]. This is translated to a five-fold increase of wind power's share since 2000. Moreover, based on a report published by the Global Wind Energy Council (GWEC), China in 2010, by having installed over 44 GW of wind power capacity, surpassed the 40 GW installed in the US and became the world leader in wind generation [13]. At present, figures are staggering with China counting over 75 GW of installed wind power capacity while the US remain the runner-up with just over 60 GW installed [14]. In the same report, the figures related to future wind generation bring hope and optimism. The projections point out that in the next few years, and in particular by the end of 2017, global wind capacity will stand at about 536 GW as compared to the 197 GW installed at the end of 2010. To mention a few *domestic* numbers (since the present research is based in the UK), up to December, 2012, 10% of the electricity in Britain was generated from renewables with almost half of it being generated from wind [15].

To connect this to the projections of the future installed capacity from wind, a recent report published by the European Wind Energy Association (EWEA) is worth to be mentioned. According to this study, EWEA has developed an electricity calculator which shows that by 2020 the cost in € per MWh will have been 67, 80, 100, 57 and 74, for gas, coal, nuclear, onshore wind energy, and offshore wind energy respectively [16]. On top of that, generating electricity from wind is considerably cheaper than photovoltaics by being a more mature technology [17].

It is also very important to mention another report by EWEA where it was found that the contribution to the EU's Gross Domestic Product (GDP) of wind energy's industry increased by 33% between 2007 and 2010 [18]. This had an immediate increase in employment by 30% with more than 240,000 jobs being created in times where EU was facing a rate in unemployment of 9.6%. In the very same report, the forecasts about the economic growth that EU will be facing are even more promising. Wind industry will have generated 0.59% by 2020 and almost 1% by 2030 of the EU's GDP. These figures are translated to an increase of 520,000 and 794,079 in jobs by 2020 and 2030 respectively. On the other side of the Atlantic Ocean, a report published by the US Department of Energy claims that by investing in offshore wind energy only, 200,000 jobs will be supported in the US which in turn can be translated to a $70 billion in annual investments by 2030 [19]. Similarly, the UK's GDP will be £20 billion higher in 2030 if investment is focussed on offshore wind rather than on gas [20]. In summary, this evidence critically suggests that wind energy not only helps the environment by tackling global warming but also contributes to the economy by creating more jobs.

However, although wind energy is available almost everywhere in the planet, it relies on climatic conditions and also depends on the distinct topography of each place. Therefore, it is considered to be a fluctuating source of energy. For that reason, there are sceptics who believe that wind energy by being intermittent[1] cannot be reliable. In the next sections, the focus will be on the feasibility to predict the output generated from wind and on the various methods/models used to perform this in different time scales. Prior to this, an explanation of the British electricity market will be given in order for the readers to gain a clear understanding of how electricity (including that generated from renewables) is traded.

[1]Intermittent and Non-intermittent Generation is defined in Engineering Recommendation P2/6 as follows: Intermittent Generation: Generation plant where the energy source for the prime mover can not be made available on demand. Non-intermittent Generation: Generation plant where the energy source for the prime mover can be made available on demand. [21]

1.3 The British Electricity Market

The wholesale electricity market created for trading electricity within Great Britain, known as the British Transmission and Trading Arrangements (BETTA) [22], obliges generators/suppliers to submit the volume (in half-hourly blocks) of their expected generation/demand to the System Operator (SO) at least an hour ahead of delivery (known as the gate closure time). Therefore, in order for the SO to ensure the delivery of an economic and safe system operation, generators are forced to adhere to their contracts. Wind power has to be traded in the same way as any other form of energy under the BETTA mechanism. However, the variable nature of wind makes the participation to an electricity market such as BETTA more complex. Thus, wind generators have been using several forecasting techniques in order to avoid being penalised for any mismatch between the power contracted to be delivered and that one actually delivered. Figure (1.2) [22] shows on a logarithmic timescale the different processes in BETTA and the actions made by its participants during each stage.

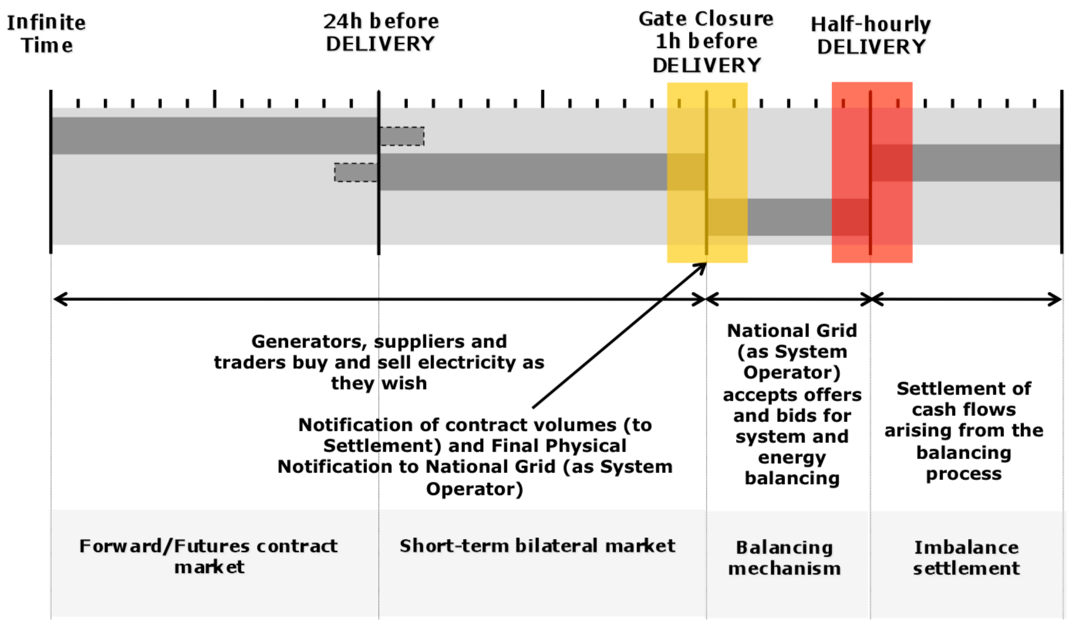


Figure 1.2: *Overview of the processes in BETTA*

In Great Britain a single trading market consists of the following components [23]:

1. The trade date, which is the date the trade is carried out;
2. The delivery date, which is the first date that the power is delivered from;

3. The delivery period, which is the length of time that the power is delivered for, which could be between 1 half-hour and an infinite time, for instance an annual contract;
4. The volume, which is the volume chosen to be sold/bought during the delivery period in MWh. At this point, it is crucial to point out that this volume is fixed and cannot be altered over the course of this specific trade, and finally;
5. The price, which is the cost of the power in £/MWh.

Obviously, generators/suppliers can carry out more than one trade to build up a profiled purchase/sale of power but each trade is constrained as mentioned above.

1.4 Wind Power Forecasting

The value of wind power forecasting has a two-fold importance. Firstly, knowledge of the expected generation output from wind power plants brings confidence to the SOs when trying to achieve reliable and secure operation of the network. Secondly, it enhances the value of wind generated electricity by providing the SOs with vital information when generators/suppliers participate in an energy market such as the market in Great Britain.

This can be achieved both by providing higher value contracts due to better bidding strategies and by minimising imbalance costs. Specifically, in a deregulated market, generators and suppliers (to whom the imbalance risk is often transferred) can avoid being penalised by choosing an optimum bidding strategy. However, this also depends on the market in question. As stated by Bathurst et al. [24], the rules can influence the selection of an appropriate bidding strategy.

So far, most of the research has focused on short-term forecasting of wind conditions. This is mainly due to the operational need for trading electricity a few days or a few hours ahead of gate closure because of the daily fluctuating nature of the demand and the finite response time of generation plants.

Findings vary depending on the selected approach (statistical or physical), the time horizon of the predictions or the area covered (single wind turbine/farm, sub-region, region or country). An up-to-date comparison and evaluation of the state-of-the-art forecasting systems can be found in a study by Martí et al. [25]. In addition, a series of studies [26–28] provide a comprehensive overview of the prediction models. However, as it will be demonstrated in the next section, SOs, generators, and suppliers have a need of longer term predictions of the power traded so that they can maximise their financial profits and schedule the maintenance of generators and power lines.

1.5 Long-Term Wind Speed/Power Forecasting

A set of motives are germane to long-term wind speed/power forecasting. Initially, long-term predictions of wind energy potential will contribute to the evaluation of the technical feasibility as well as the financial viability of wind projects for their expected operational life span. Such information can be obtained by producing annual wind speed and hence wind power estimates. A recent study reported a 5 to 15% decline of the annual wind speeds measured at 10 m above ground level (agl) across the Northern Hemisphere [29]. However, this work is regarded as being tentative since it takes into account only land surface data for a relatively small period. Nevertheless, the results showed that further work must be carried out looking at the longer timescales. In practice, annual mean wind speeds may vary at a significant rate. As a consequence, long-term variability in wind speeds can lead to misleading assessment of the wind energy yield for both operational and candidate sites. This in turn may result, depending on the specification of the turbines used, in deceptive projections of the payback period of the investment. The latter can prove to be catastrophic for the investors since large fluctuations in the annual mean wind speed may evince that such an investment is uneconomic; hence the project may be classified as non-bankable. This is because setting the terms of concessions and repayment of any loan before granting it is a common practice for bankers and external investors. In the case of wind power projects, they are interested in monthly power predictions in order to evaluate the associated risk and to determine the revenue of the project.

Another motive connected to long-term estimates is that half hourly trading, as it happens in the majority of the liberalised electricity markets (including that of Great Britain), may not be an option for small generators and suppliers due to the overhead of operating a 24-hour trading desk. Some small suppliers of green electricity, including that from wind farms, trade a month ahead and thus require an indication of the expected *windiness* of the following month. Therefore, any reference for the future state of wind conditions is critical since it would provide suppliers with vital information with regard to the optimum purchasing of base load [30].

Long-term estimates of wind speed/power would contribute to the maintenance schedule for wind farms, the whole operation of which can typically be time-consuming and expensive. Moreover, long-term wind power forecasting would allow SOs to evaluate the potential wind energy production and schedule power systems in a more effective way [31, 32]. Such a perspective would also include the potentiality for SOs to plan and manage the transmission lines.

1.6 Initial Constraints and Hypotheses

Monthly wind speed forecasts were suggested by Good Energy (who supported this research) as such information will integrate into an in-house trading system for purchasing base load months in advance. The purchase of base load has a two-fold importance. Firstly, it reassures SOs that the network will maintain its full capacity. Secondly, it prevents suppliers that generate electricity by renewable sources, such as Good Energy, from being exposed to higher risk during the imbalance settlement process. In case customers have consumed more energy than Good Energy has purchased/generated on their behalf, the company has to buy base load in order to meet the demand. However, this approach may be deemed, under special circumstances, inefficient and expensive. Thus, monthly predictions have been selected to meet the needs of the company.

Moreover, Good Energy has suggested that it may be prudent to use models based on statistical techniques. The reason for their suggestion is based on the fact that running global climate models (GCMs) or regional climate models (RCMs) is not a realistic option for the company's daily needs. All these models require high specification hardware and are time consuming. Thus, this research considered only models based on statistical techniques.

Two of the major groups of statistical models have been favoured so far in the limited literature of long-term wind speed/power forecasting. The first is the so called autoregressive models and the second group is inspired by artificial intelligence (AI). Several reasons played a catalytic role in choosing to test the autoregressive models as shown below:

- considerably less need in computational power;
- less complexity for the common user;
- access to a vast amount of data that makes it more sensible to use autoregressive models as opposed to AI models that are not dependent on large datasets.

However, as it will be shown in more detail in section 2.2, autoregressive models have not employed so far explanatory variables other than wind speed for generating long-term forecasts. This study aims to test the hypothesis that,

independent meteorological variables are highly correlated with wind speed for different time lags and when they are utilised under a single model that treats wind speed as a dependent variable then the model shows higher accuracy and performance compared to a model which does not.

Prior to presenting a model which, depending on the results, would add value to the daily operations of Good Energy, assessing the variability of wind speed from historic measurements was also deemed prudent. This target was driven solely because wind speed variations can affect the energy yield from wind farms and thus can have an impact on the risk weighed in the return of the investment. Another factor that underpins the assessment of the variability of wind speed derives from a recent survey by DNV KEMA [33] that brought to light evidence that historic wind resource contributes 18% to the total energy production uncertainty. At this point, it should be noted that in 2010 the UK experienced an unprecedented period of particularly low wind speeds which had in turn a huge impact on wind industry [34]. On the contrary, in early 1990s the UK experienced a period of high wind speeds above the usual variation in long-term mean wind speeds [35]. The aim of this line of research is to,

- *generate a reliable and cost effective wind index that accounts a considerable long period of historic onshore wind speed data, and*
- *translate the variability on wind speeds to a variation in typical wind turbine capacity factor.*

1.7 Research Objectives and Project Milestones

This project has several distinct objectives:

1. to assess the wind speed variability over the UK.
2. to compare the trends with ones from other sources.
3. to convert the changes in wind speed to changes in terms of capacity factor for a typical wind turbine.
4. to investigate the correlation between wind speed and other meteorological variables.
5. to identify a spatial association between the two variables of interest.
6. to take into account different lags in the association between the variables for the development of the model.
7. to compare the proposed model with ones that do not take into account exogenous variables.
8. to introduce an initial framework of model comparison that will contribute to the overall research of long-term wind speed forecasting in the UK.

To accomplish the above objectives the following phases were employed:

- Phase 1: A literature review on the background theory of long-term wind speed forecasting was conducted. This phase included thoroughly any relevant information about past research on the subject. This step also involved the presentation of any relevant theoretical framework. Most importantly, identifying the limitations of the current literature led to conclusions and determined the boundaries of this research. This, as a consequence, shaped the rationale of the research presented herein by highlighting its contribution to the existing body of knowledge.
- Phase 2: Data were collected from onshore historic measurements as well as from other sources such as reanalysis data. During this phase, the raw wind speed records were quality-assessed by identifying and analysing several factors that might affect the

measurements. Several criteria and flagging rules were set to deal with stations that had erroneous, missing or duplicate values. Preventing data contamination due to data entry errors as well as outlier detection were achieved by visually inspecting the time series. Afterwards, the data were stored and maintained thenceforth in a database within the premises of the Centre for Renewable Energy Systems Technology (CREST) for future continuity and other research purposes. This phase resulted in drawing remarks about the limitations, the validity, and the criteria that someone should take into account when dealing with long-term historic onshore measurements.

- Phase 3: Two representative wind indices were created for the UK in terms of both temporal and geographical coverage. This phase employed two groups of data; one consisted of 57 stations covering 29 years of measurements and another one consisted of 7 stations covering 55 years. The calculated standard deviations of these indices were used to estimate the equivalent change in the average UK wind power factor. This phase begot the estimation of the variability of wind speed in the UK and the assessment of how these changes in wind speed translate into changes in wind power. This phase also provided useful insights related to the variability of the wind climate over a period representative of the standard operational life span of a wind project.

- Phase 4: Monthly mean wind speed values were generated. This phase comprised the use of advanced statistical models that consider wind speed as an endogenous input while other variables were used as exogenous inputs. This step resulted in a proposed model that takes into account wind speed over the UK as well as meteorological variables over the Atlantic for different time lags. The high correlation coefficients between wind speed over the UK and the exogenous variables reveal the presence of a strong relationship that, when taken into account, increase the accuracy of monthly mean wind speed forecasts.

1.8 Contribution to Knowledge

The research reported herein is the first known analysis of a UK wind index using surface station observations for a period of greater than 50 years and with an analysis of regional variation. It also presents a contrast between wind indices derived from spatially smoothed datasets, e.g. reanalysis data, and point values from meteorological stations. Moreover, the model proposed for predicting monthly mean wind speeds is the first known autoregressive model that takes into account exogenous meteorological variables and autocorrelation to generate monthly mean wind speed forecasts.

1.9 Thesis Outline

Chapter 2 defines the research context by discussing the main statistical methods used in time series forecasting and further specifies the direction of this work by presenting a literature review related to the use of these models in long-term wind speed/power forecasting.

Chapter 3 provides information about the source of the data used in this study as well as the necessary actions for eliminating discrepancies and missing or duplicate values.

Chapter 4 presents two long-term wind indices for the UK based on surface stations as well as a UK annual regional wind index.

Chapter 5 presents the proposed statistical model for generating monthly mean wind speed predictions.

The main conclusions are discussed in Chapter 6. The thesis concludes with a series of suggestions for further investigation.

Chapter 2
Research Context and Literature Review

MOST of the work on wind speed/power forecasting has focused on the short-term horizon, thus the majority of the literature published also centres on the short-term predictions. However, there is a justifiable interest in predicting wind speed, and thus wind power, on a monthly, seasonal or even annual basis. This chapter is organised as follows. The first section aims to define the context of this work by presenting the prevalent statistical models used strictly in the field of wind forecasting. The second section contains a comprehensive literature review and discussion narrowing down to the statistical methods/models used in long-term wind speed/power forecasting.

2.1 Methods and Models

2.1.1 Probability Distributions

Several studies have dealt with the aspects of wind speed and power statistics. The scope of these studies was to develop appropriate techniques for evaluating the available wind resources. Estimating the availability of wind resources can lead to the evaluation of the performance of current or candidate sites. A way is to determine the probability distribution of wind speed and hence mathematically describe wind's frequency distribution. So far, different methods have been employed to fit a variety of distributions to wind speed data [36–38] with the Weibull distribution being the most dominant.

The Weibull distribution for wind speed u is expressed by a probability density function. According to Justus et al. [39], the probability density function is:

$$P(u)du = (k/c)(u/c)^{k-1} exp[-(u/c)^k]du \tag{2.1}$$

where c is the scale factor expressed in wind speed units and k is a dimensionless parameter called the shape factor. Integrating equation (2.1) between zero and a specific value of wind speed, u_x gives its cumulative probability distribution:

$$P(u \leq u_x)du = \int_0^{u_x} P(u)du = 1 - exp[-(u_x/c)^k] \tag{2.2}$$

The cumulative probability distribution is used to calculate the probability $P(u \leq u_x)$ where u is less than u_x. If the shape factor k is equal to 2 then Weibull becomes the Rayleigh distribution and c is defined as:

$$c = \frac{2\overline{u}}{\sqrt{\pi}} \tag{2.3}$$

Therefore the Rayleigh distribution is also known as a special case of the Weibull distribution. To summarise, the Weibull is a generalised distribution in which by knowing the scale and the shape factors for a given height it is feasible to adjust these parameters to another desired height [39].

A comprehensive review of the different probability distributions used in wind energy applications along with their mathematical expressions is provided by Carta et al. [40].

2.1.2 Time Series Forecasting Models

A time series is an indexed sample $y_1, y_2, \ldots, y_t$ where the indices $1, 2, \ldots, t$ represent time spaced at invariant and consecutive intervals. If the data points $y_1, y_2, \ldots, y_t$ are assigned to a random variable Y as a finite sequence of values then Y is a discrete variable. If the data points are assigned as an infinite sequence of values then Y is a continuous variable which has an associated probability distribution and a probability density function. A sequence of random variables is called stochastic or random process. A deterministic process, as opposed to stochastic, is a process whose future states over time do not involve random phenomena.

Forecasting is the prediction of an actual value at time t for a lead time $t + l$ and its error, thus its accuracy is determined by the difference between the actual value and the forecasted one. This research focuses only on statistical techniques and does not take into account methods based on models of atmospheric physics.

2.1.2.1 Regression Analysis

Regression analysis is a statistical forecasting model which determines the relationship between a dependent variable and one or more independent variables [41]. Suppose that Y is a dependent (or endogenous) variable and X is a independent (or exogenous) variable. A regression model then can be expressed as follows:

$$Y = \alpha + \beta X + \epsilon \tag{2.4}$$

where α, β are the coefficients of the dependent variable and ϵ stands for the residuals or else the so called white noise.

The model expressed in equation (2.4) aims to minimise the sum of the squared errors by fitting different values of the variables to the set of observations. The sum of the squared errors (often expressed in the language of statistics as SSE) is expressed as follows:

$$\mathrm{SSE} = \sum_{t=1}^{n} (\epsilon_t)^2 \overset{(2.4)}{=} \sum_{t=1}^{n} (y_t - (\alpha + \beta x_t))^2 \tag{2.5}$$

The units of the SSE are expressed in ms^{-1} when looking at forecasting errors in wind speed and in kW when looking at the prediction errors in wind power. Henceforth, this applies accordingly throughout this thesis unless what the SSE is measuring is not explicitly stated.

However, the technique expressed in equation (2.4), also known as regression analysis fit, has a major flaw since the choice of the fitting values is not for the original function. Instead, prior to the selection of the optimum values, the original function is linearised. This justifies the assumption that the variables are not random.

To provide an example, Connor [42] predicted wind speed and direction from equations that individually minimise the least squared errors in wind components. It was then found that the independent variables were highly correlated (intercorrelation) and therefore a forward stepwise regression method was employed. Afterwards, the model was tested against persistence[2] and climatological models. The analysis showed that the proposed model surpassed both persistence and climatological models in predicting wind speed and direction.

2.1.2.2 Box-Jenkins Methodology

This methodology, firstly presented by Box and Jenkins [43], uses an autoregressive moving average model (ARMA) for forecasting time series by fitting past data of the same series to the model. It is expedient, prior to an explanation of ARMA models, to lay out the rationale underlying the formulation of these models.

A type of a stochastic model which depends on its time-lagged forecasts of the series is named autoregressive (AR) [44]. Essentially, AR is a regressive model. The generalised form of an AR model of order p is given by Box and Jenkins [43]:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \epsilon_t \tag{2.6}$$

where c is a constant, $\varphi_1, \ldots, \varphi_p$ are the parameters (AR coefficients) of the model, $y_{t-1}, \ldots, y_{t-p}$ are the time-lagged values of the series y_t, and ϵ_t is the error term at time t with mean zero and constant variance σ_ϵ^2. The notation p in AR_p indicates the order of the autoregressive process which is expressed as a polynomial that takes into account only the previous terms of the process and the error term. Hence, the term *order* is the polynomial's degree or else the highest order power in the polynomial.

Miranda and Dunn in [45] used a probabilistic approach to develop a model which would treat wind speed time series as an autoregressive process. The proposed model was an AR of 6^{th}

[2] A persistence model assumes that the value of a variable for time t will be the same to the predicted value for time $t + l$ made at the time origin t, where l represents the forward time steps. It is the benchmark model which every other model must compete with in order for the forecasters to assess its performance.

order and when it was compared to the persistence model it showed marginal improvement in its predicting accuracy. When lower order AR models were also tested they proved to be inaccurate. However, the proposed model, despite its simplicity, served as a precursor for other more complex statistical models to follow.

Another type of time series model which regresses against the past errors of the series is also available. This model is called Moving Average (MA) and, similarly to AR, it is a type of stochastic process. Its generalised form of order q is given by:

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \tag{2.7}$$

where $\theta_1, \ldots, \theta_q$ are the parameters (MA coefficients) of the model, and $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$ are the time-lagged values of the error. Similar to the AR model, the term order q refers to the highest order power in the polynomial.

Let L operate on y_t, and ϵ_t as the *lag operator*.

$$L^k y_t = y_{t-k}, \quad \forall k \in \mathbb{Z} \tag{2.8}$$

The lag operator shifts the data and the errors either forward (when $k < 0$) or backward lags (when $k > 0$), and is defined as:

$$\begin{aligned} Ly_t &= y_{t-1} & L\epsilon_t &= \epsilon_{t-1} \\ L(Ly_t) &= L^2 y_t = y_{t-2} & L(L\epsilon_t) &= L^2 \epsilon_t = \epsilon_{t-2} \\ &\vdots & &\vdots \\ L^p y_t &= y_{t-p} & L^q \epsilon_t &= \epsilon_{t-q} \end{aligned} \tag{2.9}$$

Also, let the differencing operator ∇ be defined as:

$$\nabla y_t = y_t - y_{t-1} \tag{2.10}$$

$$= (1 - L) y_t \tag{2.11}$$

With the lagging notation, expressions (2.6) and (2.7) become:

$$y_t = c + (\varphi_1 L + \varphi_2 L^2 + \ldots + \varphi_p L^p) y_t + \epsilon_t \tag{2.12}$$

$$y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q) \epsilon_t \tag{2.13}$$

Setting the AR polynomial of L of order p as:

$$\Phi_p(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p \tag{2.14}$$

and the MA polynomial of L of order q as:

$$\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q \tag{2.15}$$

then, from expressions (2.14) and (2.15), expressions (2.12) and (2.13) become:

$$\Phi_p(L) y_t = c + \epsilon_t \tag{2.16}$$

and

$$y_t = c + \Theta_q(L)\epsilon_t \tag{2.17}$$

When these two models are coupled together they produce an ARMA model. This particular combination has been used for forecasting wind speed for short-term horizons [46–48]. In these initial studies, ARMA models were compared against persistence confirming the superiority of the latter only for periods larger than 1 h.

An ARMA model can be employed by simulating hourly averages of wind speeds for different sites, as Balouksis showed [49]. In this particular work, it became apparent that the ARMA model gives good agreement between the simulated and the measured data.

The accuracy of ARMA models for different time periods and different orders (p, q) was assessed by Milligan et al. in [50]. It was concluded that the best model was of the order of $(1, 24)$. When this model was compared to persistence it showed a 7% and 18% improvement in the first and sixth hour respectively.

An ARMA model with an order (p, q) is written:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{2.18}$$

With the lagging and polynomial notations, expression (2.18) becomes:

$$\Phi_p(L) y_t = \Theta_q(L)\epsilon_t \tag{2.19}$$

A limitation in all the time series models, mentioned so far, is that they require the time series to be stationary[3]. However, wind speed is not a stationary process and, therefore, another model has been proposed [43]. This is called autoregressive integrated moving average (ARIMA) model. ARIMA outperforms the aforementioned models since it can use non stationary time series by differencing them until stationarity is achieved [43].

Equations (2.6), (2.7) and (2.11) can be combined to give an ARIMA $(p, d, q) model$:

$$\begin{aligned} &\underbrace{(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)}_{\text{AR(p)}} \underbrace{(1 - L^d)}_{\text{I(d)}} y_t = \\ &= c + \underbrace{(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)}_{\text{MA(q)}} \epsilon_t \end{aligned} \tag{2.20}$$

The notation d refers to the order of the differencing operator ∇ which is expressed in (2.8). The order of the differencing operator d refers to the number of times that the process is transformed. Each time the differencing operator is applied the transformed series contain one point less data than the original series.

A seasonal autoregressive integrated moving average model (SARIMA), with degree p, q, P, Q, is an extension of an ARIMA model, which also takes into account seasonality s, and can be written as a SARIMA $(p, d, q) \times (P, D, Q)_s$. These models are ideal when a seasonal behaviour features the series. Hence the seasonal part may be identical with the non-seasonal one and may well include a seasonal autoregressive term, a seasonal moving average term and a seasonal differencing operator. To distinguish them, the seasonal parts are denoted with capital letters where P, D, Q are the seasonal AR, the seasonal differencing operator and the seasonal MA term respectively.

By combining equations (2.6), (2.7), (2.9), and by substituting both AR and MA polynomials of L of order p, q, P, Q we get:

[3]A stochastic process, whose attributes do not alter over time, is called a strictly stationary process. This means that the variance and the mean of the data do not alter as the process evolves over time.

$$\varphi_p(L)\Phi_P(L^s)\nabla^d\nabla_s^D y_t = \theta_q(L)\Theta_Q(L^s)\epsilon_t \tag{2.21}$$

where ∇_s^D is the seasonal differencing operator.

SARIMAX are SARIMA models with exogenous input. The exogenous input X, integrates an ordinary regression model that uses external variables into the SARIMA model. Thus, from equation (2.17) we get the generic form of a SARIMAX $(p, d, q, b) \times (P, D, Q)_s$ model:

$$\underbrace{\varphi_p(L)\Phi_P(L^s)}_{\text{AR(p, P)}}\overbrace{\nabla^d\nabla_s^D}^{\text{I(d, D)}} y_t = \overbrace{\eta_b(L)\xi_t}^{\text{exogenous input (b)}} + \underbrace{\theta_q(L)\Theta_Q(L^s)\epsilon_t}_{\text{MA(q,Q)}} \tag{2.22}$$

A SARIMAX model discloses a specific weakness which limits its use. This model assumes that the effect the independent variables have on the dependent one occurs at the current time. Thus, it fails to consider effects that take place at different time lags. However, as it will be shown later on in Chapter 5, there are cases that the dependent variable responds to the changes of the exogenous inputs several lags ahead from the time step that the independent variables were changed.

Autoregressive with exogenous variable models have been employed in the past for predicting wind power on the short-term horizon. Such examples include studies like the one published by Durán et al. [51]. In that study the authors developed an autoregressive with exogenous inputs (ARX) model which predicts wind power while it uses information about wind speed as an exogenous input. Wind speed information was retrieved from the European Centre for Medium-range Weather Forecasts (ECMWF). When the proposed model was tested against an AR and a persistence model, it was found that the ARX model outperforms the rest of the models. For a time horizon up to 24 h the exogenous input improves the forecasts on the wind speed up to 14.1% and 26.3% respectively.

Jensen et al. [52] presented the Wind Power Prediction Tool (WPPT). WPPT is also an ARX model which uses the input of Numerical Weather Prediction (NWP) models and wind speed measurements as exogenous variables. Like most statistical models, WPPT tries to determine the relationship between historical records of predicted wind speed and on-line power measurements. In order to do so, it is based on a self-calibration technique which allows the model to adapt to any changes with respect to time. Forecasts of wind speed and wind direction are produced for wind farms and sub-regions for a time horizon from 0 - 48 h up to 120 h and they are updated 2-4 times per day [53, 54]. WPPT has been operational since 1998 from the Danish transmission system operator (TSO) for predicting the power output from wind farms. The results published by Nielsen and Madsen [55] suggest that WPPT, although it cannot evaluate directly the economical value of the predictions, produces reliable forecasts that, in turn, can be used for load dispatch and day-to-day electricity trade.

2.1.2.3 Artificial Neural Networks

The architecture of artificial neural networks (ANNs) morphologically resembles the human nervous system. Processing is propagated across many units, similar to the way biological neural structures depend on the activity of many neurons.

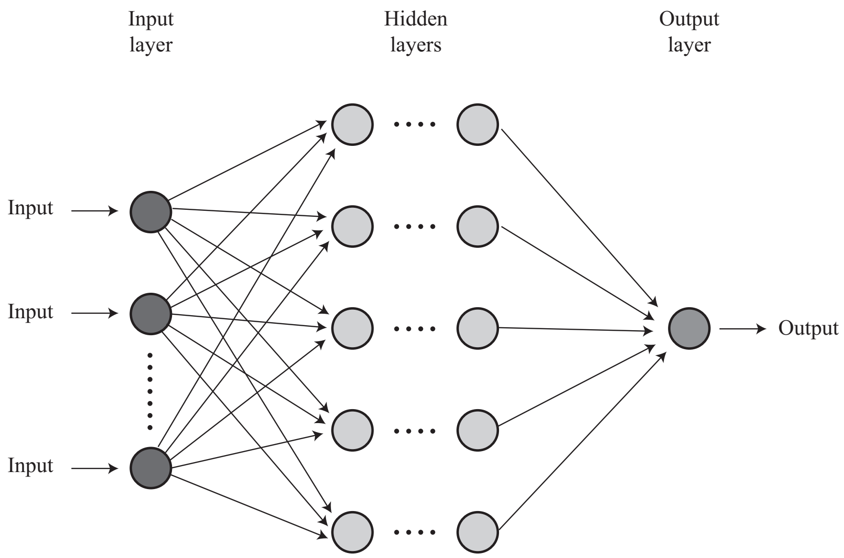


Figure 2.1: *Illustration of an artificial neural network*

Figure (2.1) illustrates an artificial neural network of multiple inputs, multiple hidden layers and one single output. The inputs of a neuron, often known as *source nodes*, are responsible for receiving the signals from either environmental stimulus or output signals from other neurons. Consequently, the inputs of a neuron are not responsible of any calculation process other than interceding the signal to the neurons in the hidden layers. The inputs x of an ANN can be expressed as a vector $\mathbf{x} \in \mathbb{R}$. Each neuron has a set of parameters that are being adjusted during the learning procedure. These parameters are the so called weights and they are associated to each input of the neuron. Similar to the inputs, the weights w that are associated to each neuron can be expressed as a vector $\mathbf{w} \in \mathbb{R}$. The neurons in the hidden layer(s) weigh up and adjust each input by multiplying the received signals with appropriate weights. The associated weighted signals are summed and the cumulative signal is being passed as an argument to an activation function. The result of the function for this argument is the output neuron for the current inputs and for the specified weights, and ends up at other neurons through the output synapses [56]. The output of each neuron can also be expressed as a vector $\mathbf{o} \in \mathbb{R}$, and it is defined as:

$$\mathbf{o} = f(\mathbf{w} \cdot \mathbf{x}) \tag{2.23}$$

There is a variety of functions used in the activation procedure. Among the most frequent functions used are the linear, the threshold, and the sigmoid functions. Below, the expression for each activation function is given [57]:

- $f(x) = x$, when the linear function is used,
- $f(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$, when the threshold function is used, and
- $f(x) = \frac{1}{1+e^{-ax}}$, when the sigmoid function is used.

There are two distinct characteristics that classify ANNs: their topology and their learning algorithm.

The topology of a network is the layout by which each neuron is connected to other neurons. Different architectures have been proposed with the major groups being the feedforward and recurrent ANNs.

- Feedforward ANNs:
  The most popular feedforward ANN is as the one illustrated in Figure (2.1) and consists

of multiple inputs, hidden layers and outputs. These networks are known as Multilayer Perceptrons (MLPs) whose transmitted signals are passed from the input(s), through the hidden layers, ending up at the output(s). As opposed to the recurrent ANNs that are described below, the reverse flow of the signal is not permitted in feedforward MLPs. The distinct difference between MLPs and single layer Perceptrons is that the neurons in MLPs can use activation functions that single layer Perceptrons cannot. This is due to the inability of Perceptrons to implement differential equations during their optimisation process [58].

- Recurrent ANNs:
  To understand the concept of recurrent networks, one would just need to visualise the connections of neurons, shown in Figure (2.1) as arrows forming loops backwards so that they become inputs to either the same neuron or to other neurons. This feed-back connectivity allows recurrent neurons to transmit their signals back and forth until a minimum error is achieved during the learning process. This feature makes recurrent ANNs ideal for use in optimisation, pattern recognition and forecasting problems [58].

Over the evolution of ANNs, several other models have been proposed such as the Hopfield and the Kohonen ones [57]. However, the scope of this section is to embody the essential characteristics and features of models that are described later on in section 2.2 and not to serve as an in-depth literature review of ANNs. Thus, detailed information about these models has been omitted.

The topological structure of ANNs as well as their training algorithm can vary depending on the application that the ANNs will be employed in. The fundamental feature of ANNs is their inherent learning ability through training. During the training of an ANN a learning algorithm is employed so that the ANN could provide users with a solution. The most popular topology used in wind forecasting is the feedforward MLPs based on back propagation (BP) training. In particular, this algorithm falls into the supervised training classification. The neuron during its training minimises the error by employing the method of steepest descent [59]. Initially, the ANN is introduced with an example that describes the problem. Then the network propagates the signal forward (*forward pass*) and calculates a potential output which essentially is a probability based on the values of the given example. The error is afterwards calculated by differencing the output of the ANN and the desired one for the given example. As a consequence, during the *backward pass* of the error, the weights of the neurons are adjusted in order to minimise the difference between the output of the ANN and the desired one. This iteration is performed for a cycle of operations until the output of the neuron converges. A comprehensive overview of neural networks, their different architectures and learning algorithms can be found in a study by Haykin [60].

Numerous studies for wind speed and power forecasting have used these models varying the architecture. Beyer et al. [61] presented feed forward ANNs based on BP and radial basis (RBF) to predict 1- and 10-minute means of wind speed. The proposed models added a 10% improvement in accuracy comparing to the Persistence model. However, when the tested networks were trained on a set of low mean wind speeds but tested on a set of high wind speeds they proved to be inferior to Persistence.

Recurrent networks have also been favoured and suggested in several studies that have investigated short-term wind speed/power forecasting, such as the one published by Kariniotakis et al. [62]. In that study advanced ANNs were utilised for predicting the power output of a wind farm for a horizon of 2 hours with a time step of 10 minutes. Two different models were presented where the first one employs wind speed and wind direction while the second model uses information only about wind speed. In the first model the inputs are treated as independent variables in order to generate wind power output forecasts. The second model generates wind speed predictions solely from wind speed measurements on site. Afterwards, these predictions are fed in another model that transforms wind speed into power by using different manufacturers' wind turbine power curves. When these two models were introduced to a real case it was found that the first model outperforms the second one and the Persistence model.

Both references constitute the first published studies in the field of short-term wind power forecasting. However, over almost 20 years of research, different architectures as well as training algorithms have also been presented. An up-to-date review on the ANNs used in wind speed prediction is available by Sheela and Deepa [63].

To summarise, in contrast to traditional statistical techniques, ANNs adapt changes as the series evolves over time. Adaptation and learning are major characteristics both attained through the training process. Therefore, ANNs have an advantage of being more robust over traditional statistical techniques especially when distributions are generated by non linear processes and are strongly non Gaussian [64].

2.1.3 Other Models Based on AI

Other models also based on machine learning have been used for time series forecasting namely support vector machines (SVMs) and Kalman Filters (KFs).

2.1.3.1 Support Vector Machines

SVMs were firstly introduced by Boser et al. [65] and since then they have been engaged in different tasks such as classification (e.g. Blanz et al. [66]) and regression problems (e.g. Drucker et al. [67]).

The goal of SVMs is to determine the optimum classifier that separates the data into different classes. The most frequent criterion for finding the optimum classifier is to determine the so-called *margin*, namely the distance between a n-dimensional hyperplane (i.e. red line) and the closest data point to each class (i.e. red and blue dots), must be the maximum from all the other margins that are created between possible hyperplanes (i.e. green lines) and the data points of each class. This is due to the inverse proportionality between the margin and the generalisation error of the classifier. Figure (2.2) illustrates the procedure in identifying the closest point of each class based on the margin criterion.

The example shown in Figure (2.2) falls into the binary classification (i.e. the blue and red dots are linearly separable). Let a training sample be of the form $\{(\mathbf{x}_1, o_1), \ldots, (\mathbf{x}_i, o_i)\}$, where $\mathbf{x}_i$ is the input pattern and o_i is the target output.

The hyperplane that separates the data into the two classes is expressed as by Haykin [68]:

$$\mathbf{w}^\intercal \mathbf{x}_i + b \leq 0, \text{ for } o_i = +1 \tag{2.24}$$

$$\mathbf{w}^\intercal \mathbf{x}_i + b < 0, \text{ for } o_i = -1 \tag{2.25}$$

where $\mathbf{w}^\intercal \mathbf{x}_i$ is the inner product of these two n-dimensional vectors on $\mathbb{R}$, equal to $\sum_{i=1}^{n} w_i x_i =$

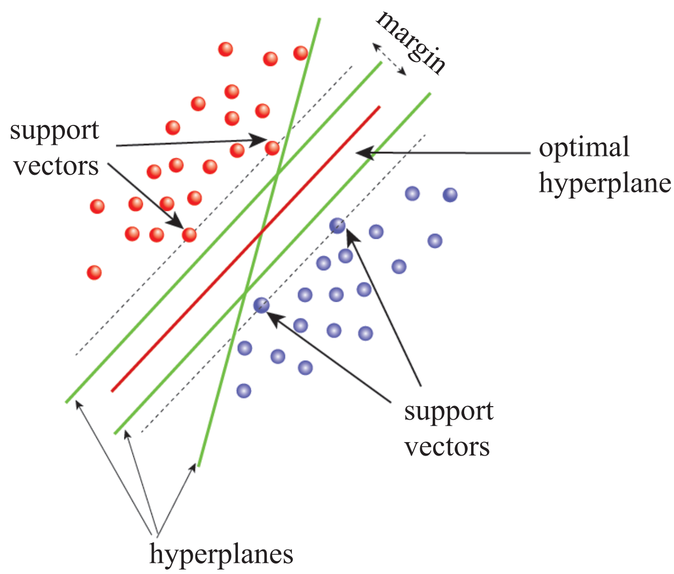


Figure 2.2: *Illustration of an SVM classifier based on the margin criterion*

$w_1x_1 + w_2x_2 \ldots w_nx_n$ and w_i and x_i are the i^{th} coordinates of $\mathbf{w}$ and $\mathbf{x}$ respectively.

Similar to ANNs, the inputs x, and weights w, are expressed as vectors while notation b, is used to represent bias. Haykin [68] by knowing *a priori* the classes of the data defined the optimal hyperplane as:

$$\mathbf{w}_o^T\mathbf{x} + b = 0 \tag{2.26}$$

He also concluded that the optimum weight $\mathbf{w}_o$ provides the maximum distance between the classes [68].

However, different problems may yield different training algorithms. For example, there are also algorithms that aim to optimise either the margin distribution or the number of support vectors [69]. Moreover, if it is found that a training set under linear separation falls into errors, then other techniques should be employed to separate the data. This can be achieved by constructing a hyperplane in a higher dimensional feature space where the separation will be easier. For this task a non linear classifier should be used by applying a kernel function. Such an example is shown in Figure (2.3).

Different types of kernel functions have been used so far such as linear, polynomial, radial basis functions, etc. One can review the different types of kernel functions as well as their pros and cons in a study presented by Gunn [70]. In the same report it is highlighted that SVMs show

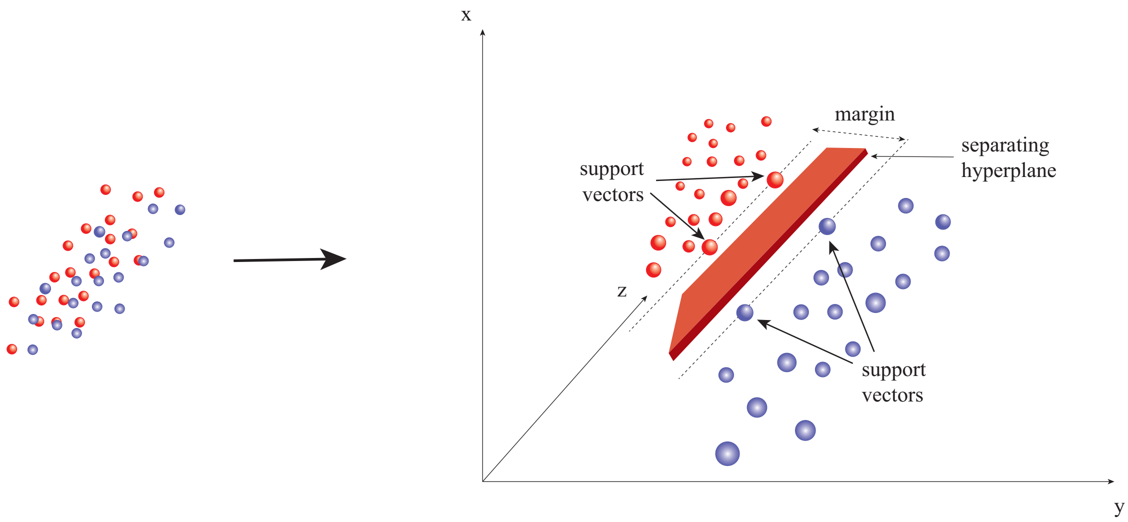


Figure 2.3: *Illustration of a higher dimensional feature space for classifying two different sets of data*

high generalisation ability over other statistical models such as ANNs. The author points out that unlike ANNs, which face massive problems with generalisation, and, in turn, can cause models to overfit the data, SVMs show resilience to overfitting. This is due to the learning algorithm chosen to optimise the parameters and statistical metrics so that the best model is selected. ANNs determine the associated risk from the error rate on the training set, also known as Empirical Risk Minimisation (ERM). On the other hand SVMs utilise another criterion function known as Structural Risk Minimisation (SRM) [71]. The latter aims to minimise an upper bound on the expected risk and, as it is shown in Ref. [72], outperforms ERM.

SVMs were introduced for the prediction of the daily mean values of wind speed and tested against an MLP neural network [73]. The results showed that SVMs outperformed MLP achieving the smallest mean square error (MSE) in wind speed predictions. The units of the MSE, in a similar way to the SSE, are expressed in ms^{-1} when looking at forecasting errors in wind speed and in kW when looking at the prediction errors in wind power. Henceforth, this applies accordingly throughout this thesis unless what the MSE is measuring is not explicitly stated. The mathematical expression of MSE is given as:

$$\mathrm{MSE} = \frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n} \tag{2.27}$$

Following the aforementioned study, an SVM regressor was applied on mean hourly wind speed data and an SVM classifier was trained to estimate the forecasting error [74]. When this approach was evaluated compared to traditional SVM regression algorithms it was highlighted that it produces the lowest MSE and mean absolute percentage error (MAPE) scores. MAPE, as its name indicates, is expressed as a percentage and is defined as:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\epsilon_t}{y_t}\right| \tag{2.28}$$

Another study compared an SVM to a BP neural network for predicting wind speed [75]. From the results it became apparent that the correlation coefficients values were very close and that both models predicted wind speed similarly. However, when these models were tested using the MSE and mean absolute error (MAE) metrics it was found that SVM surpassed the BP model. MAE measures how close the forecasts are to the real values and as such is expressed in ms^{-1} or in kW when looking at the prediction errors in wind speed or wind power respectively. MAE is mathematically expressed as:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_t - y_t| = \frac{1}{n}\sum_{t=1}^{n}|\epsilon_t| \tag{2.29}$$

2.1.3.2 Kalman Filters

The Kalman Filter (KF), firstly introduced by Kalman [76], is an algorithm which has the ability to estimate past, present and future values of measurements. To do so, KF depends on several assumptions similar to the ones taken into account in Gauss' least-squares problems [77]. For example, the residuals are considered to be independent as time series evolves over time. However, the distinct difference between these two methods is that KF demonstrates flexibility in dealing with time series whose state changes between sequent time intervals.

More formally, KF is a linear recursive algorithm which aims to estimate the state of a process without requiring explicit information about the process itself. Moreover, KFs are featured as powerful and robust since storage and reprocessing of past values are not required each time a new measurement is taken [78]. However, time series may disclose linear and (or) non linear patterns. Therefore, in that case, the non linearities of the process are evaluated by a linearised variant of the KF model named extended KF (EKF). Several studies offer a comprehensive overview of Kalman Filters [79, 80].

KFs have been employed several times since the flourishing stages of wind speed/power forecasting. Geerts [46] (although this study is mentioned previously in section 2.1.2.2), apart from an ARMA model, proposed also the use of a KF for predicting wind speeds on an hourly basis. The proposed models were then tested against persistence where the results manifested the superiority of the ARMA and KF models. This specific study also conducted a comparison between the ARMA and KF models which elected the former as the optimal model.

Another early work used KF for very short wind speed forecasting from three different sites [81]. The main finding was that KF performed better than persistence on a minute-by-minute basis in terms of root mean square error (RMSE[4]). On the other hand, persistence gave better results for hourly wind speed forecasts. It was also stressed that evaluating the optimal technique is site-dependent. This is of great importance since choosing a forecasting model does not depend on the prediction timescales solely but it is also confined to the geographical location and datasets used.

2.1.3.3 Hybrid Approach

Time series are neither pure linear nor pure non-linear. Therefore, a hybrid model which combines the strengths of the linear and non-linear methods may be the optimum technique for capturing both patterns. Hybrid models have been under scrutiny in several studies for time series forecasting with the most commonly used model being a combination of an Autoregressive Integrated Moving Average - Artificial Neural Network model (ARIMA-ANN) [82, 83]. Firstly, the ARIMA part models the linear component of the series in order to generate the forecasts. Afterwards, the ANN part is appointed to determine the non-linear relationship of the residuals and hence to improve the predictions. A similar approach has been proposed by He et al. [84] with the non-linear component consisting of an SVM. Approaches have also been considered as hybrids when they incorporate algorithms such as KFs for training or tuning other models. Examples of this particular approach can be found in the literature where, for instance, KFs are combined with ANNs [85] or SVM models [86]. Other studies have also been published in which the proposed hybrid architecture is based on coupling ANNs with SVMs [87, 88].

[4]RMSE is another statistical metric that derives by calculating the square root of MSE which is expressed in equation (2.29).

A study for predicting hourly mean wind speed is presented by Cadenas and Rivera [89]. The model used is composed of an ARIMA and an ANN model and thus falls into the class of the hybrid approach. The authors tested the proposed model for different areas and compared to pure ARIMA and ANN models. It was then shown that the model achieved lower scores in statistical measures than the rest of the models. This suggests that the hybrid methodology captures better the predictor's behaviour hence it highlights its superiority over conventional techniques.

In another study two different techniques were combined in order to forecast wind power 3 h ahead [90]. The proposed model employed a multilayer feed forward ANN to generate the forecasts. Prior to this, wind time series were reconstructed through a wavelet transformation; hence wind's signal was decomposed into different frequency components. As a result, the proposed model dominated when tested against pure ARIMA, ANN, and persistence models.

KFs have also been used in short-term wind speed forecasting as a post processor in NWP models [91]. The data produced from the NWPs were filtered from a KF algorithm in order to correct any systematic errors. From the results, the theoretical as well as computational benefits were highlighted since both the systematic errors and the time needed for the computations were reduced.

Similarly, the output of a mesoscale forecasting system was processed both in MLP and SVM algorithms in order to produce mean hourly wind speed forecasts [92]. Salcedo-Sanz et al. [92] followed the metric of MAE to evaluate the performance of the models and showed that the MLP algorithm performs slightly better than the SVM.

2.2 Statistical Models in Long-Term Wind Speed/Power Forecasting: A Literature Review

Over the last two decades, a small body of research has been carried out in the field of long-term wind speed/power forecasting. The term *long-term* refers to monthly means and even longer predictions rather than to forecasts for a time horizon, for example, up to 72 h ahead, as is described in other studies such as the one by Barbounis et al. [93].

The existing body of research centres on generating monthly wind speed forecasts being conducive to the evaluation of technical feasibility of current or future wind projects. Looking closely at the similarities between the studies, evidently the majority of them have used onshore data while, due to lack of data, they have drawn on short datasets. In parallel, there is a prominent lack in evaluating the uncertainty - apart from some exceptions - while most of the studies presented below have avoided examining whether the forecasts are biased.

On the other hand, there is a degree of diversity across the literature in many respects. This manifests in the use of different statistical metrics (MSE, MAPE, and RMSE). While these metrics are primarily used to measure/assess the error, still there is not a converging evidence base to aid the decision making on which statistical metric is best to use. Using MAE for instance, is the best choice for assessing the average performance of the model of interest. Yet the metrics' role rests on both the researchers' approach and their work objectives, namely what they want to test. In other words, in reality, these measures are adjusted to serve the research objectives rather than constitute a constant and solid base for explaining the models' properties and gaining more answers about them. At this point, it is worth noting that this diversity seems to be opposed to the short-term wind speed forecasting work. For example, throughout the ANEMOS project [94] (a project that looked at short-term wind speed/power forecasting), there had been an "unsaid" tendency, derived actually from a *status quo*, to utilise RMSE and the coefficient of determination, R^2, which measures how well the regression line fits the data.

What is also prominent in the literature is the source of data differing in terms of location, topography, and local microclimate. This, in combination with the different methodologies employed in the studies, contributed in making it infeasible to decide, by comparison of the different studies, which is the optimum model. This diversity in methods does not facilitate the statistical homogeneity and the interpretation of the results as an integral part of the picture. In terms of the character and the role of the data used overall in these studies, the current research's perspective dictates to ascertain, and highlight as problematic, that different data would have yielded different results, further affecting both the interpretation and the conclusions with regard to the models.

A tentative study, and similar in a way to the present research regarding its scope, has been recently presented by Lynch et al. [95]. The authors have been planning to employ reanalysis data for use as a climatology and verification dataset for monthly/seasonal forecasts. It is obvious that if the skill of the predictions was found to be high then it would add economic value in the monthly/seasonal forecasts. According to the initial results, both temporal and spatial variations (at scales from 300/400 km) at longer timescales were well represented. In addition, when data were extrapolated to the standard hub height of turbines (~100m), minimum bias and low RMSE was found. The aforementioned research is a follow up to the study presented by Brayshaw et al. [32]. In that work the authors investigated the existence of a possible link between wind speeds and large-scale climate patterns such as the North Atlantic Oscillation (NAO). Time series of wind speed were associated with different classes of NAO. Afterwards, hourly wind speeds were synthesised for each class by using a Markov Chain model. It was then noticed that the model tends to underestimate wind speed beyond the 24-hour threshold. The resulted synthesised wind speed series were then converted to hourly power outputs and finally to monthly power outputs. A comparison between the models that take into account NAO with the ones that do not, shows that the first model is the best by achieving the lowest scores in the statistic metrics.

García-Bustamante et al. [96] compared three different methods in generating monthly wind power production estimates. Four years of hourly data of wind speed, direction as well as wind power were employed for five different sites in Spain. The first method employed hourly records of wind speed and fit them to a Weibull distribution. Then a transfer function determined the relationship between wind speed and wind power. The second method assumed that a theoretical power curve is valid at monthly timescales and used monthly wind speed data to estimate the power on a monthly timescale. For doing so, the authors interpolated wind speed information to a theoretical power curve. However, it is noteworthy that using the theoretical power curve leads to underestimation of the generated power at lower wind speeds and to overestimation at higher wind speeds. Finally, the third method that the authors compared in their study used an average power curve which was calculated by averaging all the effective power curves over the period of the 4 years of the measurements. Then, by employing a simple linear regression method, which used monthly wind speeds, they generated wind power on a monthly basis. The comparison between the aforementioned methods established the linear regression model, despite its simplicity, for use in estimating wind energy at monthly timescales since its relative error is smaller by 15%, as opposed to the other approaches.

SARIMA models have been employed for monthly electricity production from wind farms, as it is shown by Cadenas and Rivera [97]. The authors used six years of wind speed measurements for training the model and one year of data for validation purposes. The information was retrieved from SCADA systems and it was generated by accumulating the data from the sensors. The proposed model was of the order of $(0,1,1)(0,1,1)_{12}$ to obtain both the seasonal and non-seasonal features of the time series. As it was mentioned in section 2.1.2.2, a SARIMA model can be written as $(p,d,q) \times (P,D,Q)_s$ where p,P are the non-seasonal and seasonal autoregressive terms, q,Q are the non-seasonal and seasonal moving average terms, d,D are the non-seasonal and seasonal differencing, and the notation s represents the number of periods within a season (e.g. $s = 12$ periods, 4 periods, and 1 period for monthly, quarterly, and annually data within a year respectively). The term order refers to the polynomials' degree or else the highest order power in the polynomials as it is explained in section 2.1.2.2. To determine the accuracy of the proposed model the authors compared it with an ANN similar to a simple Perceptron. The difference of the tested ANN in this study is that the activation function is linear and hence the values of its output can be any value as compared to the Perceptron which can have a binary number as output. From the comparison it was shown that the lowest statistical errors were achieved in the case where the SARIMA model was used. For brevity reasons, we refer only to the values of MSE which were 2.70 and 5.10 in ms^{-1} for the SARIMA and ANN models respectively.

AR as well as ARMA models were tested by Kennedy and Rogers [98] for monthly predictions of wind speed. The authors used two time series of monthly wind speeds from a buoy station in the US for a 16- and 5.5-year period. The authors concluded in their study that although wind speed may be a stochastic process, it is not purely random with time series exhibiting seasonal components. After removing the seasonal component from the series, they normalised the variance of the residuals and they used a first order AR to synthesise wind speed time series. Testing different orders of ARMA model did not produce better results, therefore the authors concluded that an AR(1) is sufficient for generating the new series. The synthesised series were then transformed to power by applying a wind turbine's power curve. The authors applied a spatial smoothing procedure to extrapolate wind power over a wide geographical area. Finally, the proposed model was used to generate a distribution function for the output of a hypothetical offshore wind farm. However, it is noteworthy that the smoothing procedure prevents the combined power output from being either zero or reaching the full power nominal output of the hypothetical site. In reality, this is not correct since there are cases where the wind across an entire region is either below the cut-in wind speed and above the cut-out wind speed.

Kritharas and Watson [99] used 52 years of wind speed records for seven different geographical locations in the UK. A statistical analysis of the data resulted in subdividing the sites into different classes based on the seasonal component. The classes created were sites where seasonality was weak, average and very strong. Afterwards, four parsimonious models were employed for predicting wind speed on a monthly basis. The results indicated that the model which takes into account monthly seasonality and autocorrelation of lag one, gives the best performance in terms of MSE. The same authors continued their work and proposed an autoregressive moving average with exogenous inputs (ARMAX) model [100] for predicting wind speeds for a time horizon up to one month ahead. After the data analysis, the order of the ARIMA model was determined and the forecasting model was tested against persistence and the conventional Holt-Winters' method. Further, they compared the forecasts of wind speed with observed weather data and statistically analysed the errors using analysis of variance (ANOVA). Following the ANOVA, it was showed that the meteorological parameters which affect wind speed the most is atmospheric pressure and air temperature. Thus, both parameters were used as exogenous parameters to the proposed model. The results testified that the proposed model which uses air temperature as an exogenous input gives better results, in terms of MSE, than the one which uses atmospheric pressure. Comparing the accuracy of the best models of each study between each other, it is shown that the ARMAX model proposed by Kritharas et al. [100] achieves a reduction in terms of MSE equal to 11.5% when compared to the simple seasonal model presented by Kritharas and Watson [99].

The two aforementioned studies do not undermine the purpose of the research presented herein as Kritharas et al. [100] used meteorological variables that were retrieved from data at the location of the sites under investigation. These independent variables were then fed to the model for the same time index as the dependent variable. Nonetheless, the hypothesis tested in the present study is that there is a high correlation between the independent variables and wind speed at different time lags. Moreover, the hypothesis aims to test that both the dependent and independent variables are geographically dispersed.

The majority of the studies mentioned so far compared their models against ANNs. These comparisons show that the proposed method (linear models) surpasses the competitive one (non-linear models) by achieving the lowest scores in the statistic metrics. However, a body of literature doubts the efficacy of linear models when compared to non-linear ones such as ANNs.

Different architectures of ANNs have been used for monthly wind forecasting. Examples of such architectures include feed forward as well as recurrent networks based on BP and cascade correlation algorithms [101]. The authors used daily averaged wind speed records for 12 years. The first 10 years were used as a training set whereas the last 2 years have been used to validate the accuracy of the models. Any gaps in the time series were filled by interpolating between neighbouring values. Afterwards, the time series were averaged to obtain the monthly mean values. The authors developed several topologies of ANNs for predicting wind speed on a monthly basis. However, the optimum model is a recurrent one based on Jordan's training algorithm which achieved a marginal accuracy in monthly wind speed predictions of 0.2% and 0.4% when compared to other topologies. It is also noteworthy that when the proposed model was tested against an ARIMA one, the ANN showed a decrease of 1.6% in the mean percentage error.

Gao et al. [102] proposed an MLP feedforward BP ANN for forecasting wind speed on a monthly basis. Data of monthly wind speed for 7 years has been used to train the ANN while 2 years have been used to test the accuracy of the model. The proposed model used information related to mean air pressure and mean relative humidity. The model was then tested against Persistence. The comparison showed that the proposed method outperforms the naive predictor by achieving the lowest statistic metrics.

From the overall work on feedforward ANNs one could distinguish two major research groups: one that used feedforward ANNs based on BP learning algorithms [103–106] and another one that proposed feedforward ANNs based on resilient propagation algorithm [107].

Fadare [103] used monthly mean wind speed data from 28 meteorological stations for a period of 20 years. In that study, data from 18 stations were used for training the model and 10 stations for testing it. The proposed ANN consists of 4 inputs, 2 hidden layers and one output. The author used information about latitude, longitude, altitude and month of the year as inputs of the proposed model while wind speed was the output of the ANN. The results indicate that the proposed topology shows high accuracy in forecasting the monthly mean wind speed by achieving a score in terms of MAPE of 8.9%. Most importantly, the correlation coefficient between the predictand and the measured wind speed is 0.983 which testifies how well this model performs.

Kalogirou et al. [104] presented two network architectures of MLP ANN for predicting the monthly mean wind speed in Cyprus. Wind speed data were recorded at 2 m and 7 m agl from three meteorological stations over an 11-year period. The data used as the input of the network were the month of the year and the mean monthly values of wind speed at the two different heights. The output of the network was the mean monthly values of wind speed of the third station. The authors also tested another architecture which employs the easting and northing coordinates of each station. However, the best results were generated when the 5-input ANN was used.

Campbell and Adamson [105] used data from a wind farm in Ireland over a period of 5 years. For predicting the monthly mean wind speed, a 4-layer MLP was developed. Hourly wind speed data along with the time/day of the records was averaged to obtain the monthly values. Since the wind farm consisted of 25 turbines the proposed network used as the input data the monthly mean wind speed values for two turbines along with the timestamp of the measurements to predict the monthly wind speed at another turbine. Similarly to the previous studies, the proposed model used 4- year data for training and one year for testing it. The results in the statistic metric indicate that this model performs extremely well having an RMSE of 0.11078 in ms^{-1}.

A feedforward MLP network based on the resilient propagation algorithm was proposed by Bilgili et al. [107] for generating monthly mean wind speed at a target station using the monthly mean wind speeds from neighbouring stations similarly to the previously mentioned studies. The distinct difference between BP and the resilient one is in the way the weights are adjusted. As it is shown in section 2.1.2.3, every time an input value is introduced to the ANN the weights are adjusted to minimise the error. On the other hand, in resilient BP prior to the adjustment of the weights the complete training set must be introduced to the network. Hourly wind speed data, recorded at 8 meteorological stations were used in this study. Similarly to previous work, the monthly mean wind speed data was calculated from the hourly records. The proposed network consisted of an input layer which used as source nodes the monthly mean wind speeds of the reference stations and the timestamp of each measurement. However, the architecture of the model (i.e. the number of neurons in hidden layers) was not the same for all the stations assessed in this study. The performance of the model for all stations varied with MAPE values spanning from 4.49% to 14.13%. The reason for this *diversity* in the results was due to the fact that, overall, the correlation coefficients between the measuring stations was significantly low. From the 56 pairs of correlation coefficients between the stations, only 11 pairs recorded correlation coefficients larger than 0.5. However, this study concluded that it may be prudent to employ ANNs for forecasting monthly mean wind speed as this method is robust and pretty much straightforward.

Similarly, an MLP ANN based on a Bayesian regularisation training algorithm was employed for generating annual wind speed predictions as presented by López et al. [108]. The main difference compared with the above studies was the learning procedure of the network. The training algorithm chosen to train the proposed model falls into the Bayesian regularisation class. The advantage of this specific algorithm over the ones mentioned previously is its resilience in over fitting the model and thus, while it produces small errors during the training phase, it also attains good agreement when new data are fed to the inputs of the network. As a consequence, employing the aforementioned training algorithm delivers a network that is relatively small and hence has fewer patterns to be required for the model to be trained. The authors tested several architectures that affected the number of the inputs for the network in order to determine which one performs the best. They also experimented around the type of information that would have been the input of the model. After several trials, they concluded that the optimum approach was to use wind speed data for five days of each month at the target station for a period of one year and wind direction at nearby locations. They then identified that the best neighbouring stations were those who showed high correlation with the target station. For that reason, they trained the ANN with inputs having been the data from the nearby stations while the data from the target station served as output to the model. The results of this research highlight the importance of using the direction as, in that case, the network revealed a decrease in terms of RMSE of up to 23%.

Bilgili and Sahin [109] developed different topologies of ANNs and used different training algorithms to predict the monthly wind speed at four different sites. For doing so, the authors used as inputs of the proposed model the information regarding the wind speed at any three reference stations and as an output of the network the wind speed at any target station. The period of the measured data covered a period of five years, though the authors highlighted that there were several gaps in the time series due to missing data. In a similar approach to other studies mentioned above, in this study the authors averaged the raw hourly data to obtain monthly mean values. After averaging the hourly values, the available data covered a period of 47 months. Therefore, the authors used 35 months for training the model and 12 months for testing it. Findings indicate a good agreement between predicted and measured data with their correlation coefficients varying between 0.89 to 0.98 depending on the training algorithm. Most importantly, all models developed in that paper did not use any other topographical detail or additional meteorological parameters except the wind speed series for the tested sites. Finally, the score in terms of MAPE undulated between 3.22% and 7.57%. Furthermore, in the same study the authors also generated daily and weekly wind speed predictions with the results for the proposed model/architecture favouring the monthly timescale.

Mohandes et al. [110], developed an ANN model and compared it with an AR model. The network used in this study is an MLP BP with 6 hidden units. The observed data were retrieved from one location in Saudi Arabia and covered a period of 12 years. From the total 144 months of monthly data, 120 months were used for training the ANN, while the last 12 months were used for evaluating its performance. Surprisingly, the authors used a single input and single output for the network. This could be down to the fact that the aforementioned study is one of the earliest ones that employed ANNs for long-term wind speed forecasting. Nonetheless, the results indicate the superiority of the ANN model compared with an AR(1) for predicting wind speed on a monthly basis. The statistical metrics testified the previous statement since the RMSE was found to be 1.87 and 2.88 in ms^{-1} for the ANN and AR model, respectively.

Similar conclusions derived from another study where linear regression, non-linear regression, and ANN models were used for predicting monthly wind speed at three different sites [111]. The authors used several monthly meteorological data such as wind speed, atmospheric pressure, atmospheric temperature, relative humidity and rainfall. The unique difference between this study and the others mentioned previously is that, in this case, the authors employed AI to predict wind speed by using the independent variables just mentioned. The proposed model consisted of a three layer feedforward network based on a BP algorithm. Examining the correlation coefficient between the independent variables and the dependent one, the authors have drawn similar conclusions with Kritharas et al. [100]. From the 5 years of data, 4 years have been used to train the model. When the remaining year was used for evaluating the model's performance it was found that MAPE values ranged from 7.92% to 10.48% in the case of the ANN model. When the other models were also tested against the proposed one it was found that MAPE values were higher ranging between 9.07% and 16.62%. From the direct comparison of the errors each model produced, it was concluded that ANNs outperformed linear and non-linear regression methods.

Guo et al. [112] decomposed time series from three years of wind speed records using an empirical mode decomposition (EMD) method, and then integrated the series to ANNs for generating monthly predictions of wind speed. The basic idea in EMD is to transform series to be stationary by shifting the non-linear and non stationary parts of the series. Monthly mean wind speed data were retrieved for 3 years in China. Afterwards, the time series were decomposed using the EMD technique producing a finite number of distinct sub-series. Using the EMD method the authors managed to convert these signals into stationary and to feed them as input nodes to the proposed ANN. The architecture of the ANNs presented in that study included feedforward and EMD-feedforward networks. The results from the statistical errors showed that, depending on the horizon of the predictions, each method surpasses the other. For instance, when a multi step forecasting was set up, the EMD-feedforward ANN performs better than the single model. However, when one step was employed the results pointed that the EMD-feedforward model performed poorer than the simple model. In addition, the authors tweaked slightly the EMD-feedforward model by calculating the partial autocorrelation of each decomposed signal before feeding it to the ANN. With this method, they allowed the previous value for lag = 1 to be used as input for the network. Finally, it was shown that the modified model was better compared to others. MSE, MAE and MAPE were reduced by 0.0173 in ms^{-1}, 0.1497 in ms^{-1} and 12.55% respectively compared to the feedforward ANN.

Various studies have also employed artificial intelligence other than ANNs, such as the one presented by Chengwei et al. [113], where grey theory and, in particular, the *Grey Model First Order One Variable* model, GM(1,1), were used in order to generate annual wind power forecasts. The fundamental feature of Grey models is their inherent ability to rely on less historical data than other statistical models do. However, the authors highlighted that Grey models are more appropriate for dealing with problems where the time series are growing exponentially. To achieve this, they multiplied the original wind power generation data with geometric series. Initially, daily averages of wind speed over a period of 56 years were collected for a whole region in China. Afterwards, the authors by using the power curve from a commercial wind turbine they converted wind speed to wind power. Multiplying the resulted annual power data with a series with a constant ratio between successive and consecutive intervals they managed the series to increase monotonously. The proposed model used 6 years of annual wind power data to forecast the next year's value. Then, it adopted the new calculated value and, by forgetting the first one, just so in every step using 6 years of data, it predicted the next year's annual wind power. The normalised MAE, which was the result of MAE over the total installed capacity of a farm for example, was found to be improved by 0.7679% than the case where the time series were non monotonous. However, as the authors implied, selecting the appropriate ratio needs further investigation in order to obtain higher accuracy in the forecasts.

Similarly, Gao et al. [114], developed a metabolic GM(1,1) model capable of predicting the annual wind generation in a wind farm for a period of 56 years. The proposed model took into account 8 years of data to forecast the next year's value. As the model moved through time, the oldest year from the series became less important and hence it was omitted. Therefore, the model adopted a new year so that the training period remained constant in order to forecast what would happen the year after. The authors used the same data as in the work presented by Chengwei et al. [113], with the studies being slightly different. For instance, the dataset used to set up the model covered a period of 7 years than 6 years that were used previously. The other difference is that, instead of multiplying the series with a constant ratio, the authors applied a differential model in order to get the regular historical data series with an exponential growth. When the normalised MAPE was calculated it was found to be equal to 7.8806% which shows that the model performs very well. However, when the generated energy for the year 2009 was calculated, then the relative error between the predicted and the calculated value was found to be quite high equal to 6.833%. The final result does not provide confidence in using this method until further and in-depth investigation is taken forward. The last statement leads to the conclusion that grey theory, although it is based on model's uncertainty and relies less on sufficient and complete information, cannot produce accurate results for long-term forecasting. In time grey theory would play a key role to wind speed/power forecasting but, for the time being, this approach is considered to be tentative.

To conclude, it seems that each study has made a standalone contribution to our understanding of long-term wind speed forecasting. However, chiefly due to data and methodology diversity, the picture is far from complete. This conclusion is further underpinned by the extracting remark grounded in the available research that no model appears to dominate. Therefore, it is not deemed wise to draw any conclusions about the dominance of a specific model as time series vary depending indissolubly on each location. Moreover, the conflicting findings supporting the linear versus the non-linear models add more to this judgment.

This controversy has recently driven researchers towards the use of hybrid models. Shi et al. [115] and Zhang et al. [116] provide a comprehensive overview of hybrid models for short-term wind speed/power forecasting.

Nevertheless, only a handful of studies has been published so far employing such models for long-term wind speed/power forecasting and, as a result, no robust conclusions can be drawn. Specifically, to the best of our knowledge[5], only one study has used hybrid models in long-term wind speed forecasting [117]. In this study, Guo et al. developed a hybrid model which consisted of a SARIMA part and a least square SVM part (LSSVM) in order to predict monthly mean wind speeds. Monthly wind speed data were retrieved from two locations in China over a period of 6 years 5 of which were used for training the model and the last year was used for evaluating its performance. Initially, the SARIMA model transformed the data, as it is described in section 2.1.2.2, in order for the series to become stationary. Then, by using the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF), the authors identified the model's parameters and generated monthly predictions. In a similar way, as the one described in section 2.1.3.3, the next step was to correct the residual of the forecasts by using an LSSVM model. The results showed that the proposed model managed to produce reliable forecasts. SARIMA-LSSVM model proved to be superior than the conventional forecasting models tested in the same paper by obtaining the lowest statistical errors. The most striking conclusion was that the MSE for the LSSVM was found to be equal to 0.12 in ms^{-1} and 0.11 in ms^{-1} for the two sites tested in the study. Equally, the SARIMA model resulted in an error of 0.36 and 0.21 in ms^{-1} correspondingly. The latter equated to a decrease in terms of MSE of 200% and 90.91% respectively. Another advantage of this method was that it did not require a large dataset to correct the residuals, and hence to improve the predictions.

2.3 Chapter Summary

The research context thoroughly discussed in the first section of this chapter was the inspiration which defined this work. The statistical models employed in the research of wind forecasting have proved to be advantageous for capturing the characteristics of wind speed. The major drawback though is that there is not converging evidence on which is the optimal method. This is due to the fact that these models depend on the time series and on the timescale of the

[5]This is valid up to the time this study is submitted and the statement refers to papers that have been published only in English.

forecasts. Consequently, although they perform satisfactorily in each individual study, providing better understanding of wind time series forecasting, the picture is yet far from complete. Similar conclusions are drawn from the use of the same models specifically in the long-term topic.

Chapter 3
Data Collection and Analysis

This chapter contains identical parts from three papers [99] (published journal) [118] (published conference) [119] (published journal)

The main thrust of this chapter is to provide information about the data used in this study. The needs of the research called for drawing on data from various sources. Such sources consisted of historic long-term onshore measurements from several Met stations of the UK Met Office across the country as well as reanalysis data. The succeeding procedure of setting criteria for including/excluding particular stations is amply discussed, while details about factors that possibly affect wind speed measurements are provided. Information then is presented on designing the structure of a database which was a critical step for the conduct of the research. Overall, this chapter is organised in sections based on the chronological phases carried through for the purposes of collection, filtering, analysis, and organisation of data.

3.1 Sources of Data

3.1.1 Surface Observations from the UK MIDAS via the BADC

The UK Met Office produces a data set of land surface observations from 1853 to the present date which is freely available for research purposes. This argument as it stands alone was sufficient for engaging with the Met Office in order to get hold of the data. The vast amount of historic onshore records in combination with the well spread coverage over the UK constituted the drive for using primarily that source. Another reason that limited the present research to the use of data provided by the Met Office was a combination of the hypothesis aimed to be tested with the particular family of statistical models which was chosen. As mentioned in the literature review, the autoregressive models require more data in order to capture the characteristics of time series. At the same time getting hold of data for a concurrent period that derive from tall masts (e.g. above 60 m agl) was not feasible as there are a few meteorological masts in the UK that combine the desired height (i.e. hub height) and the long period of measurements that was necessary in the study. An alternative solution would be to engage developers that have a large fleet of 10 m masts across the country but the drawback under this scenario would be that these masts would not record data for the same period. Also, by engaging developers or owners of wind farms, delays would be added in the progress of the research. This would be inevitable as non disclosure agreements would need to be signed off before any data were sent for analysis. For these reasons it was decided that a flexible, fast and reliable solution that was not compromised in terms of both spatial and temporal coverage would be the data produced by the UK Met Office. This data set is held on the Met Office Integrated Data Archive System (MIDAS) and is available via the British Atmospheric Data Centre (BADC) [120]. The MIDAS data set contains a large number of observations covering a variety of meteorological parameters including mean wind speed [121]. Wind observations are typically 10-minute means in knots (kt) and are typically made at a height of 10 m above the land surface. However, this has not necessarily been the case historically as will be discussed later in this chapter. Since this research has two different objectives the data used for each case differs slightly.

- The part that deals with the variability of wind speed and will be presented in Chapter 4 uses two different groups of data. Although both are retrieved from MIDAS, their distinct difference lies in the number of the stations contributing and the period of the series. One group used records for the period 1983 - 2011 and the second for the period 1957 - 2011 (see Chapter 4 for details). In order for the readers to be able to identify which dataset or which stations each time this study is referring to, these will be known henceforth as BADC-57 and BADC-7 respectively.
- Similarly to above, the part that deals with the long-term forecasting, and will be presented in Chapter 5, uses data from surface stations. In order to accomplish homogeneity between the different stations in terms of time, the year 1957 was set as the starting year for all stations used to generate the long-term forecasts. The specific year was set because, after applying several criteria for the stations that would be included in

the analysis (see section 3.2 below), the earliest common year for all the stations was found to be 1957. Hence, this part uses also the BADC-7 dataset.

3.1.2 ERA-40 Reanalysis Dataset

The hypothesis set to be tested in this study aimed to investigate whether other atmospheric variables except wind, can increase the efficacy in the long-term forecasts. Reanalysis data comprise a useful source that is rich in both spatial and temporal coverage. Most importantly, reanalysis datasets incorporate a vast amount of observations that, through an assimilation system, offer useful insights related to climate. Similarly to the onshore stations, the present study required data that were dated back to the starting date of the onshore measurements. One of the reanalysis data that covered a period longer than 40 years is the ERA-40 which is also freely available for someone to download. Thus, it was decided for the purpose of the study to use the ERA-40 dataset. The ERA-40 reanalysis was produced by assimilating a large number of different meteorological datasets, including satellite measurements, ship-borne and buoy observations, land-based surface observations, upper air measurements and remote sensing observations [122]. The assimilated data have been output onto both a $2.5^\circ \times 2.5^\circ$ grid and a $1^\circ \times 1^\circ$ grid at six hourly intervals covering the period 1957-2002. It should be noted that land-based surface wind speed observations were not used as input to the assimilation, though the assimilating model produces output surface wind speeds on the regular array of grid points, including those over land. In this work, 10 m values of the u' and v' wind components from the $1^\circ \times 1^\circ$ grid were extracted. u' and v' are the zonal and meridional wind components. For vector fields, such as wind velocity, the zonal component refers to eastwards wind while the meridional component refers to northwards wind. Then, the magnitude of wind speed, u, was calculated using Pythagoras theorem:

$$u = \sqrt{(u')^2 + (v')^2} \tag{3.1}$$

3.2 Criteria for Station Selection

There are circa 50,000 UK sites which report data to the Met Office; these are organised into a number of categories according to the type of data message produced. The most appropriate stations to meet the research objective are in the synoptic network. These stations have an average spacing of less than 50 km. Some of these stations are part of the global synoptic network and data are exchanged internationally in near real-time. It is thus expected that the synoptic network will have the most complete observational record and that the majority of these stations will continue to provide observations in the future [123]. This is because the availability of a complete record of observations is crucial for observing and determining the variability in wind. Future continuity is important as any information related to the wind climate should be continually updated with new data. However, the observations from the synoptic stations within the MIDAS database did not alone meet these criteria and so the synoptic data were augmented with observations from selected stations within the Met Office climatological station network.

Prior to the selection of the stations, a quality assessment procedure was applied to the data to avoid discrepancies and erroneous, missing or duplicated values. The MIDAS data contain a number of quality flags relating to the recorded values. For the mean wind speed, a value of zero in the associated quality flag indicates an unreliable observation. In order to exclude any unreliable records, a condition was set that observations would be included only if they met the non-zero quality flag criterion. However, it was then observed that there are several occasions where unique timestamps have double rows of data. Frequent communication with the MIDAS team lead to the conclusion that the correct rows are those with a value of ”6” in the associated quality flag [124]. Nonetheless, BADC archives ended up (and still remain) with duplicate values as those who maintain the database tend not to overwrite existing entries (as the Met Office does). As a consequence, this resulted in the BADC database suffering from double records each time the database was being updated in the Met Office. The problem was tackled by storing the data on a database as it is discussed in section 3.4. To do so, proper queries were performed that excluded records with different values of quality flags.

As a consequence of this criterion, there were significant gaps in the data for some stations. One option was to fill in the gaps using several statistical techniques. An example of such techniques is to replace any missing records with the median or the average value of their previous and next values. However, one of the goals of this study was to use as much raw information as possible without any intervention and without having to artificially synthesise the time series. Therefore, a supplementary criterion for selecting the stations was the completeness of the data recorded and stored. The requirement set was for the total available recorded hours to be ≥75% of the total theoretical hours (i.e. ~481,500 hours for the 55 years). The availability figure of 75% was chosen so that there were enough data to accurately represent the wind climate at each individual station but not to exclude too many stations from the study. Similarly, the available recorded hours per annum had to be ≥75% of total number of hours in the year.

A further criterion was that any station that met the previous requirements had to have no more than 15 days per month (i.e. half of each month) of missing consecutive records. If so, the station was excluded. This was necessary as long periods of missing data would skew the results by missing some of the seasonal variation. In fact, once this initial criterion had been set, the actual average availability of data for the stations filtered varied between 96.2% and 98.3%. When all the criteria were applied, from over 50,000 stations, varying in type and location, BADC-57 and BADC-7 stations remained to be included in the analyses. The reason for the much lower number of stations was simply that fewer stations met the required criteria over these much longer periods. Each station was assigned to one of six UK regions, as Figure (3.1) illustrates. The negatives signs in longitude λ indicate that the domain is located in the western hemisphere. Table (3.1) shows only the names of the BADC-7 stations used in the study along with their total number of records (in hours) and their data availability expressed in %.

Stations	**Total hours**	**Availability%**
Lerwick	473,169	98.1
Stornoway Airport	465,761	96.6
Valley	469,321	97.3
Aldergrove	473,956	98.3
Boscombe Down	463,732	96.2
Aberporth	473,761	98.3
Tiree	469,380	97.4

Table 3.1: *Data Quality Assessment of the BADC-7*

The BADC-57 stations selected are as shown in Table (3.2) along with their UK Met. Office station identifier and their region number to which they have been allocated. The UK region numbers are explained in Table (3.3). Figure (3.2) shows the geographical location of all the stations.

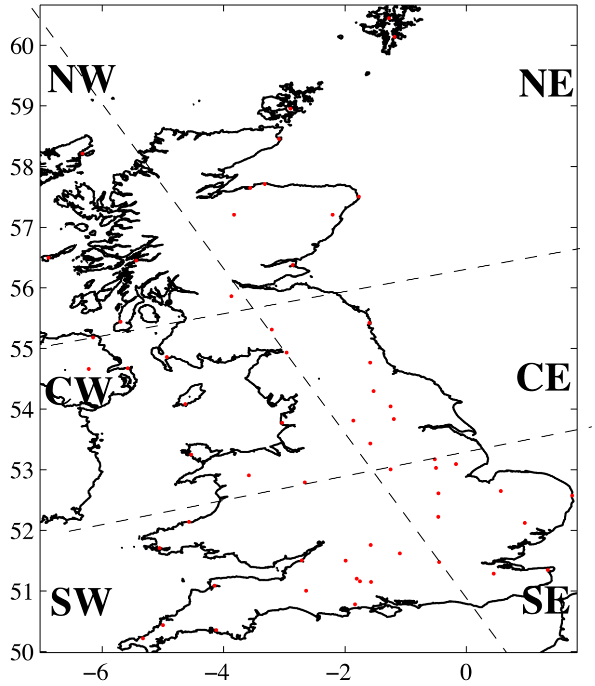


Figure 3.1: *Regional distribution of stations used in the study*

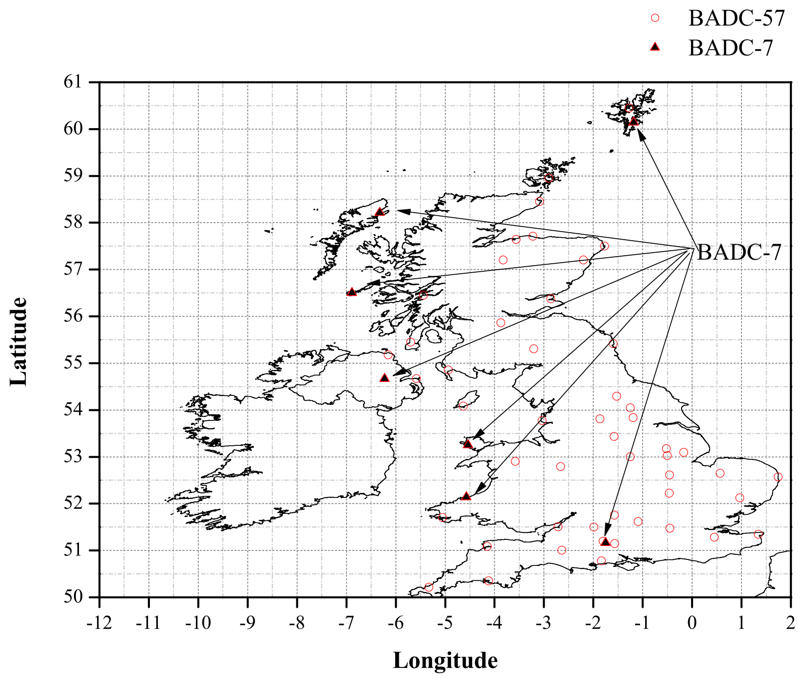


Figure 3.2: *Geographical location of stations used in the study*

Identifier	Name	Region
9	LERWICK	5
10	SELLA NESS	5
32	WICK AIRPORT	5
54	STORNOWAY AIRPORT	6
113	AVIEMORE	5
132	KINLOSS	5
137	LOSSIEMOUTH	5
161	DYCE	5
170	PETERHEAD HARBOUR	5
235	LEUCHARS	5
315	BOULMER	3
326	DURHAM	3
346	LINTON ON OUSE	3
384	WADDINGTON	3
386	CRANWELL	3
393	CONINGSBY	3
409	MARHAM	1
432	GORLESTON	1
440	WATTISHAM	1
461	BEDFORD	1
513	BINGLEY, NO 2	3
527	HIGH BRADFIELD	3
533	CHURCH FENTON	3
556	NOTTINGHAM, WATNALL	3
583	WITTERING	1
605	BRIZE NORTON	2
613	BENSON	1
643	SHAWBURY	4
674	AVONMOUTH	2
708	HEATHROW	1
744	EAST MALLING	1
775	MANSTON	1
842	HURN	2
847	MIDDLE WALLOP	2
886	LYNEHAM	2

continued on next page

continued from previous page		
Identifier	Name	Region
888	LARKHILL	2
889	BOSCOMBE DOWN	2
908	MACHRIHANISH	6
918	DUNSTAFFNAGE	6
982	SALSBURGH	6
1023	ESKDALEMUIR	4
1039	WEST FREUGH	4
1046	RONALDSWAY	4
1090	BLACKPOOL, SQUIRES GATE	4
1145	VALLEY	4
1180	BALA	4
1198	ABERPORTH	2
1215	MILFORD HAVEN CONSERVANCY BOARD	2
1302	YEOVILTON	2
1336	PLYMOUTH, MOUNTBATTEN	2
1346	CHIVENOR	2
1395	CAMBORNE	2
1450	ALDERGROVE	4
1467	BALLYPATRICK FOREST	6
1529	ORLOCK HEAD	4
17314	LEEMING	3
18974	TIREE	6

Table 3.2: *List of of BADC-57 stations*

No	Region
1	South Eastern UK
2	South Western UK
3	Centre Eastern UK
4	Centre Western UK
5	North Eastern UK
6	North Western UK

Table 3.3: *Key to Regions*

3.3 Factors that may Affect Wind Speed Measurements

The principal difficulty in producing a reliable wind analysis based on observed surface wind speed measurements is ensuring consistency and homogeneity of the data. Over time, site exposure can change, instruments can be replaced or re-calibrated and measurement heights or locations can be altered.

3.3.1 Site Exposure

In producing the 55−year and the 29−year analyses, the BADC-57 and BADC-7 stations were partly chosen to ensure that they were isolated rural sites to avoid changes associated with urbanisation. For brevity reasons, Figure (3.3) shows the wind roses for BADC-7 stations only. In each case, the wind is shown to come predominantly from between the South and the West which is common for UK sites, though Valley, on the West coast of Anglesey, North Wales, shows a relatively large proportion of winds from the East. This is due to the different topographical features of the sites. According to Lapworth and McGregor [125], the high ground over Wales, Northern England and Scotland has a significant effect on the pressure gradient, as the isobars *back* (turn anti-clockwise) over the western coasts and *veer* (turn clockwise) over the eastern half of the country. There is also no evidence for significant sheltering at any of the sites.

In addition, the mean wind speed by direction was examined over time to see whether there was any evidence of changes in site exposure. Figure (3.4) to Figure (3.6) show the annual mean wind speed by 30° direction sector for Stornoway Airport, Aldergrove and Tiree stations respectively. The grey shading on the plots represents one standard deviation of uncertainty (also known as error bars) calculated from the hourly values in each year. In each case a standard least squares linear fit is made to the data over the period 1957 to 2011, and 95% confidence limits are shown. Stornoway Airport (Figure 3.4) shows an overall decline but in all direction sectors. The wind speed declines from 1957 to the early 1990s and then shows an increase. An analysis of gale days for Stornoway Airport over the period 1884-1996 [126] shows a steady decline from around 1940 to the early 1980s and then provides evidence of an increase thereafter which is broadly consistent with the present analysis. However, Dawson et al. in [126] mentioned, the upturn seems to start around a decade earlier. This in turn implies that any trend is synoptic

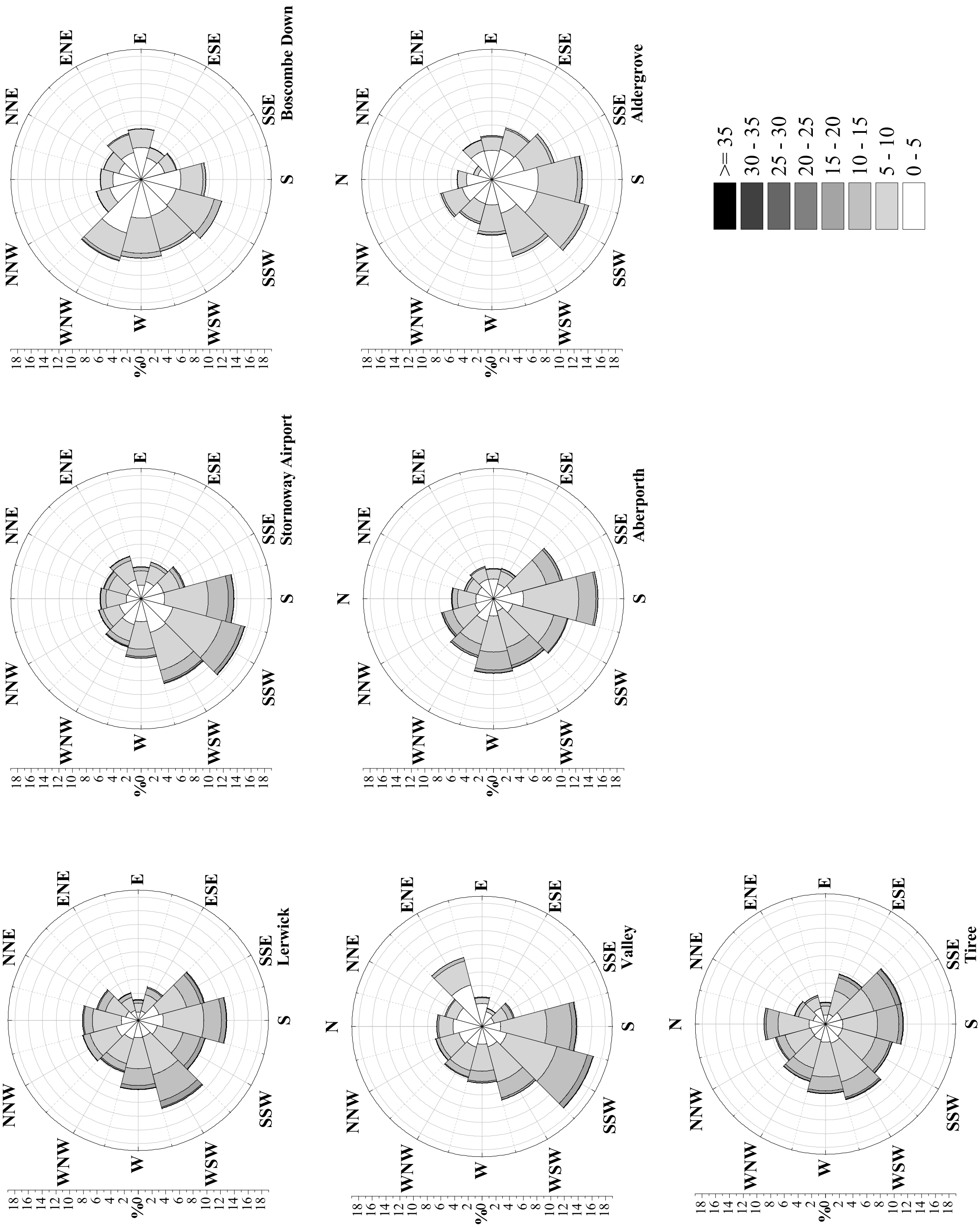


Figure 3.3: *Wind roses for the BADC-7 sites*

rather than exposure related. Aldergrove (Figure 3.5) shows some decline in wind speed in all directions. There is evidence of some urbanisation to the South-East of the site but for directions between the South-West and the North the site is open with Lough Neagh 3 km to the West. Therefore, changes in wind speed at this site are unlikely to be due to changes in exposure. Tiree (Figure 3.6) showcases evidence of some decline in wind speed from the East to the South-East. However, this site is extremely exposed, so it is unlikely that such a change is due to increased shelter. The remaining sites are presented in Appendix A and do not show any significant trends by direction sector when considering the level of inter-annual variation.

3.3.2 Instrument and Height Measurement Changes

Using information from the UK Meteorological Office Archive, instrument changes were deduced for the BADC-7 and BADC-57 stations. Table (3.4) summarises the changes for BADC-7 stations while information about the BADC-57 stations can be found in Appendix B. Over the 55-year period of measurements four different types of anemometers were successively used. It is also worth noting that the effective height of measurement has generally decreased over times, though there are some exceptions, e.g. Lerwick.

3.3.2.1 Vertical Extrapolation and Roughness Length

In order to ensure a consistent height for all wind speed values, wind speed data at all sites were corrected to an effective height of 10 m using the adiabatic logarithmic profile:

$$u = \frac{u_*}{\kappa} \ln \frac{z}{z_o} \tag{3.2}$$

where u is wind speed at the height of interest z, u_* is the friction velocity in the surface layer (assumed constant), κ is von Kármán's constant (=0.4) and z_o is the surface roughness length assumed to be 0.03 m for all sites. This value is appropriate for short grass which is typical of rural meteorological stations. Figure (3.7) shows an example of the annual mean wind speed values for Boscombe Down with and without correction to an effective height of 10 m. It can be seen in this case that the correction makes a significant difference to wind speed values in the

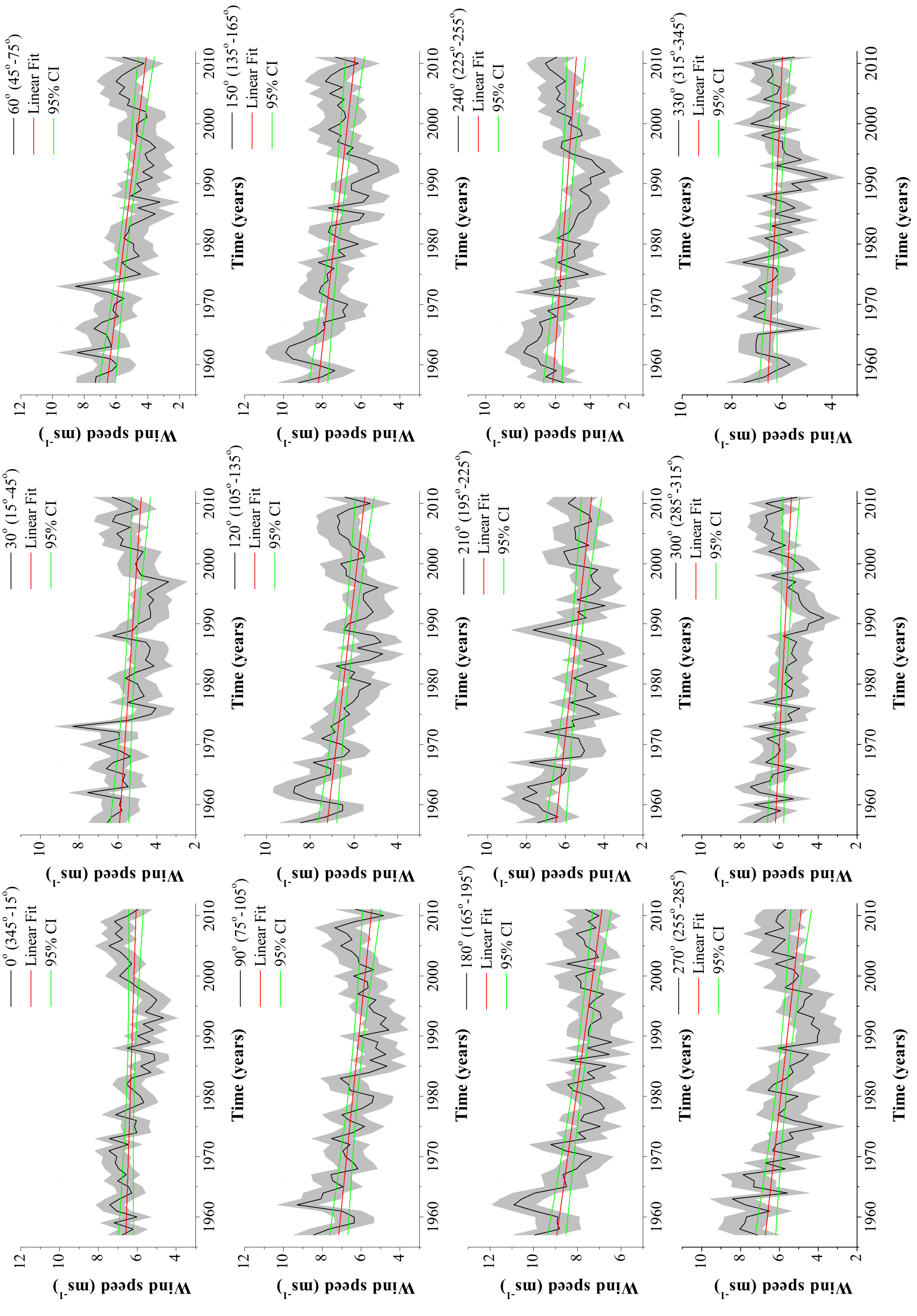


Figure 3.4: *Mean annual wind speed by 30° direction for Stornoway Airport*

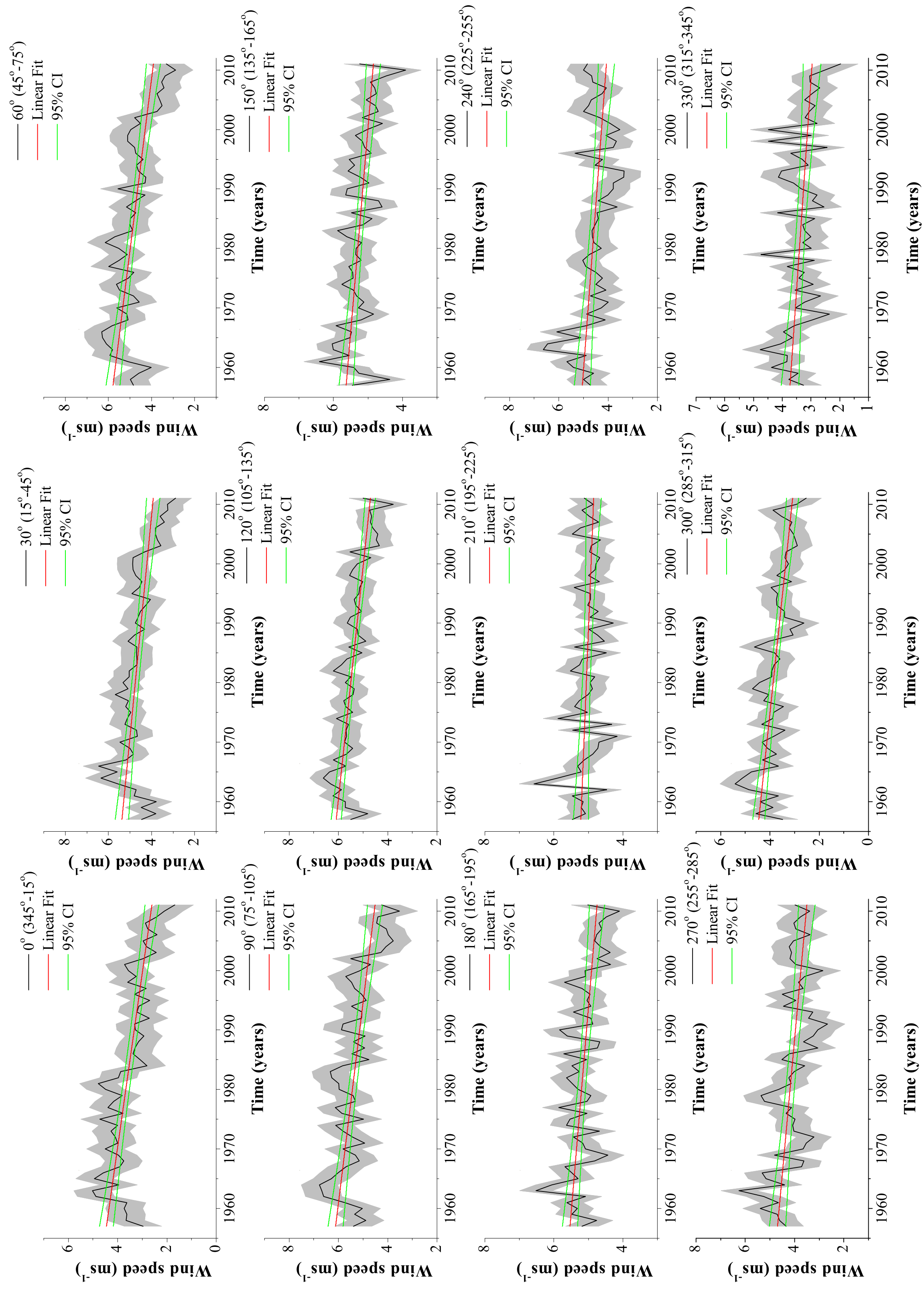


Figure 3.5: *Mean annual wind speed by 30° direction for Aldergrove*

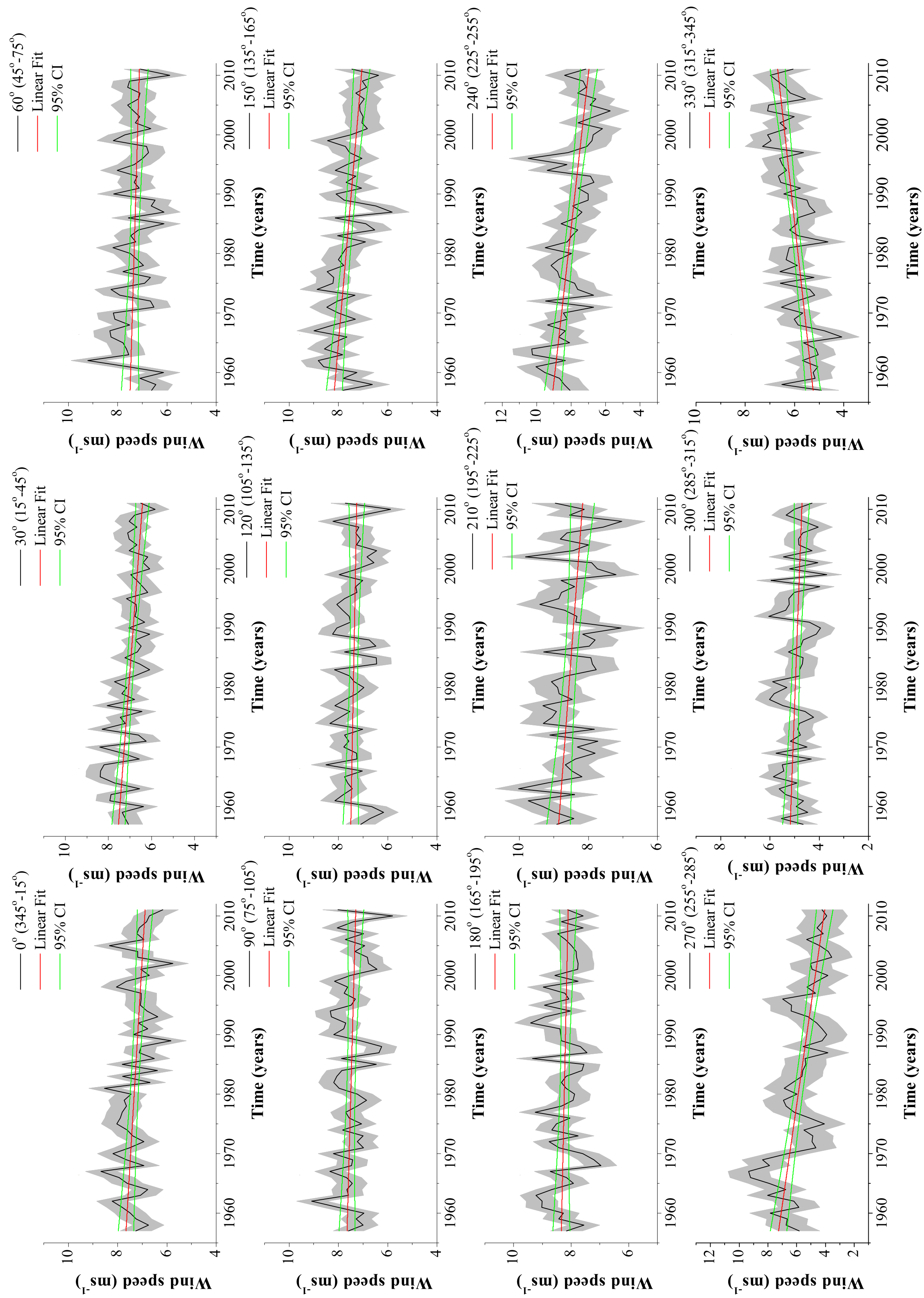


Figure 3.6: *Mean annual wind speed by 30° direction for Tiree*

Sites	Dates	Instrument Type	Effective Height (m)
Lerwick	Jan 1957 - Oct 1960	Assman Mk2 anemometer	6
	Oct 1960 - Jun 1984	Munro Mk4 anemometer	10
	Jun 1984 - Oct 1999	Munro Mk4 anemometer	10
	Oct 1999 - Dec 2009	Munro Mk6 anemometer	10
Stornoway Airport	Jan 1957 - Oct 1967	Anemograph †	14
	Oct 1967 - Oct 1971	Munro Mk4 anemometer	14
	Oct 1971 - Dec 1974	Munro Mk4 anemometer	10
	Dec 1974 - Aug 2002	Munro Mk5 anemometer	10
	Aug 2002 - Dec 2009	Munro Mk6 anemometer	10
Valley	Jan 1957 - Mar 1988	Munro Mk4 anemometer	12
	Apr 1988 - Mar 1995	Munro Mk4 anemometer	10
	Apr 1995 - Jul 2007	Munro Mk5 anemometer	10
	Jul 2007 - Dec 2009	Munro Mk6 anemometer	10
Aldergrove	Jan 1957 - Aug 1963	Pressure tube	17
	Sep 1963 - Jan 2003	Munro Mk4 anemometer	10
	Jan 2003 - Dec 2009	Munro Mk6 anemometer	10
Boscombe Down	Jan 1957 - Jun 1964	Anemograph †	16
	Jun 1964 - Dec 1972	Assman Mk2 anemometer	16
	Jan 1972 - Nov 2009	Munro Mk4 anemometer	10
	Nov 2009 - Dec 2009	Vector Mk6 anemometer	10
Aberporth	Jan 1957 - Jun 1966	Anemograph †	11
	Jun 1966 - Dec 1969	Anemograph †	11
	Feb 1970 - Apr 2000	Munro Mk4 anemometer	10
	Apr 2000 - Dec 2009	Vector Mk6 anemometer	10
Tiree	Jan 1957 - Sep 1970	Pressure tube	16
	Sep 1970 - Sep 1980	Munro Mk4 anemometer	12
	Oct 1980 - Jul 2001	Munro Mk4 anemometer	10
	Jul 2001 - Dec 2009	Vector Mk6 anemometer	10

* December 31st, 2009, is the last record used in this study.
† Particular anemometer type not specified.

Table 3.4: *Instrument changes since 1957 for the seven stations (BADC-7) used in the study*.*

earlier years of the time series. It is also worth noting that if the effective height of measurement had not been applied, then wind speeds would have been overestimated, leading to misleading results/conclusions.

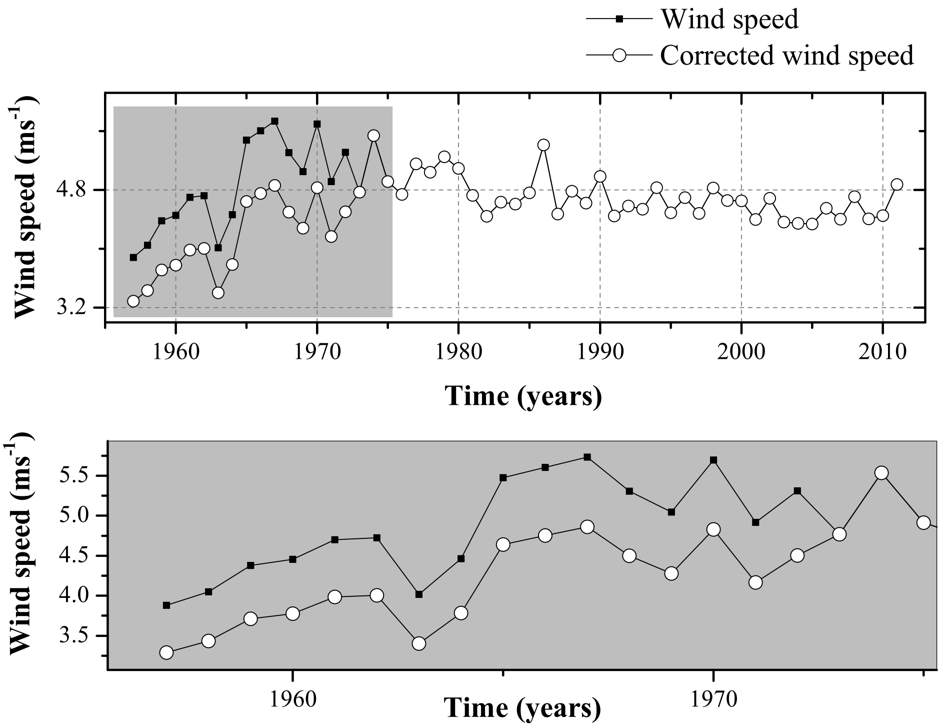


Figure 3.7: *Correction in instrumentation's effective height*

The value of roughness length (0.03) was chosen as being fairly representative of a rural site. It is possible that this might vary anywhere in the range of 0.01 m to 0.08 m for the sites studied. Therefore, a sensitivity analysis of the change in wind speed as a function of roughness length between 0.01 m and 0.08 m was performed. However, it was found that there was negligible difference to the wind speed when corrected to 10 m using this range of roughness values. Although the topic of *wind indices*[6] will be discussed in full details in Chapter 4 it is very important at this stage to illustrate that the differences between the tested roughness values are minor. Figure (3.8) shows the box plot of the the annual wind speeds for Lerwick station. Changing the roughness values affects slightly the percentiles and the mean value of each index. It is clearly shown that such a change does not affect massively the measurements, and no major differences are observed. However, the next section will further question the choice of setting a value of roughness length equal to 0.03 and it will be debated against other proposals.

[6]A wind index provides an indication of the mean wind speed, usually annually and/or monthly, relative to the calculated long-term mean wind speed [127].

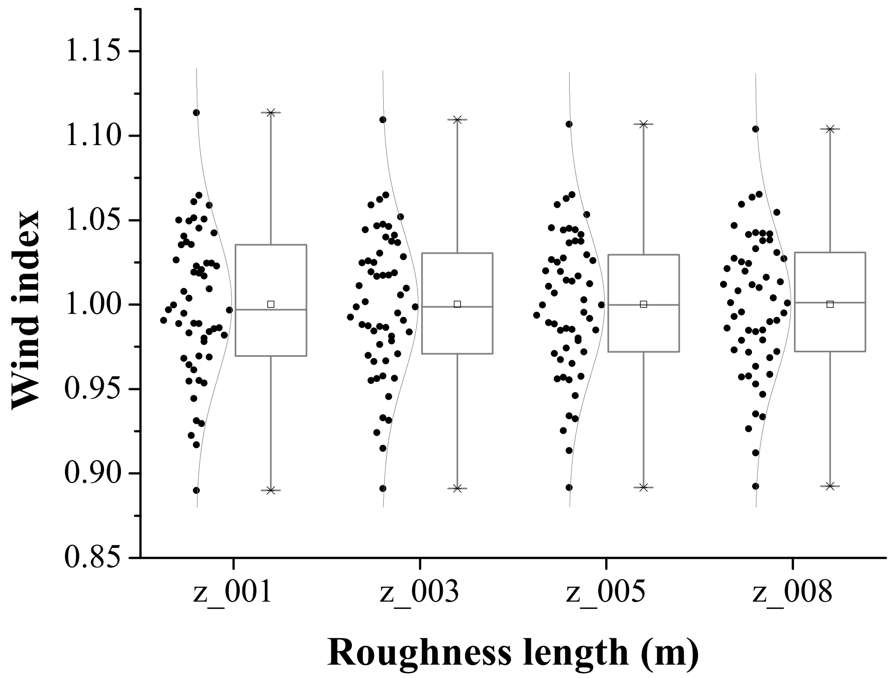


Figure 3.8: *Box plot of 55-year index for Lerwick stations*

3.3.2.2 CORINE Dataset

In 1985 the European Environment Agency (EEA) of the European Commission launched the COordinate INformation for the Environment (CORINE) programme which aimed to provide information on several issues related to the state of the environment and natural sources within the European Union. From this European framework a database of the CORINE land cover (CLC) was created. CLC covers geographically 12 countries within Europe with a mapping scale at 1:100,000 and minimum mapping unit (MMU) fixed at 25 hectares (ha). The aim of this project is to provide cartographic information about land coverage and land occupation. One can found more details published by Bossard et al. [128] and by the EEA [129].

Silva et al. [130] used a roughness length classification based on CLC data and concluded that their classification shows good agreement with the results from site inspections. Table 3.5 shows the 44 classes of CLC with their corresponding roughness length values. However, when satellite images were used to inspect the sites used in the present study, several discrepancies were found that discouraged the use of this dataset. For example, Aberporth's CLC code is 121 which means that the roughness value is 0.5 m. However, it is very clear from the satellite image that the land covering that meteorological station is short grass (0.03 m). The Figure (3.9) was retrieved from the free web service Google Earth [131]. Given the inconsistency observed between the CLC classification and the satellite image, the option of using the CORINE dataset was dismissed, and instead values for roughness length were used, as it was described in the previous section.

CLC codes	Type of Land	Roughness length (z)
111	Continuous Urban Fabric	1.2
112	Discontinuous Urban Fabric	0.5
121	Industrial or commercial units	0.5
122	Road and rail networks and associated land	0.075
123	Port areas	0.5
124	Airports	0.005
131	Dump sites	0.005
132	Mineral extraction sites	0.005
133	Construction sites	0.5
141	Green urban areas	0.6
142	Sport and leisure facilities	0.5
211	Non-Irrigated Arable Land	0.05
212	Permanently irrigated land	0.05
213	Rice fields	0.05
221	Fruit trees and berry plantations	0.1
222	Vineyards	0.1
223	Olive groves	0.1
231	Pastures	0.03
241	Annual crops associated with permanent crops	0.1
242	Agro-forestry areas	0.3
243	Complex cultivation patterns	0
244	Land principally occupied by agriculture	0.3
311	Broad-leaved forest	0.75
312	Coniferous forests	0.75
313	Mixed forests	0.75
321	Sclerophyllous vegetation	0.03
322	Moors and heathland	0.03
323	Natural grasslands	0.03
324	Transitional woodland-shrub	0.6

continued on next page

continued from previous page		
CLC codes	Type of Land	Roughness length (m)
331	Beaches, dunes, and sand plains	0
332	Bare rocks	0.005
333	Sparsely vegetated areas	0.005
334	Burnt areas	0.6
335	Glaciers and perpetual snow	0.001
411	Inland marshes	0.05
412	Peat bogs	0.0005
421	Salt marshes	0.05
422	Salines	0.0005
423	Intertidal flats	0.0005
511	Water courses	0
512	Water bodies	0
521	Sea and Ocean	0
522	Estuaries	0
523	Coastal lagoons	0

Table 3.5: *Value of roughness length based on CORINE land cover classes*

3.3.3 Data Contamination due to Human Errors

The operational period of the stations used in this study goes back to the end of 1940's. The actual measurements in those days, instruments logs as well as inspectors' reports were kept in hard copies. As technology improved, the files in the Met Office were also updated into a digital format to help scientists and the public utilise this information in a more efficient way. However, the original information is still maintained at the Met Office Archives in Exeter, UK. An inspection and comparison between the hard copies and the digital ones was decided in this study to gain information related to the collection and storage of data. This action revealed any incompatibilities between the hard copies and the digital ones. This visual inspection served as a quality control technique to prevent data contamination due to data entry errors. From this comparison only a few cases were found to be in a need for special action. Figure (3.10) illustrates a photo taken from the Archives for Lerwick station. The numbers shown in the snapshot refer to monthly values in kt. Similarly, Table (3.6) shows the same information from MIDAS database. However, it is noteworthy that the monthly values in the year 1978 for the

Figure 3.9: *Satellite image for Aberporth station*

Mean wind speed (kt)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
1957	22.5	15.0	16.0	15.0	11.3	12.6	12.2	12.8	14.8	16.8	14.2	17.3	15.1
1958	18.0	16.5	14.2	11.0	12.3	9.6	11.7	10.9	11.4	16.2	15.2	18.2	13.8
1959	17.2	20.7	16.5	14.1	11.4	17.2	12.7	12.8	12.2	16.6	16.2	20.2	15.7
1960	14.4	17.2	19.4	17.6	11.9	13.7	11.0	11.1	12.4	13.5	17.6	16.5	14.7
1961	18.5	18.1	21.5	14.1	14.1	16.5	13.7	12.8	13.0	19.2	13.7	16.6	16.0
1962	18.2	17.4	12.4	10.8	11.6	11.7	11.0	11.8	12.0	13.7	13.3	13.3	13.1
1963	11.8	12.0	16.3	13.6	11.8	10.6	12.3	9.3	16.1	17.8	16.3	14.9	13.6
1964	16.2	16.5	17.6	13.4	13.1	13.2	14.2	12.1	11.8	12.0	14.5	15.4	14.2
1965	14.7	14.0	12.8	12.9	13.2	12.1	11.0	11.9	13.5	14.5	12.8	16.4	13.3
1966	15.0	11.7	18.0	12.8	14.0	10.9	12.2	11.0	15.1	13.1	16.1	18.9	14.1
1967	16.8	19.1	21.9	17.2	13.3	12.8	12.3	9.4	10.1	16.1	16.4	18.0	15.3
1968	17.7	13.6	15.9	12.4	13.6	11.1	10.0	8.9	11.4	13.0	15.3	16.1	13.3
1969	16.0	12.7	16.4	14.6	9.9	10.5	12.7	10.2	15.0	16.2	14.0	17.0	13.8
1970	17.4	13.3	15.1	13.2	12.0	8.8	11.8	10.0	11.2	14.2	14.4	15.9	13.1
1971	14.8	15.8	14.5	12.7	11.4	10.9	11.8	10.5	11.4	14.6	17.3	18.0	13.6
1972	17.1	13.4	13.8	12.2	10.8	10.3	7.9	11.2	10.0	12.9	17.3	16.8	12.8
1973	15.3	13.9	15.2	13.9	10.4	10.0	8.6	9.1	11.7	10.6	16.5	16.7	12.7
1974	20.2	13.6	11.2	8.2	12.5	10.6	10.3	10.6	12.5	12.1	12.4	19.3	12.8
1975	17.5	12.6	12.0	13.2	12.5	11.1	9.9	10.6	14.7	12.6	15.4	20.8	13.6
1976	18.9	18.5	20.1	15.5	13.8	12.2	12.4	9.9	15.3	15.5	15.5	11.6	14.9
1977	14.5	13.4	15.3	14.9	10.2	11.3	9.6	10.9	15.1	14.3	15.0	18.5	13.6
1978	19.2	15.0	15.2	11.1	10.2	12.8	12.4	10.3	15.2	0	blank	blank	13.5
1979	15.1	16.5	19.0	13.0	14.1	12.3	11.5	10.8	16.3	17.8	15.3	19.5	15.1
1980	16.9	15.1	15.6	15.8	11.9	10.9	11.6	11.3	13.0	17.1	15.9	20.7	14.6

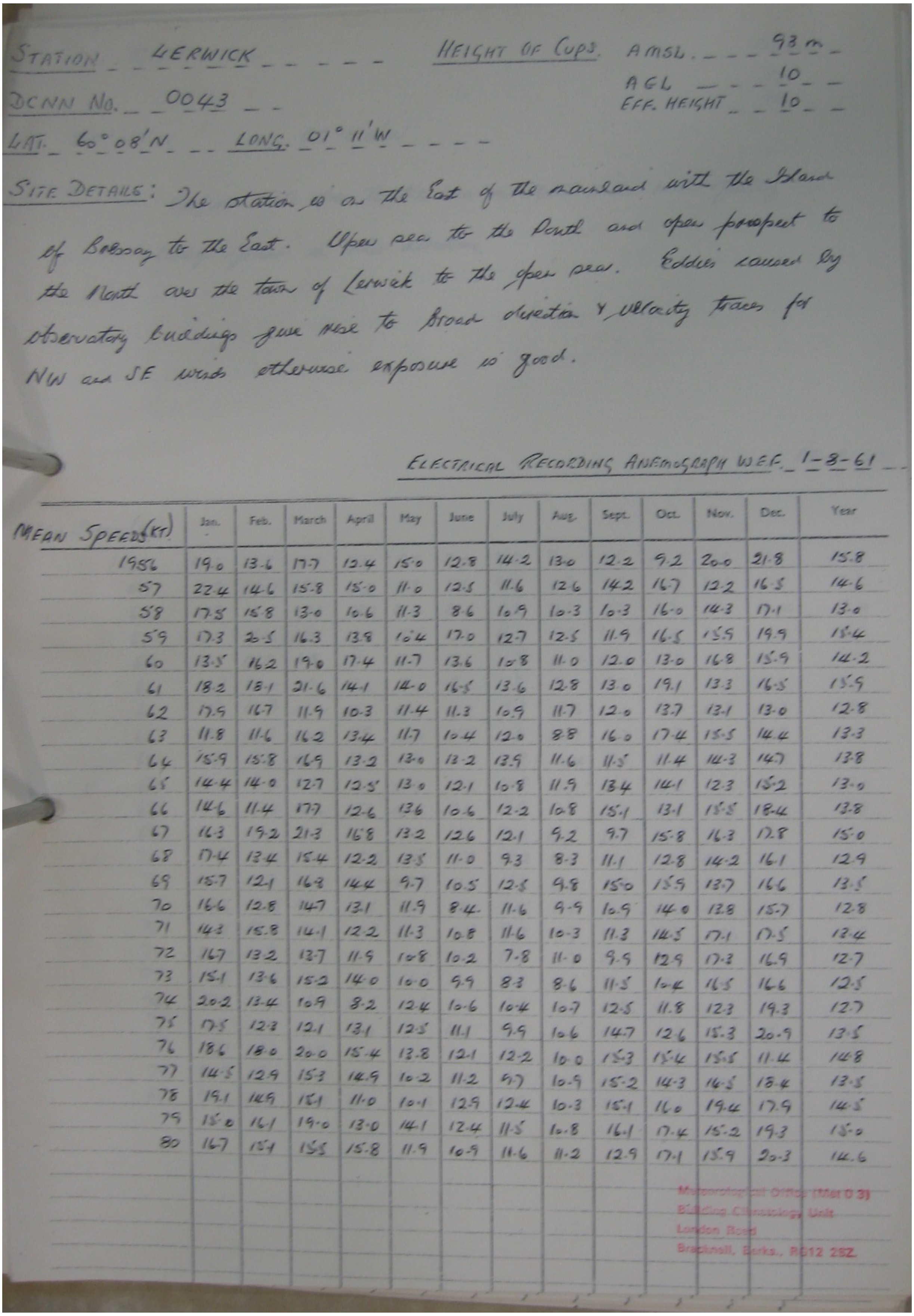
Table 3.6: *Monthly and yearly wind speed records at Lerwick station (digital copy)*

months October, November, and December are 0, blank and blank respectively. In that special case, instead of using statistical techniques to fill in the gaps information from the genuine historical logs was used. Overall, the handwritten records matched the digitised values with differences between the two formats being negligible.

3.4 Database

After the data had been retrieved from MIDAS, they were stored in a MySQL® database on a dedicated server which is consisted of two quad-core INTEL XEON X5355 processors 2.66 GHz with a total RAM of 32 GB, 667 MHz Fully Buffered DIMM (FBD).

Figure (3.11) shows an example of raw data from the ERA-40 dataset. It is clear that the data



STATION: LERWICK

HEIGHT OF CUPS. AMSL: 93 m; AGL: 10; EFF. HEIGHT: 10

DCNN No. 0043

LAT. 60° 08' N LONG. 01° 11' W

SITE DETAILS: The station is on the East of the mainland with the Island of Bressay to the East. Open sea to the South and open prospect to the North over the town of Lerwick to the open sea. Eddies caused by observatory buildings give rise to broad direction & velocity traces for NW and SE winds otherwise exposure is good.

ELECTRICAL RECORDING ANEMOGRAPH W.E.F. 1-8-61

MEAN SPEEDS (KT)	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
1956	19.0	13.6	17.7	12.4	15.0	12.8	14.2	13.0	12.2	9.2	20.0	21.8	15.8
57	22.4	14.6	15.8	15.0	11.0	12.5	11.6	12.6	14.2	16.7	12.2	16.5	14.6
58	17.5	15.8	13.0	10.6	11.3	8.6	10.9	10.3	10.3	16.0	14.3	17.1	13.0
59	17.3	20.5	16.3	13.8	10.4	17.0	12.7	12.5	11.9	16.5	15.5	19.9	15.4
60	13.5	16.2	19.0	17.4	11.7	13.6	10.8	11.0	12.0	13.0	16.8	15.9	14.2
61	18.2	18.1	21.6	14.1	14.0	16.5	13.6	12.8	13.0	19.1	13.3	16.5	15.9
62	17.9	16.7	11.9	10.3	11.4	11.3	10.9	11.7	12.0	13.7	13.1	13.0	12.8
63	11.8	11.6	16.2	13.4	11.7	10.4	12.0	8.8	16.0	17.4	15.5	14.4	13.3
64	15.9	15.8	16.9	13.2	13.0	13.2	13.9	11.6	11.5	11.4	14.3	14.7	13.8
65	14.4	14.0	12.7	12.5	13.0	12.1	10.8	11.9	13.4	14.1	12.3	15.2	13.0
66	14.6	11.4	17.7	12.6	13.6	10.6	12.2	10.8	15.1	13.1	15.5	18.4	13.8
67	16.3	19.2	21.3	16.8	13.2	12.6	12.1	9.2	9.7	15.8	16.3	17.8	15.0
68	17.4	13.4	15.4	12.2	13.5	11.0	9.3	8.3	11.1	12.8	14.2	16.1	12.9
69	15.7	12.1	16.8	14.4	9.7	10.5	12.5	9.8	15.0	15.9	13.7	16.6	13.5
70	16.6	12.8	14.7	13.1	11.9	8.4	11.6	9.9	10.9	14.0	13.8	15.7	12.8
71	14.3	15.8	14.1	12.2	11.3	10.8	11.6	10.3	11.3	14.5	17.1	17.5	13.4
72	16.7	13.2	13.7	11.9	10.8	10.2	7.8	11.0	9.9	12.9	17.3	16.9	12.7
73	15.1	13.6	15.2	14.0	10.0	9.9	8.3	8.6	11.5	10.4	16.5	16.6	12.5
74	20.2	13.4	10.9	8.2	12.4	10.6	10.4	10.7	12.5	11.8	12.3	19.3	12.7
75	17.5	12.3	12.1	13.1	12.5	11.1	9.9	10.6	14.7	12.6	15.3	20.9	13.5
76	18.6	18.0	20.0	15.4	13.8	12.1	12.2	10.0	15.3	15.4	15.5	11.4	14.8
77	14.5	12.9	15.3	14.9	10.2	11.2	9.7	10.9	15.2	14.3	16.5	13.4	13.5
78	19.1	14.9	15.1	11.0	10.1	12.9	12.4	10.3	15.1	16.0	19.4	17.9	14.5
79	15.0	16.1	19.0	13.0	14.1	12.4	11.5	10.8	16.1	17.4	15.2	19.3	15.0
80	16.7	15.1	15.5	15.8	11.9	10.9	11.6	11.2	12.9	17.1	15.9	20.3	14.6

Meteorological Office (Met O 3)
Building Climatology Unit
London Road
Bracknell, Berks., RG12 2SZ

Figure 3.10: *Monthly and yearly wind speed records at Lerwick station (hard copy)*

lacked in structure and organisation. Integers in the first column on the left (0, 6, 12, etc.) are timestamps while all decimal values below correspond to the u' wind component with each cell having a unique latitude/longitude. Therefore, it was necessary for all data to be grouped in such a way that records or relations between them were presented in a logical tabular form (i.e. columns structured as: timestamp, latitude, longitude, u', etc.). As it is mentioned by Teorey et al. [132], MySQL® has these attributes and by running as a server provides access to multiple users in a central database or other distributed databases. Further, duplicated values were detected that were tackled using MySQL commands.

0																
2,54727	3,48868	4,24942	4,39005	4,04141	3,77579	4,38419	5,62442	6,65469	6,767	6,142	4,80997	1,84415	1,18399	1,61758	2,05801	2,20645
2,68301	3,66739	4,33926	4,34122	3,96915	3,89883	4,62149	5,61661	6,46426	5,77676	5,21622	3,01602	1,23672	0,455475	0,759186	1,32266	1,67032
1,91934	2,93887	3,63907	3,71036	3,43008	3,50528	4,29532	5,20157	5,77286	4,3959	3,59415	2,18399	0,4711	-0,425385	-0,467377	-0,110931	0,331451
0,091217	1,25723	2,17911	2,74063	2,97305	3,13809	3,61758	4,75528	5,39688	4,09415	2,56778	0,501373	-0,842377	-1,2037	-1,33066	-1,4664	-0,43808
-1,26816	-0,154877	0,609772	1,26309	1,76407	2,03555	2,50919	3,97891	4,29727	2,83438	2,05899	0,0521545	-1,28964	-1,45859	-1,56113	-0,639252	0,927155
-1,78378	-0,964447	0,306061	0,585358	0,731842	0,693756	1,02481	1,76016	0,735748	0,831451	0,685944	-0,56601	-1,42831	-1,20273	-0,583588	0,666412	1,13809
-2,78378	-1,78769	-0,730072	-0,285736	-0,405853	-0,752533	-0,792572	-0,715424	-1,5416	-0,511322	-0,214447	-0,608002	-1,16855	-1,28671	-0,567963	1,69864	1,52188
-3,72031	-2,65097	-1,78378	-1,48886	-1,74277	-2,06992	-2,01425	-2,05624	-2,13535	-1,14511	-0,854095	-0,970306	-1,23007	-1,21542	-0,277924	2,22989	2,02383
-4,2955	-3,48593	-2,70078	-2,28378	-2,25351	-2,16171	-1,7789	-1,87851	-2,00449	-1,29941	-1,10995	-1,17245	-1,1666	-0,767181	0,292389	2,22696	2,38809
-4,89804	-4,25449	-3,45566	-2,66464	-1,99765	-1,30722	-0,582611	-0,355072	-0,325775	-0,264252	-0,65976	-0,882416	-0,632416	0,0248108	1,04239	2,4877	3,56876
6																
3,83463	3,88638	3,56705	3,1774	3,23697	3,84537	4,62173	5,819	6,67935	6,81509	6,1481	4,41373	0,270172	-0,760101	-0,6185	-0,475922	-0,0423279
3,65787	3,44791	2,92935	2,63736	2,86783	3,4274	4,0983	5,1481	5,9108	5,27603	4,5856	1,46841	-0,753265	-1,55014	-1,34506	-1,05991	-0,707367
2,57291	2,12369	1,65396	1,65298	2,00357	2,41959	3,06802	3,7399	3,76627	1,61295	1,65103	0,392242	-1,48471	-2,45834	-2,31674	-2,12436	-2,02084
0,674469	0,303375	0,0660706	0,265289	0,63443	0,937164	1,32193	1,32584	0,930328	0,0846252	-0,161469	-1,37143	-2,85678	-3,40854	-2,9808	-2,81284	-2,55502
-1,04428	-0,958344	-0,949554	-0,859711	-0,690765	-0,571625	-0,767914	-1,44955	-1,88998	-1,56381	-1,5892	-2,43295	-3,18588	-3,11557	-2,54819	-1,78842	-0,874359
-1,82162	-1,55795	-1,34116	-1,39194	-1,48373	-1,67123	-2,12534	-3,04526	-3,25717	-2,99057	-2,67416	-2,53354	-2,41733	-2,01205	-1,2435	-0,397797	0,334625
-2,79526	-2,51596	-2,16245	-1,8812	-1,69858	-1,71127	-2,05502	-3,1517	-3,79428	-2,80209	-2,4642	-2,10873	-1,70932	-1,43686	-1,056	0,346344	0,984039
-3,67709	-3,20346	-2,68198	-2,07455	-1,51596	-1,27475	-1,42612	-2,37827	-2,93002	-1,9183	-1,7142	-1,44174	-1,07065	-0,684906	-0,00424194	1,52896	1,68521
-4,11948	-3,67905	-3,05698	-2,2435	-1,44662	-0,92514	-0,742523	-1,01401	-1,1058	-0,702484	-0,691742	-0,693695	-0,585297	-0,201508	0,742828	2,27505	2,37173
-4,58334	-4,09604	-3,43491	-2,62436	-1,76303	-0,969086	-0,359711	0,236969	0,735016	0,489899	0,345367	0,112946	0,0592346	0,400055	1,2731	2,74576	3,54459
12																
3,78604	3,43057	3,08389	3,23428	3,85831	4,43448	4,84171	5,59073	6,08682	6,03897	5,11417	3,29092	-0,539154	-1,00693	0,162018	0,930573	1,12784
2,77432	2,66104	2,51846	2,71964	3,22842	3,62393	3,92764	4,44034	4,72061	4,35831	3,70303	0,420807	-1,41513	-1,38388	-0,464935	-0,0430603	-0,0245056
0,687408	0,938385	1,1962	1,56534	1,95792	2,09366	2,04971	2,31631	2,35635	-0,76474	-1,42294	-2,11533	-2,48349	-2,18954	-1,96884	-1,82724	-1,87802
-1,87216	-1,16415	-0,294037	0,260651	0,338776	0,0506897	-0,342865	0,00967407	-0,119232	-3,44736	-5,10068	-5,754	-5,03232	-3,9122	-3,24521	-2,99521	-3,09091
-2,93759	-2,28915	-1,62314	-1,16318	-1,10751	-1,45126	-1,76669	-1,37021	-2,23349	-5,00302	-5,8331	-6,08603	-5,44931	-4,28525	-3,20322	-2,55771	-2,32529
-2,69443	-2,31259	-2,18759	-2,01572	-1,86142	-2,09286	-2,27353	-2,16025	-3,88779	-4,56357	-4,52939	-4,45224	-3,93564	-2,81357	-1,629	-0,572357	0,171783
-3,88681	-3,55673	-3,25204	-2,76669	-2,2999	-2,20712	-2,22861	-2,04306	-2,07333	-2,83408	-2,90732	-2,86337	-2,1915	-0,977631	0,314362	1,63467	2,62881
-4,86923	-4,34677	-3,83798	-3,24716	-2,60751	-2,03232	-1,57822	-1,10654	-0,518646	-1,16513	-1,57822	-1,43759	-0,67099	0,426666	1,72745	3,13858	4,1503
-5,39853	-4,97568	-4,3165	-3,59384	-2,86142	-1,92685	-0,962006	-0,0450134	0,999908	0,0409241	-0,395599	-0,321381	0,194244	1,10049	2,51358	3,93546	5,01065
-5,75302	-5,38193	-4,69931	-3,93564	-3,18271	-2,17392	-1,00009	0,311432	1,56729	0,582916	0,158112	0,177643	0,615143	1,54288	3,0546	4,4296	5,53409
18																
5,05463	4,94525	4,81732	4,89545	5,11517	5,32806	5,57513	5,46576	5,20306	4,75775	3,49603	1,79584	-0,376038	0,254822	1,51459	2,11224	1,54291
4,69037	4,29486	3,98334	3,94916	4,00775	4,07123	4,33588	4,32123	4,3642	3,40033	2,34564	-0,437561	-0,66803	0,287048	1,16107	1,33002	0,705017
3,08978	2,70111	2,49506	2,37885	2,29095	2,41107	2,7558	3,13568	2,72748	-0,870178	-1,81061	-1,28815	-0,448303	0,121033	0,333923	0,264587	-0,203186
0,629822	0,677673	0,793884	0,703064	0,542908	0,664001	1,25189	2,85736	1,46576	-2,86627	-3,10065	-2,07526	-0,887756	-0,353577	-0,25592	-0,450256	-0,682678

Figure 3.11: *u' component raw data from ERA-40*

MySQL® was chosen as the favoured RDBS since it is an open source, cross-platform software [133]. Another feature equally important for choosing MySQL® as the selected RDBS is the fact that it is easy to be maintained. Also, its portable implementation and operation in different systems such as Unix, Linux, Windows, Solaris, MacOS X etc. gives flexibility to users without compromising or jeopardising the smoothness of any process. Another key element in choosing this specific RDBS is its speed in storing, updating, querying and retrieving data. The ability of MySQL® to communicate with other languages/softwares creates the ideal conditions for utilising in full extend the strengths of the aforementioned RDBS. Several tools are available for connecting to MySQL® from Java, C/C++, Perl, etc. In this study MatLab® has been used as the main software for all the numerical calculations. This required the use of the open database connectivity (ODBC) driver provided and documented in full details by ORACLE [134]. In addition, security is very important when dealing with sensitive data. Any breach in security may result to unauthorised modification which in turn can cause loss of data or malfunction of the database(s). However, permissions, privileges and/or restrictions to a database/table can be set for all users via MySQL® Administrator. Last, MySQL® offers easy management even to novice users. Various tools for doing so are offered such as the MySQL® command line client tool [135] and graphical user interfaces (GUI) like the MySQL® Administrator [136] and MySQL® Query Browser [137]. There are also other tools available similar to MySQL® Query Browser such as the HeidiSQL software [138]. The preference in using HeidiSQL, in contrast to other alternatives, lies in the fact that the specific software is capable of exporting results in tabular form straightforward to LaTeX which is the selected document preparation system in the present thesis. Figures (3.12) and (3.13) show the working environment for the MySQL® command line client tool and HeidiSQL respectively.

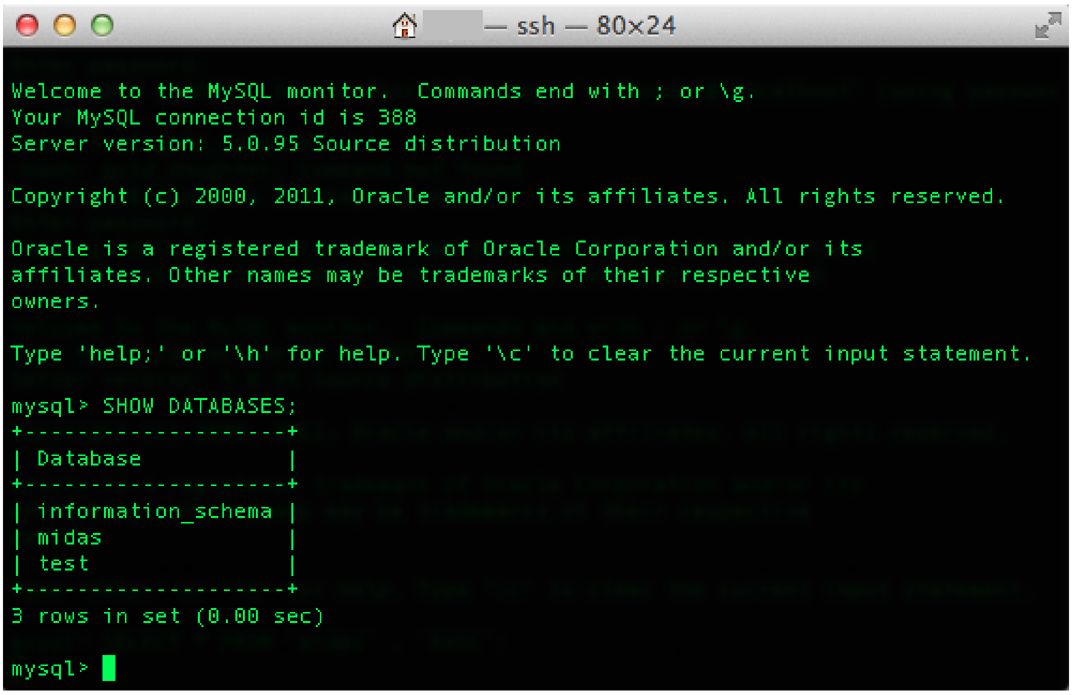


Figure 3.12: *Using MySQL® databases via the command line client tool*

3.5 Data Treatment

One crucial step towards the conduction and completion of the research presented herein involved handling big data. This necessitated the efficient design of a system which while it would be simple for the average user it would also provide accuracy and flexibility. Thus, as presented previously, the MySQL® database system was selected. This section provides information regarding the selected engine type, the structure of the database and the necessary actions for ensuring that the quality of data has not been compromised.

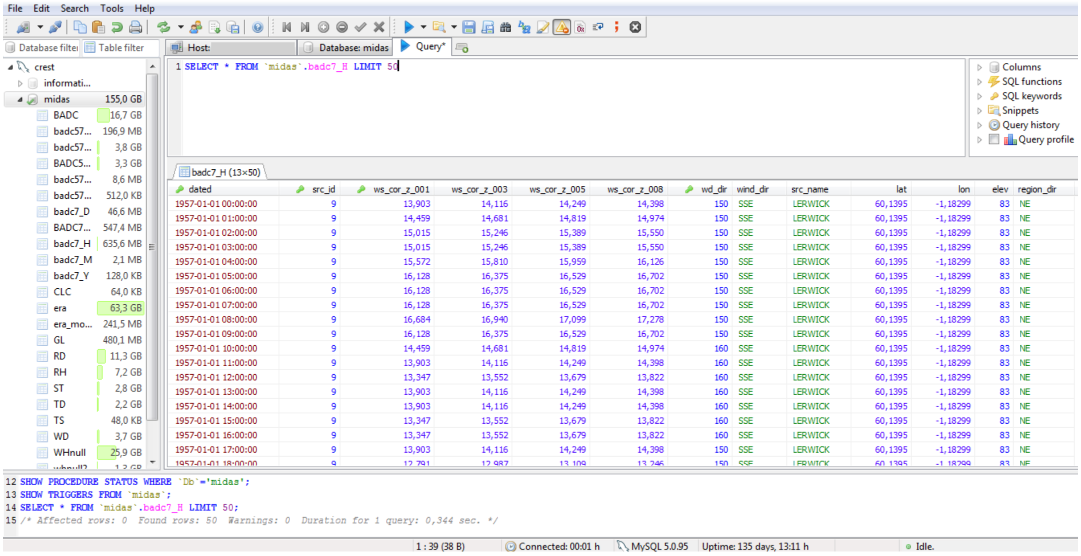


Figure 3.13: *Using HeidiSQL in order to provide an overview of the databases running within the developed RDBS.*

3.5.1 Database Structure

InnoDB was selected as the most appropriate storage engine as it offers referential integrity. This is achieved by assigning a foreign key as a reference to a primary key for the same table or other tables within the same database. Among others, InnoDB type offers to users flexibility without complicating procedures such as recovering tables on the crash and grouping multiple concurrent inserts. Most importantly, whilst InnoDB is not famous for its speed, it is ideal when one would process high volume of data [139]. The table below illustrates a comparison between different engine types. At this point, it is noteworthy to mention that only engine types under the General Public License (GPL) were taken into account. Moreover, engines such as the Berkeley DB were not included as they are not classified as RDBS.

Name	Vendor	License	Transactional	Current Development
MyISAM	ORACLE	GPL	NO	NO
InnoDB	ORACLE	GPL	YES	YES
Archive	ORACLE	GPL	NO	YES
CSV	ORACLE	GPL	NO	YES

Table 3.7: *Comparison of storage engines for the MySQL® RDBS*

3.5.2 Importing the Data

The data held at the BADC are stored in ASCII format and therefore prior to any analysis the data had to be downloaded and stored on the database. The code below serves as an example on how the raw data were imported to the table "BADC" which hosts the raw hourly data from the BADC.

```
LOAD DATA LOCAL INFILE
'/Users/elpk/BADC/midas_wxhrly_195701-195712.txt'
INTO TABLE BADC
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
(ob_time, id, id_type, met_domain_name, version_num, src_id, rec_st_ind,
wind_speed_unit_id, src_opr_type, wind_direction, wind_speed, prst_wx_id,
past_wx_id_1, past_wx_id_2, cld_ttl_amt_id, low_cld_type_id,
med_cld_type_id, hi_cld_type_id, cld_base_amt_id, cld_base_ht, visibility,
msl_pressure, cld_amt_id_1, cloud_type_id_1, cld_base_ht_id_1,
cld_amt_id_2, cloud_type_id_2, cld_base_ht_id_2, cld_amt_id_3,
cloud_type_id_3, cld_base_ht_id_3,cld_amt_id_4, cloud_type_id_4,
cld_base_ht_id_4,vert_vsby, air_temperature, dewpoint, wetb_temp, stn_pres,
ground_state_id, q10mnt_mxgst_spd, cavok_flag, cs_hr_sun_dur,
wmo_hr_sun_dur, wind_direction_q, wind_speed_q, prst_wx_id_q,
cld_ttl_amt_id_q, low_cld_type_id_q, med_cld_type_id_q, hi_cld_type_id_q,
cld_base_amt_id_q, cld_base_ht_q, visibility_q, msl_pressure_q,
air_temperature_q, dewpoint_q, wetb_temp_q, ground_state_id_q,
cld_amt_id_1_q, cloud_type_id_1_q, cld_base_ht_id_1_q, cld_amt_id_2_q,
cloud_type_id_2_q, cld_base_ht_id_2_q, cld_amt_id_3_q, cld_base_ht_id_3_q,
cld_amt_id_4_q, cloud_type_id_4_q, cld_base_ht_id_4_q, vert_vsby_q,
stn_pres_q, alt_pres_q, q10mnt_mxgst_spd_q,ccs_hr_sun_dur_q,
wmo_hr_sun_dur_q, meto_stmp_time, midas_stmp_etime, wind_direction_j,
wind_speed_j, prst_wx_id_j, past_wx_id_1_j, past_wx_id_2_j, cld_amt_id_j,
cld_ht_j, visibility_j, msl_pressure_j, air_temperature_j, dewpoint_j,
wetb_temp_j, vert_vsby_j, stn_pres_j, alt_pres_j, q10mnt_mxgst_spd_j,
rltv_hum, rltv_hum_j, snow_depth, snow_depth_j);
```

Listing 3.1: *Example of importing data on MySQL*

As previously mentioned, in this study MatLab® has been used as the main software for all the numerical calculations. Therefore, a script was written to allow communication between the dedicated server which runs on Unix and desktop users. The example below illustrates the necessary actions for a user to run a query which fetches the hourly records from the raw data for a given station.

```
function results=dbconnection_server()

dbUsername = '*****'; %superuser for the db
dbPassword = '*****'; %pwd associated with the above su

SQLquery = 'SELECT YEAR(dated) AS YearAverage,
MONTH(dated) AS MonthAverage,
DAY(dated) AS DayAverage,
HOUR(dated) AS HourAverage,
mwd,
(mws) * 0.515 AS mwsIn_ms
FROM WM
WHERE src_id=9 and MINUTE(dated)=00
GROUP BY YearAverage, MonthAverage, DayAverage, HourAverage';

conn = database('myodbc',dbUsername,dbPassword); %connects to the database

results = fetch(conn, SQLquery);

if (isempty(results))
    errordlg('The table is empty')

end

close(conn);
```

Listing 3.2: *Hourly query to MySQL*

If, however, the monthly records are required from a specific station for a specific period then one should run the following script that includes the right MySQL query. This was very crucial in the present study as different parts of the research necessitated different periods from the records. For example, the part of the wind indices required both monthly and yearly data as well as seasonal. On the other hand, the forecasting part of the research required monthly data for generating the predictions.

```
function results=dbconnection_server()

dbUsername = '*****'; %superuser for the db
dbPassword = '*****'; %pwd associated with the above su

SQLquery = 'SELECT YEAR(dated) AS YearAverage,
MONTH(dated) AS MonthAverage,
AVG(mwd),
AVG(mws*0.515) AS mws
FROM WM
WHERE src_id = 1006 AND MINUTE(dated)=00
AND YEAR(dated)≥1983 AND YEAR(dated)≤2007
GROUP BY YEAR(dated), MONTH(dated)';

conn = database('myodbc',dbUsername,dbPassword); %connects to the database

results = fetch(conn, SQLquery);

if (isempty(results))
    errordlg('The table is empty')

end

close(conn);
```

Listing 3.3: *Monthly query to MySQL*

3.5.2.1 Ante-process Analysis

3.5.2.2 Data Transformation

A critical step towards the creation of the database included restructuring the data in such a format which would be appropriate for the purpose of this study. As mentioned at the beginning of section 3.4 the raw data had to be transformed and in a way to be reshaped in order to match the basic principles of an RDBS. The following code illustrates how the ERA-40 dataset was converted so that all data were grouped in a format that would allow relations between them to be presented in a logical tabular form. As the data were initially downloaded in a txt format, the VBA language which is available in Microsoft Excel was chosen to reshape them.

```
Sub convert()
'
Dim datefield As String
Dim daterow, i, j, activerow As Long
activerow = 1

Worksheets.Add after:=ActiveSheet

For daterow = 1 To Worksheets(1).Cells(1, 1).End(xlDown).Row - 10 Step 11

datefield = Worksheets(1).Cells(daterow, 1)
For j = 1 To 17
For i = 1 To 10
Worksheets(2).Cells(activerow, 1) = datefield
Worksheets(2).Cells(activerow, 2) = 9 + i
Worksheets(2).Cells(activerow, 3) = -26 + j
Worksheets(2).Cells(activerow, 4) = Worksheets(1).Cells(daterow + i, j)
activerow = activerow + 1
Next
Next
Next

End Sub
```

Listing 3.4: *Restructure of raw data*

3.5.2.3 Directional Sectors

In BADC, information is also recorded and kept regarding the direction of wind speed. Values of "0" indicate periods that wind speed is assumed to be calm and thus no direction or wind speed is recorded. As it is shown in Figure (3.3), 12 directional sectors have been used in order to present information related to wind speed direction. The table below shows the classification and hence the relationship between azimuth degrees and cardinal directions.

Azimuth Degrees	Cardinal Directions
345° to 15°	N
15° to 45°	NNE
45° to 75°	ENE
75° to 105°	E
105° to 135°	ESE
135° to 165°	SSE
165° to 195°	S
195° to 225°	SSW
225° to 255°	WSW
255° to 285°	W
285° to 315°	WNW
315° to 345°	NWN

Table 3.8: *Classification of wind speed direction*

This classification caused two problems which had to be tackled. Firstly, one should be able to distinguish the difference between "North" and "calm". This was achieved by running the following query.

```
CREATE TABLE DirWind ADD generic_wind_dir VARCHAR(3),
ADD INDEX (generic_wind_dir);

UPDATE DirWind
SET generic_wind_dir='N'
WHERE wind_dir
BETWEEN 345.001 AND 360
OR wind_dir BETWEEN 0.001 AND 15;
UPDATE DirWind
SET generic_wind_dir='NNE'
WHERE wind_dir
BETWEEN 15.001 AND 45;
UPDATE DirWind
SET generic_wind_dir='ENE'
WHERE wind_dir
BETWEEN 45.001 AND 75;
UPDATE DirWind
SET generic_wind_dir='E'
WHERE wind_dir
BETWEEN 75.001 AND 105;
UPDATE DirWind
SET generic_wind_dir='ESE'
```

```
WHERE wind_dir
BETWEEN 105.001 AND 135;
UPDATE DirWind
SET generic_wind_dir='SSE'
WHERE wind_dir
BETWEEN 135.001 AND 165;
UPDATE DirWind
SET generic_wind_dir='S'
WHERE wind_dir
BETWEEN 165.001 AND 195;
UPDATE DirWind
SET generic_wind_dir='SSW'
WHERE wind_dir
BETWEEN 195.001 AND 225;
UPDATE DirWind
SET generic_wind_dir='WSW'
WHERE wind_dir
BETWEEN 225.001 AND 255;
UPDATE DirWind
SET generic_wind_dir='W'
WHERE wind_dir
BETWEEN 255.001 AND 285;
UPDATE DirWind
SET generic_wind_dir='WNW'
WHERE wind_dir
BETWEEN 285.001 AND 315;
UPDATE DirWind
SET generic_wind_dir='NWN'
WHERE wind_dir
BETWEEN 315.001 AND 345;
```

Listing 3.5: *Creating Table for Directions*

The aforementioned query allowed to create a new table which would contain the cardinal directions by excluding the value ”0”. The name of the table ”DirWind” and the field ”generic_wind_dir” serve as an example and they do not, by any means, reflect the names for the tables/fields used in the present study. The new table that included the cardinal directions could also be a new column within an existing table and not a separated one. The second challenge with the cardinal directions was to perform aggregate functions either these were to call upon the raw data (hourly) or weekly, monthly and yearly data. It was noted that the azimuth degrees linked to the North would return false results when one would try to use aggregate functions. For example, if there were 2 records, one with a recorded direction of 350° and the other with azimuth degree equal to 13° it would have been difficult to get the average direction between those values. The average function would perform $\frac{(350°+13°)}{2} = \frac{363°}{2} = 181.5°$ which clearly is wrong. The average of these two values, while mathematically correct, is not what happens in reality. A value of 181.5° indicates a direction pointing to the South and not to the North. This problem was overpassed by using the following query.

```
SELECT YEAR(dated) AS YrAvg, MONTH(dated) AS MthAvg, DAY(dated) AS DAvg,
AVG(wind_dir) as TempAvgWind,
 if(TempAvgWind≥360, TempAvgWind+360, TempAvgWind) as AvgWind
FROM WIND
WHERE src_id = ..... AND YEAR(dated) .....
GROUP BY YEAR(dated), MONTH(dated), DAY(dated);
```

Listing 3.6: *Retrieving the correct averages for azimuth degrees between* 345° *and* 15°

The above code corrected the problem. Assume there are two azimuth degrees. One is equal to 345° and the other equal to 15°. The query would run $\frac{(345°+15°+360°)}{2} = \frac{720°}{2} = 360°$ which is North. Similarly, if the azimuth degrees are 350° and 11° then the query would fetch as a result $\frac{(350°+11°+360°)}{2} = \frac{721°}{2} = 360.5°$ which again is correct since the arc between 350° and 360° is equal to 10° while the arc between 0° and 11° is 11°. The average between 10° and 11° is 10.5° which is equivalent to 360.5°.

3.5.2.4 Bi-linear Interpolation

Bi-linear interpolation is a statistical method for determining the value of a point when its corresponding values on the x- and y-axis are unknown. To perform a bi-linear interpolation someone needs the values of four neighbouring points, as Figure (3.14) illustrates.

These four points (blue) are then averaged on a distance-weighted method in order to estimate the value of the unknown point (red). Bi-linear interpolation is the result of performing linear

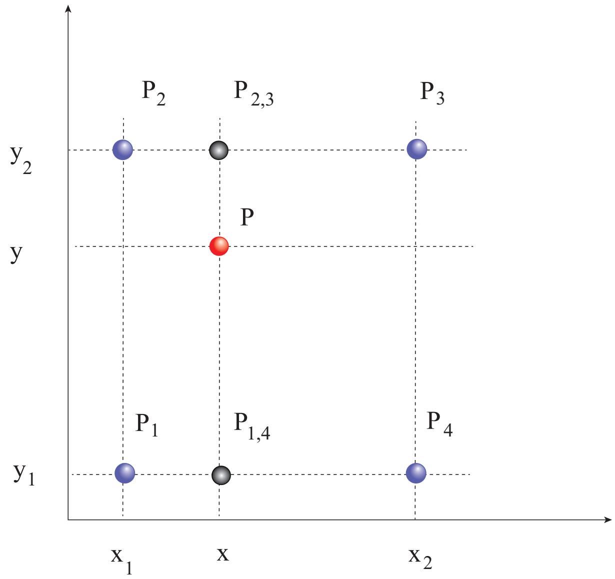


Figure 3.14: *Illustration of a bi-linear interpolation*

interpolations for the x- or y-axis separately and then perform a final linear interpolation between the values of the two previous interpolations (if previously the interpolations took place on the x-axis the final interpolation will be on the y-axis and *vice versa*). The mathematical expression of a bi-linear interpolation is given as follows:

$$\begin{aligned} P_{1,4} &= \tfrac{x_2-x}{x_2-x_1}P_1 + \tfrac{x-x_1}{x_2-x_1}P_4 \\ P_{2,3} &= \tfrac{x_2-x}{x_2-x_1}P_2 + \tfrac{x-x_1}{x_2-x_1}P_3 \\ P &= \tfrac{y_2-y}{y_2-y_1}P_{1,4} + \tfrac{y-y_1}{y_2-y_1}P_{2,3} \end{aligned} \tag{3.3}$$

Where, x_1, x_2, x are the values on the x-axis, y_1, y_2, y the values on the y-axis and $P(x, y)$ are the points in a Cartesian coordinate system.

In this research, as mentioned in section 3.1.2, 10 m values of the zonal and meridional wind components from the $1^\circ \times 1^\circ$ grid have been used. To translate these components to wind speed values firstly the wind speed was calculated by using the equation (3.1) and, afterwards, by performing a bi-linear interpolation. Each cell in the ERA-40 dataset is a square of $1^\circ \times 1^\circ$ grid. The coordinates (x, y) of each station mentioned in Chapter 3 and used in this work are also known. Consequently, by employing the equation (3.3) the corresponding wind speed value for each station from the ERA-40 dataset was calculated.

3.5.3 Data Cleaning Guide

This section acts more as a general guide for handling wind data rather than as a description of a procedure that took place during the present study. The reason why this part is missing from this thesis is the inconsistencies and lack of information about loggers and programs historically used by the Met Office as well as information about the hardware that the Met Office used for storing the records from the stations. Hence it was not feasible to acquire necessary information in order to apply several cleaning rules. Moreover, admittedly it is assumed that the Met Office provide data of high quality. However, this *a priori* argument could add uncertainties on the research itself or, even worse, bias the person who would handle the data. Therefore, this section deals with the extra level of certainty that the cleaning of the data would offer from a theoretical perspective.

3.5.3.1 Icing Events

Icing events could easily knock off instruments especially those that operate with low inertia such as wind vanes. In present this can be avoided by using heated anemometers and wind vanes that prevent the disruption of normal operation. In order for an analyst to capture possible icing event several rules could be set up to flag the data as potential icing events. Nonetheless, it would be prudent the analyst to visually check the data and not to accept the results from the queries/flagging rules. One rule to capture possible icing incidents is as the one shown below.

```sql
SELECT YEAR(dated) AS YrAvg, MONTH(dated) AS MthAvg, DAY(dated) AS DAvg,
AVG(wind_dir) as TempAvgWind,
FROM WIND
WHERE src_id = ..... AND RH BETWEEN 90 AND 95 AND Temp ≤ 3 AND std_ws = 0
GROUP BY YEAR(dated), MONTH(dated), DAY(dated);
```

Listing 3.7: *Detecting Icing Events*

That rule searches the table for a specified station and finds periods where the standard deviation of wind speed is zero, the air temperature is equal or less than 3 °C and the relative humidity is between 90% and 95%. Another useful rule is to check the table for events where for periods greater than 2 hours the wind speed changes less than 0.25 ms^{-1} while the temperature is equal or less than 3 °C. The use of value ”3” for the air temperature is supported by a study published by the International Energy Agency (IEA) where the authors identified two different types of atmospheric icing [140]. One is the *in-cloud icing* and the other is the *precipitation icing*. One form of precipitation icing is the so-called *wet snow*. Wet snow ultimately is partially melted snow crystals that contain high water level which can become sticky and slow down the instruments. Wet snow events may occur when the air temperature is between 0 °C and +3 °C. Therefore, the value of ”3” can be used for running the query. Most importantly, a good practice in flagging icing events is to also include periods up to 5 hours before and after any suspect incident.

3.5.3.2 Instrument Degradation

Another crucial factor to be taken into consideration when relying on records is instrument degradation. Historically and throughout the present days, the instruments have been undergone replacements at frequent intervals by the Met Office. These replacements were accompanied by the appropriate documentation with regards to the installation date, make, serial and height of the new instruments. The stations used in the present study did not include double boom instruments hence it is assumed that no instrument degradation is apparent. However, in case the meteorological masts included double boom instruments then this could be checked by using the ratio of the anemometers for the same height.

3.5.3.3 Battery Voltage

A quick way to identify problems on the data is by checking the battery voltage of the logger. When the voltage battery is less than 12V then it is very likely the data to be classified as erroneous. This is especially for cases that use powered anemometers like the Vector RK. A query for the voltage could flag potential problems and could guide the analyst to be extra careful when checking those periods.

```sql
SELECT YEAR(dated) AS YearAverage,
MONTH(dated) AS MonthAverage,
AVG(mwd),
AVG(mws*0.515) AS mws
FROM WM
WHERE src_id = 54 AND MINUTE(dated)=00
AND bat_volt ≤12
AND YEAR(dated)≥1983 AND YEAR(dated)≤2007
GROUP BY YEAR(dated), MONTH(dated)
```

Listing 3.8: *Detecting Low Battery Voltage*

3.6 Chapter Summary

A detailed description of the data management was presented. Data from various sources were collected including surface measurements as well as reanalysis data. The vast amount of measurements necessitated the use of several filters in order to exclude any erroneous records and increase the reliability of the data. As a follow up step, the remaining stations were examined in order to reveal any factors that might affect the measurements over the long period of records. This, also included sensitivity analyses with regard to data contamination due to subjective errors, site exposure, and changes in both instrument's type and height. The resultant outcome was the organisation of the data using a relational database management system that achieved maximum efficiency and contributed to the completion of this research.

Chapter 4
Wind Speed Variability Across the UK

This chapter contains identical parts from a published paper (conference) [118] and a published paper (journal) [119]

This chapter describes the methodology of calculating a national and regional wind index for the UK for two different periods and two different groups of data. The calculations have used surface data of BADC-7 and BADC-57 meteorological stations as it was explained in full details in the previous Chapter. When both indices are calculated the standard deviation is found to be equal to 4% for both the 29-year period and the 55-year period respectively. These values in turn can be translated to a change in average UK wind power capacity factor equal to 7% for each case. As mentioned in Chapter 3, several factors have been taken into account during the analysis such as site exposure, instrument bias and change in the effective height in instruments. When the generated indices are compared to ones from other sources it is found that the inter-annual variability is similar to high observed correlation coefficients. The indices presented in this chapter have also taken into account the CLC codes for the CORINE dataset that was explained in section 3.3.2.2. The results show slightly steeper decline in wind speeds comparing to the indices that used a roughness length of 0.03 m but overall the same trends are observed. Last, when regional differences in the index are investigated, it is seen that wind speeds show a very slight decline across the UK in all regions except the South-East, which shows a slight increase. It is noteworthy that the greatest decrease is seen in the North-West. This is consistent with the tentative predictions given by climate models for future changes in wind speed across the UK, though the uncertainty is high given the large degree of inter-annual variation.

4.1 Introduction

An understanding of variation in long-term wind speed is important in several sectors such as those where wind loading on structures needs to be assessed or those where long-term energy yields for wind farms needs to be evaluated. Knowledge of wind variability is particularly important for wind farm developers and operators to minimise long-term risk due to fluctuations in annual revenue.

A wind index provides an indication of the mean wind speed, usually annually and/or monthly, relative to the calculated long-term mean wind speed [127]. An alternative method is the concept of a wind index based on *significant wind events*, e.g. Mason et al. presented the concept of the wind indices for the purpose of predicting certain oceanic events relevant to fish reproduction [141]. However, an index based on anomalous events is not useful for the prediction of long-term energy yields from wind farms. What is required is a time history of mean wind speed values at regular (say annual) intervals.

By using a wind index, it is possible to estimate the energy production of a wind farm on the basis of a historical record of mean wind speed. The wind index can also provide an indication of long-term trends in mean wind speed and can be used to provide a financial estimate of the impact of any periods of unavailability [142]. Trend information is particularly useful given that a wind farm might have a 20-30 year production lifetime, and therefore, the wind climate might change significantly from that assumed at the planning stage.

The aim of this work was to reconstruct, from surface observations of wind speed, a wind index which would indicate the variability of the wind climate on a timescale at least as long as the expected lifetime of a wind farm site (≥ 25 years) and would allow a regional analysis of wind speed variability across the UK. This was extended (> 50 years) using as long a period as possible from the observations available, whilst still giving a representative UK average. The index was analysed both nationally and regionally and compared with wind indices calculated from other sources including reanalysis data and surface observations interpolated onto a regular grid. A comparison was then made with other studies in this field. Some conclusions are drawn including how well the recent studies in the wind index align with present climate change predictions.

The research reported here is the first known published analysis of a UK wind index using surface station observations for a period greater than 50 years and with an analysis of regional variation. It also provides a contrast between wind indices derived from spatially smoothed datasets, e.g. analysis data, and point values from meteorological stations.

4.2 Background

This section aims to provide the necessary background information for the concept of the wind indices. It was decided this literature review to be separated from the main literature review presented in Chapter 2 as it would make easier to readers to distinguish the two different topics and hence goals set in this research.

4.2.1 Historic Long-Term Wind Speed Trends

There has been a volume of work investigating historic wind speed trends using indirect indications of wind speed, including reanalysis datasets and historic pressure fields. McCabe et al. [143] presented an analysis of six-hourly cyclone activity between 1959 and 1997 from the National Centers for Environmental Prediction (NCEP)-National Center for Atmospheric Research (NCAR) reanalysis. Their analysis suggested a decrease in cyclone frequency but a slight increase in intensity for mid-latitudes (30°N-60°N) and an increase in cyclone frequency and intensity for high latitudes (60°N-90°N). Another study published by Paciorek et al. [144] and looked at the six-hourly NCEP-NCAR reanalysis dataset between 1949 and 1999 produced six winter indices including extreme wind speed. This work suggested an increasing wind speed trend of between 1.5 and 2.5 ms^{-1} per 50 years for the UK over the period. Wang et al. [145] in a previous work, they studied a number of pressure triangles over the North-Eastern Atlantic to infer the geostrophic wind over the period 1874-2007. The results of this study indicated maximum storminess in the North Sea around 1990. A steadily increasing trend in the geostrophic wind was determined in the North-Eastern extent of the region studied, whereas a decline was observed in the Western extent. Over Great Britain, the geostrophic wind was found to be reasonably steady, except during the summer where a slight decline was observed. For the North-Eastern Atlantic region as a whole, it was found that during the summer, storminess appeared to have declined, except that the South-Western areas (including North-West Scotland) showed no noticeable trends. Bärring and Fortuniak [146] undertook a similar analysis of pressure data from Sweden for the period 1780-2005 using a Eulerian framework. The authors concluded that no overall change in storminess was observed, although significant decadal swings were observed. This work highlighted the importance of studying a long enough period when trying to determine climate changes. Atkinson et al. [35] presented a study of wind speed trends in North-Western Europe and concluded that there was a reasonable degree of correlation between indices in the UK, Germany, Denmark and the Netherlands with a similar declining trend over the 15 year period of 1990-2005. Pryor and Barthelmie [147] published a study of 850 hPa winds from the NCEP-NCAR reanalysis dataset, focussing primarily on the Baltic. Their research showed for the grid cell at 55°N, 5°E, that the time series displayed a number of features, namely a peak in 1967, an increase during 1970s and 1980s, a low around 1987 and a declining trend during the 1990s.

4.2.2 Future Projected Trends

There is an increasing interest in projected changes to the wind climate over the next century, and a number of authors have studied the output from several GCMs and RCMs under different CO_2 forcing scenarios. The output of the ECHAM5 GCM for several climate change scenarios was studied by Pinto et al. [148], indicating an increase in extreme winds with higher wind speeds over Britain, the North Sea, the Baltic Sea and nearby coastal regions during the 21^{st} century. In a previous study presented by Pryor et al. [149], when Hadley Centre Coupled Model version 3 (HadCM3) output data were examined under the Special Report on Emissions Scenarios (SRES) A2 high-emissions scenario, it was found that there was a high degree of correlation of latitudinally integrated wind indices in Europe, balanced around 45°N. This research also found that there was evidence for a slightly reduced annual cycle amplitude European wide, comparing 1990-2001 with 2088-2099. It was concluded by Pryor et al. [149] that although there is generally little evidence for future changes in spatial or temporal variability of wind indices, this is uncertain because of model biases. In the work of Cradden et al. [150] future projections for wind speed are shown comparing HadCM3 and ECHAM5 for several climate change scenarios. Both models indicated decreasing wind speeds during the summer and increasing during the winter by the end of the 21^{st} century. However, this work also highlighted significant uncertainties and discrepancies between the models.

Leckenbusch and Ulbrich [151] analysed the HadCM3 output data, considering the SRES A2A and B2A scenarios over the period 1990-2099. This study showed projected increasing winter (Oct-Mar) 10 m wind speeds over South-Eastern England and slightly decreasing wind speeds over the North-West. However, in the same research, it was found that the regional HadRM3H model output gave a slightly different pattern with an increase to the North-East and a slight decrease in the central and southern UK. Although differences are seen for the different climate change scenarios, research conducted by Nolan et al. [152] which looked at the Rossby Centre's RCA3 RCM suggests projected future change in 60 m wind speeds across the British Isles for the period 2021-2060 with wind speeds increasing in the South-East and decreasing in the North of Scotland under the A2 scenario, albeit with some localised inhomogeneities across this South-North gradient.

Donat et al. [153] presented an analysis of nine simulations from six GCMs under the SRES A1B scenario which showed a projected future increase in storms in the Eastern Atlantic, near the British Isles and in the North Sea. The research also suggested an associated increase in storm intensity over large parts of central Europe towards the end of the 21^{st} century.

Beniston et al. [154] simulated the percentage change in the 90^{th} percentile of winter (Dec-Feb) daily maximum wind speed in Europe, between the 1961-1990 and 2071-2100 periods by using the Rossby Centre's coupled atmosphere-ocean model RCAO. The results from the simulation showed a projected gradient of South-East to North-West with an increase in South-East Europe (5-10%) and a much lower increase in the North-West (0-2.5%). A further RCM analysis presented by Rockel and Woth [155] showed a projected change in the total number of storm peaks from 1961-1990 to 2071-2100, as simulated by the ETH Zurich's Climate High Resolution Model (CHRM) and the Climate Limited area Modelling Community's COSMO-CLM model with a decline in the North-West and an increase in the South-Eastern UK.

Recent UK climate projects have considered a range of meteorological variables, although projected changes to wind speeds over the UK are particularly uncertain. Nevertheless, the UK Meteorological (Met.) Office Hadley Centre has produced two reports that are relevant in this regard. The first published by Brown et al. [156] suggests that there is evidence for an overall slight reduction in wind speeds over the UK, but largest in the North-West with possibly a slight increase in the South-East, most pronounced in the summer, by 2070-2099. The report does, however, note an RCM bias to lower than observed wind speeds in Scotland and Wales and higher in low-lying regions of England. The second report published by Sexton and Murphy [157] suggests that averaged over the entire UK, there is expected to be a small reduction in mean wind speeds, although regional differences are not so clear. However, there is some evidence for a slight reduction in Scotland and a smaller reduction or no change in southern England by 2070-2099. An analysis of the UK Climate Impacts Programme (UKCIP)02 climate change scenarios conducted by Harrison et al. [158] suggested, by 2080, significant reductions in wind speeds over Northern Ireland particularly in the summer and smaller increases in Northern Scotland most notably in late spring/early summer. In England and Wales, wind speeds were projected to slightly reduce in the autumn and increase in winter.

4.2.3 A Comparison of Wind Indices

With the growth in wind power worldwide, there has been an increased interest in assessing inter-annual variation in temporal *windiness* of a region for the purposes of evaluating variability in wind energy generation. This is important from an economic viewpoint to know the variation in likely annual revenue from a wind farm over its lifetime. Wind indices fall into four broad categories:

- those derived directly from surface wind speed observations (e.g. Watson and Kritharas [118, 119], Früh [159] and Hodgetts [34]);
- those derived from observations of wind energy generation, e.g. the Danish wind energy index published by Nielsen [160], the German Ingenieurwerkstatt für Energietechnik (IWET) index and the Statistics Netherlands (CBS) Windex-CBS in the Netherlands;
- those derived from pressure gradients in the form of geostrophic winds triangulated from site pressure observations (e.g. Wang et al. [145], Bärring and Fortuniak [146] and Bakker and van den Hurk [161]); and
- those derived from NWP models such as reanalysis data, with or without regional downscaling (e.g. Pryor and Barthelmie [147, 149]).

The first type of index relates directly to observations of surface wind speed and is thus capable of indicating variations in regional wind climate. The disadvantage of this type of index is the sensitivity to very localised effects. The second category has similar benefits and also relates directly to wind energy. However, changes over time of the portfolio of turbines used to generate such indices and non-availability of machines due to outages can create inhomogeneities [162]. The third and fourth categories have the advantage over the first two of filtering out most local-scale anomalies due to factors such as local microclimates, local orography, changes in site exposure, changes in instrumentation and so on. On the other hand, using such spatially smoothed datasets can also mask local and regional differences in wind variability. RCMs can be used to try and downscale reanalysis datasets, but it has been found that the datasets so derived significantly underestimate inter-annual variability [163]. In addition, such models do not capture the variable spatial characteristics of the wind [164–166].

Differences in surface wind speeds and large-scale circulations have been reported by Vautard et al. [29] and much of the difference has been attributed to changes in surface roughness. Wever used site-specific gust factors of meteorological stations in the Netherlands to assess the impact of surface roughness on long-term wind speed trends [167]. In this study an approximately equal influence of climate, large-scale (mesoscale) surface roughness changes and local roughness changes was suggested. It is possible to distinguish to some extent the influence of these factors by comparing trends inferred from different indices such as in the research published by Bakker et al. [162]. For example, indices inferred from pressure gradients and numerical weather prediction models are relatively insensitive to changes in surface roughness.

4.3 Data

The construction of the wind indices necessitated the use of both BADC-57 and BADC-7 stations as well as data from the ERA-40 dataset as it is amply described in Chapter 3. However, for testing the different indices against other studies data from other sources were required. It was therefore decided to separate the data sources into two categories. One for describing the data used in this thesis in order to calculate the wind indices and for generating the monthly predictions and another one for presenting the data used for comparing previous studies with the one presented herein. Thus the following sections are dedicated to describe the data which were used for comparing different studies and hence are missing from Chapter 3.

4.3.1 UKCIP Met. Office Gridded Dataset

Monthly and annual averages of a range of meteorological parameters, initially for the period 1961–2001, were generated by the UK Met Office using data observed from all of the available surface stations, with gaps in data filled to avoid biasing the results. Gap filling was carried out by spatially interpolating the missing data at a site from the six closest stations for 13 of the meteorological parameters, although not wind speed [168]. In the case of wind speed, some data substitution was carried out, but the details of this are not reported [169]. The resulting dataset was produced for the UKCIP to provide a consistent time series of climatological variables that could be used in long-term climate studies. The data have been interpolated onto a 5km x 5km grid using spatial interpolation and multiple linear regression, taking into account a number of parameters including easting/northing, altitude, proximity to the coast and local urbanisation. Errors in interpolating to this grid were assessed, and for monthly mean wind speed, the root mean squared error was found to be 5.5 kt (2.8ms^{-1}). It should be noted that time-varying urban effects were not considered. No mention is made in the work of Perry and Hollis [168, 169] of any correction made for time-varying changes in instrument height. Station openings and closures over the period used to generate this climatological dataset meant that the number of stations used to infer the gridded climatological data has changed over the period. This will have some implications for continuity of the wind speed data. The average number of stations available for the generation of the wind speed gridded data over the period was 70. The present study has used the mean monthly wind speed values that were available for the period 1969–2006.

4.3.2 The Garrad Hassan Wind Index

The renewable energy engineering consultancy company Garrad Hassan (GH, now known as DNV GL-Garrad Hassan) formerly published a UK wind index. When compared with other European indices, this showed a reasonable level of agreement [170]. Although the methodology behind this index and the exact list of stations used to generate the index were not publicised, it was stated by GH that this index was calculated using data from 50 meteorological stations spread throughout the UK [34]. The index used in this work was a 13 year index normalised over the period 1995–2007.

4.4 Calculation of the Different Indices

A wind index, I_j, is normally defined as the average wind speed over a region for a given averaging period divided by the overall average wind speed over that region over the normalisation period of interest, i.e.:

$$I_j = \frac{\sum\limits_{k=1}^{n} \sum\limits_{i=1}^{m_j} U_{i,j,k}}{\sum\limits_{j=1}^{l} \sum\limits_{k=1}^{n} \sum\limits_{i=1}^{m_j} U_{i,j,k}} \tag{4.1}$$

Where $U_{i,j,s}$ is the wind speed at station (or grid point) k and time i within averaging period j. The value m_j depends on the number of values in the averaging period j and the number of the sites (or grid points) used to generate the index is n. The normalisation period consists of l averaging periods. For example, if an annual index is being generated over a period of l years, then m_j would be the number of hours in year j.

Indices were calculated using equation (4.1) from:

1. the hourly wind speed values from the BADC-7 and BADC-57 stations;
2. six-hourly wind speed values blinearly interpolated to the BADC-7 and BADC-57 stations from the $1^\circ \times 1^\circ$ u' and v' 10m wind components of the ERA-40 reanalysis dataset;
3. monthly wind speed values at the 5km x 5km grid points covering the UK from the UKCIP Met. Office gridded dataset;
4. the hourly wind speed values from the BADC-7 stations for values of roughness length from CLC known henceforth as CLC-7 and CLC-57 respectively.

Table (4.1) summarises the parameters used in equation (4.1) for the generation of the different indices. The BADC indices were calculated over several normalisation periods for comparison with the indices calculated using the ERA-40 reanalysis data and the UKCIP Met Office gridded dataset. In the case of the GH index this was pre-calculated. As mentioned previously, details of how this was done are not published, but given that it is based on observed hourly values, the calculation were done using an equation similar to equation (4.1) with a normalisation period l =13 years.

Index	i	j	k	l years	m_j
BADC-7	Hourly values	Year	7 (stations)	38, 44, 55	Hours in year
BADC-57	Hourly values	Year	57 (stations)	13, 19, 29	Hours in year
ERA-40	Six-hourly values	Year	Bi-linear interpolation to either BADC-7 or BADC-57 stations	44	Six-hourly values in year
UKCIP	Monthly values	Year	All grid points covering UK	38	Months in year
CLC-7	Hourly values	Year	7 (stations)	55	Hours in year
CLC-57	Hourly values	Year	57 (stations)	29	Hours in year

Table 4.1: *Summary of the parameters used in equation* (4.1) *for calculation of the different indices*

4.5 Results

4.5.1 The Annual 29-year and 55-year UK Wind Indices

Figure (4.1) shows the 29-year and 55-year wind indices based on the BADC-57 and BADC-7 stations respectively. The grey shading represents one standard deviation of uncertainty (also known as error bars) calculated from the hourly values and standard least squares linear trend fits to each series are shown with 95% confidence limits. The values of each index are tabulated in Appendix (C). In the case of the BADC-7 index, the value is tabulated normalised to the full 55-year period and the 29-year period concurrent with the 29-year index derived from the BADC-57 stations.

There is a good degree of agreement between the two indices for the period 1983-2011 with a Pearson correlation coefficient of 0.89. Averaged over the longer 55-year period, there does not seem to be any significant change in annual mean wind speeds, though there is a slight decrease when averaged over the shorter 29-year period. It can be seen that from the late 1950s until the late 1960s, there was a significant increase in annual mean wind speeds followed by a rapid decrease and another increase throughout the 1970s into the early 1980s. Between the

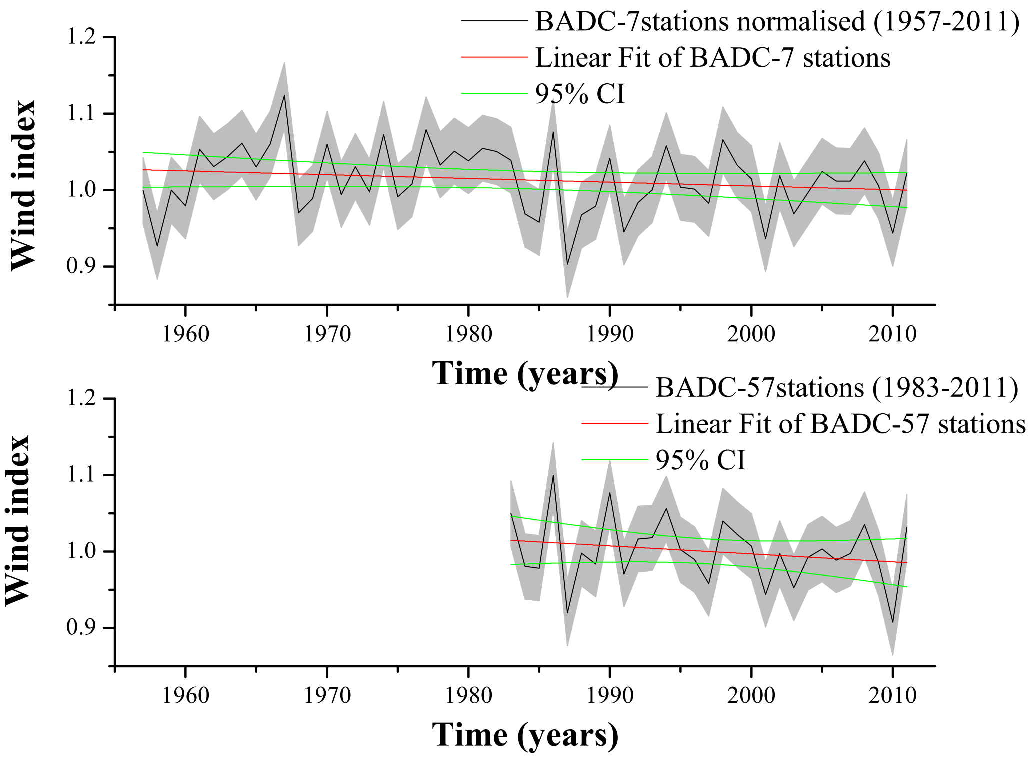


Figure 4.1: *The 29-year and 55-year wind indices with linear trend line fits to each series*

mid-1980s and 2011, there have been some significant low wind speed years including 1987, 2001 and 2010. Although 2010 was the lowest wind speed year over the 29-year period, it was not unusual when compared with the 55-year period, an observation consistent with other references such as the studies published by Fruh [159] and by Ebsworth et al. [171].

Inclusion of data for periods such as the one in 29-year analysis and even shorter may introduce bias to the long-term estimates. This is likely to happen especially when over the period of the analysis unprecedented periods exist of high or low wind speeds that consequently are above or below the usual variation in long-term wind speed. This confirms the necessity to include a considerable period of data such as the one used in the 55-year analysis. Over both periods, the standard deviation of the two indices is 0.04.

Figure (4.2) shows a comparison between the BADC-7 station wind index, this time calculated over the shorter normalisation period 1958-2001 for comparison with the wind index calculated using the ERA-40 reanalysis data interpolated to the locations of the seven stations over the period 1957-2001. Once again, a linear trend line is fitted to both indices over the period 1958-2001. The two indices show similar behaviour, though in the latter half of the period, there is some difference. The ERA-40 index shows a greater declining trend than the BADC-7 station index.

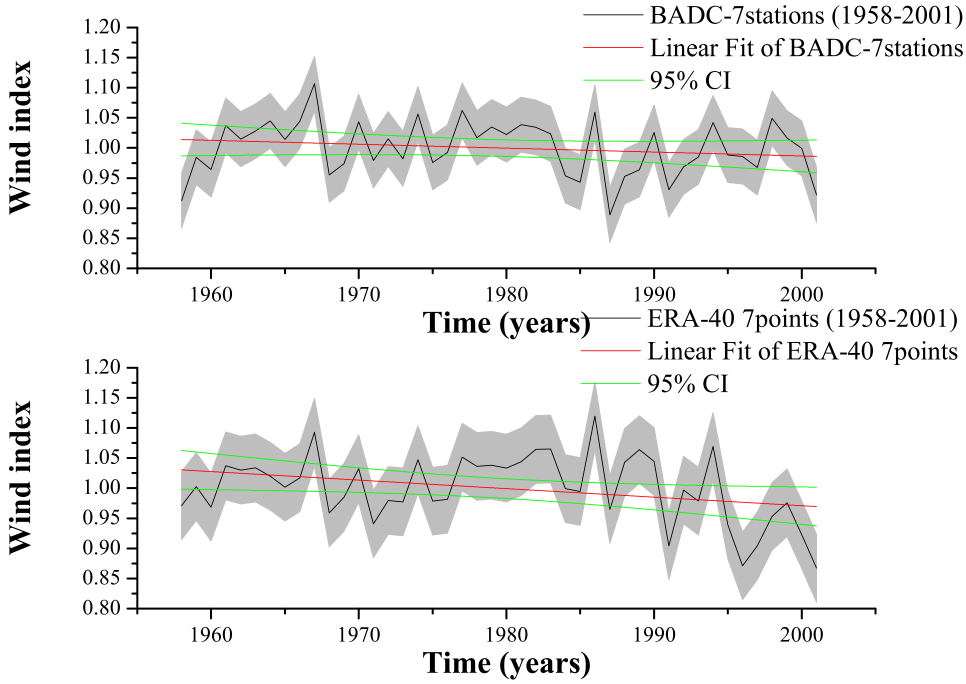


Figure 4.2: *A comparison between a wind index calculated using the BADC-7 stations and ERA-40 data interpolated to the same sites over the period 1958-2001*

Figure (4.3) shows a similar comparison, this time between the index calculated using the BADC-57 stations and one using the ERA-40 data interpolated to the locations of the 57 stations over the period 1983-2001. In this case, the difference in the trend over the period is more marked with the ERA-40 data showing a steeper decrease.

Figure (4.4) shows a comparison between the index calculated using the BADC-7 stations and the one generated using the UKCIP Met Office gridded dataset for the period 1969-2006. The agreement is relatively good, as would be expected given that the UKCIP dataset was generated using observed surface station wind speed data, though there is a more steeply declining trend in the latter index. This also might be expected given that the UKCIP dataset includes data from a large number of stations including more urbanised stations whose exposure is likely to have changed over time with increasing shelter more likely.

Figure (4.5) shows a comparison between the index calculated using the BADC-57 stations and the GH index over the period 1995-2007. Here agreement is good but with a more sharply

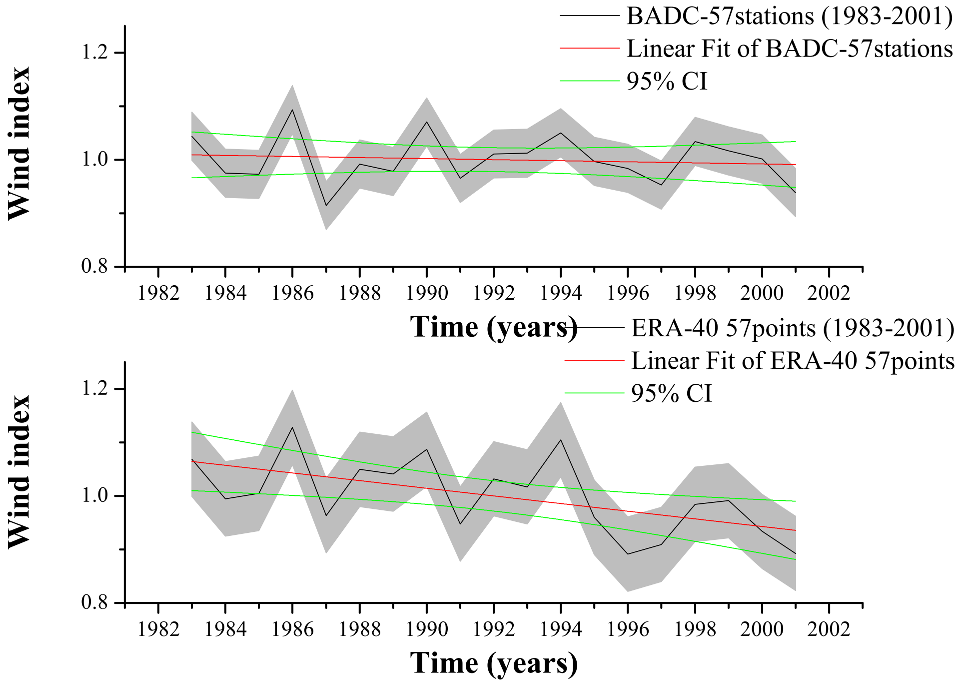


Figure 4.3: *A comparison between a wind index calculated using the BADC-57 stations and ERA-40 data interpolated to the same sites over the period 1958-2001*

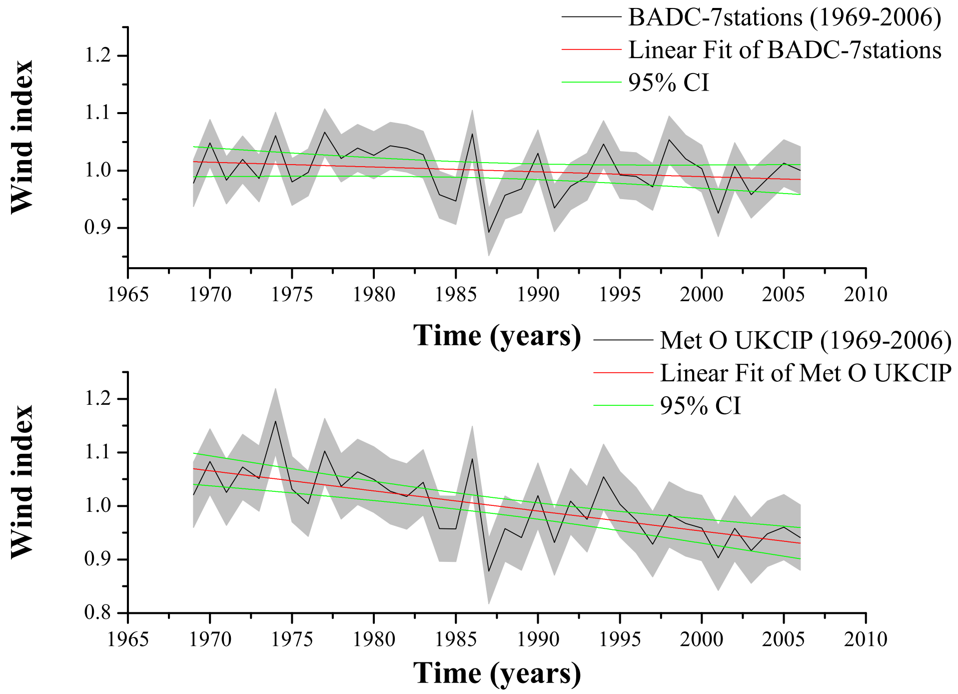


Figure 4.4: *A comparison between a wind index calculated using the BADC-57 stations and ERA-40 data interpolated to the same sites over the period 1958-2001*

declining trend for the GH index.

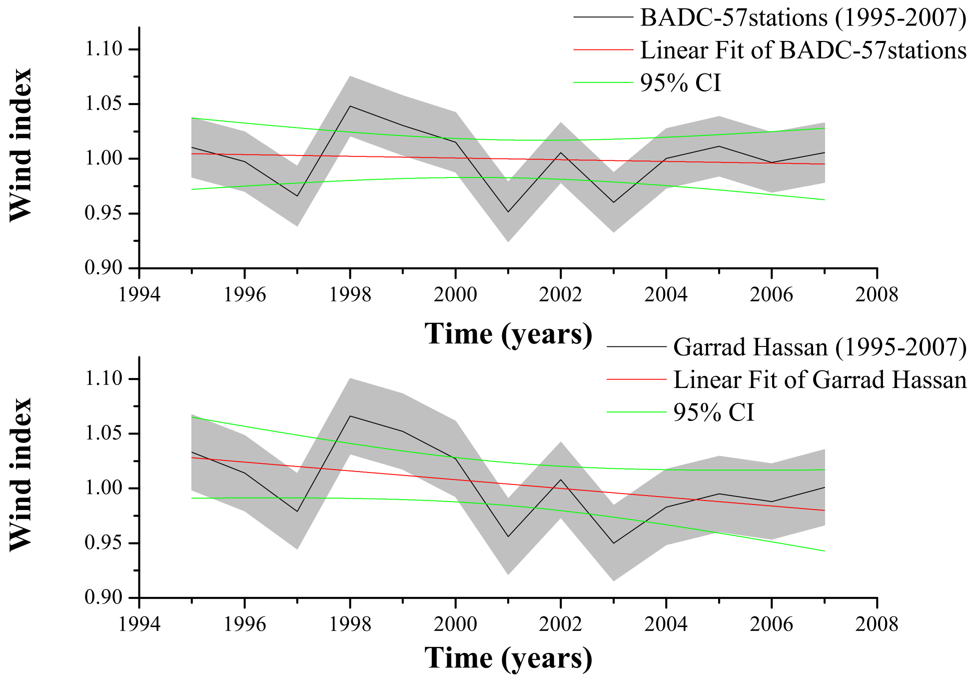


Figure 4.5: *A comparison between a wind index calculated using the BADC-57 stations and the GH index over the period 1995-2007*

Finally, Figures (4.6) and (4.7) depict a comparison between the indices shown in Figure (3.14) and the indices of the BADC-7 and BADC-57 stations based on the CLC roughness classification, as mentioned in section 3.3.2.2.

The initials *"W&K"* on the above indices in each figure correspond to the authors Watson and Kritharas [118, 119] and refer to BADC-7 and BADC-57 indices. There is a good degree of agreement between *"W&K"* and *"CLC"* indices with a Pearson correlation coefficient of 0.942 and 0.856 for the periods 1957-2011 (Figure 4.5) and 1983-2011 (Figure 4.6) respectively. The most sharply declining trend is observed for the longer period where most of the sites have changed the effective height in instruments.

4.5.2 Correlation between Different Wind Indices

In order to quantify the degree of agreement between the different indices, the Pearson correlation coefficient was calculated using selected combinations of the concurrent annual index values. The results of the analysis are shown in Table (4.2) below.

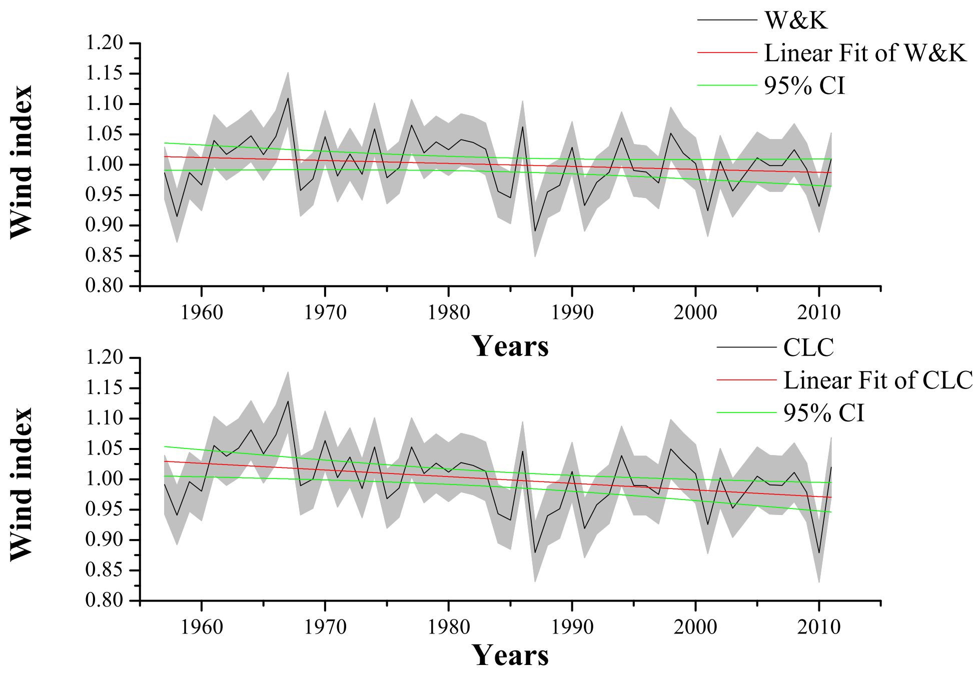


Figure 4.6: *A comparison between a wind index calculated using the BADC-7 stations and the same index based on CLC classification*

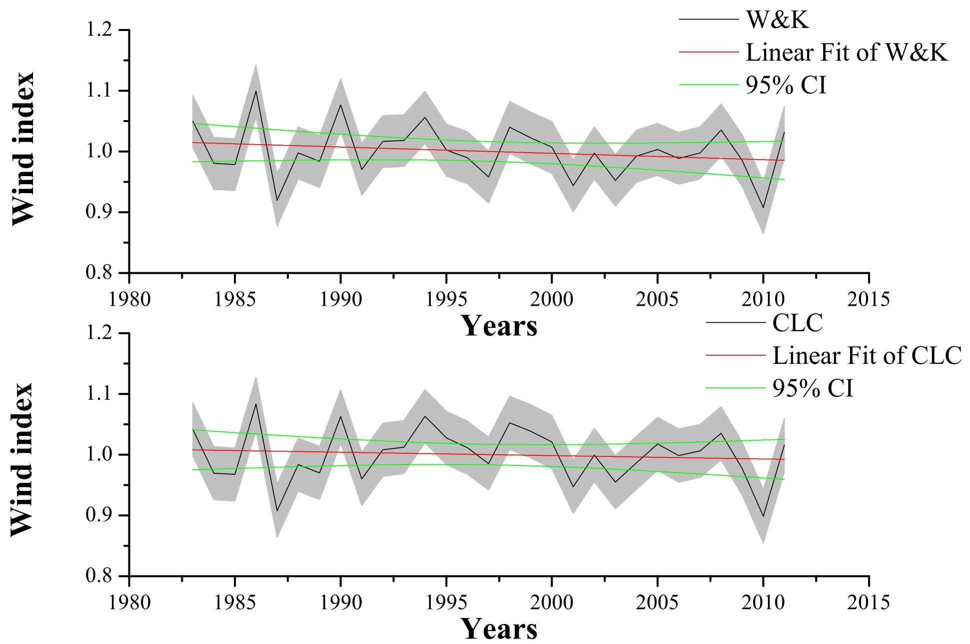


Figure 4.7: *A comparison between a wind index calculated using the BADC-57 stations and the same index based on CLC classification*

Index 1	Index 2	Concurrent years	Correlation coefficient
BADC-7	BADC-57	29	0.887
BADC-7	ERA-40 (UK)	44	0.629
BADC-7	UKCIP	38	0.781
BADC-7 ("*W&K*" in Figure 4.6)	CLC	57	0.942
BADC-57	ERA-40 (57)	19	0.755
BADC-57	ERA-40 (UK)	44	0.876
BADC-57	GH	13	0.919
BADC-57 ("*W&K*" in Figure 4.7)	CLC	29	0.856

Table 4.2: *Pearson correlation coefficient calculated using concurrent annual values for different indices*

There is a high degree of correlation between the indices calculated using the surface data, i.e. the BADC-7, BADC-57, CLC, and GH indices. The highest correlation coefficients are found between BADC-7 and CLC. This supports the decision to use a value of roughness length equal to 0.03 m. Figure (3.8) in section 3.3.2.1 shows that the difference between wind indices calculated for different values of roughness length is negligible. Though, the range of the tested roughness lengths varied between 0.01 m and 0.08 m. Here, it is shown that when the corresponding values are taken into account for roughness length (Table 3.3.2.2) of the stations used in the study, the difference in the results remains undetectable. The most striking finding though is the high correlation between BADC-57 and the index provided by GH. This agreement is of great importance for both industry and academic community. It provides evidence that it is feasible to calculate a cost effective wind index which would give an indication about the long-term mean wind speed for the UK from surface stations. The correlation between the BADC-7 and ERA-40 index is somewhat lower. The BADC-7 index is calculated using a relatively small number of point observations whereas the ERA-40 index is more spatially homogeneous. In addition, the ERA-40 index is based on six-hourly rather than hourly data, though this should still capture the main features of diurnal variation and is unlikely to introduce bias. This is consistent with the previous observation in that 57 stations will provide a higher degree of spatial smoothing. The correlation between the BADC-7 stations and the UK Met Office gridded dataset lies somewhere in between. Both are generated with surface observations, though the latter would be expected to exhibit a greater degree of spatial smoothing given the much larger number of stations used to generate the 5km x 5km grid.

In order to analyse the degree of spatial smoothing of the BADC-7 stations, the Pearson correlation was calculated between each combination of the seven sites using the annual mean wind speeds at each site. The results of this are shown in Table (4.3).

	Lerwick	Stornoway Airport	Boscombe Down	Valley	Aberporth	Aldergrove	Tiree
Lerwick	1	0.002	-0.277	-0.089	0.186	-0.071	0.168
Stornoway Airport		1	-0.317	0.119	-0.130	0.270	0.241
Boscombe Down			1	0.680	0.409	0.351	0.376
Valley				1	0.441	0.529	0.508
Aberporth					1	0.373	0.547
Aldergrove						1	0.731
Tiree							1

Table 4.3: *Pearson correlation coefficient calculated using annual mean wind speed values for combinations of the BADC-7 stations*

In addition, the correlation is plotted as a function of distance in Figure (4.8)

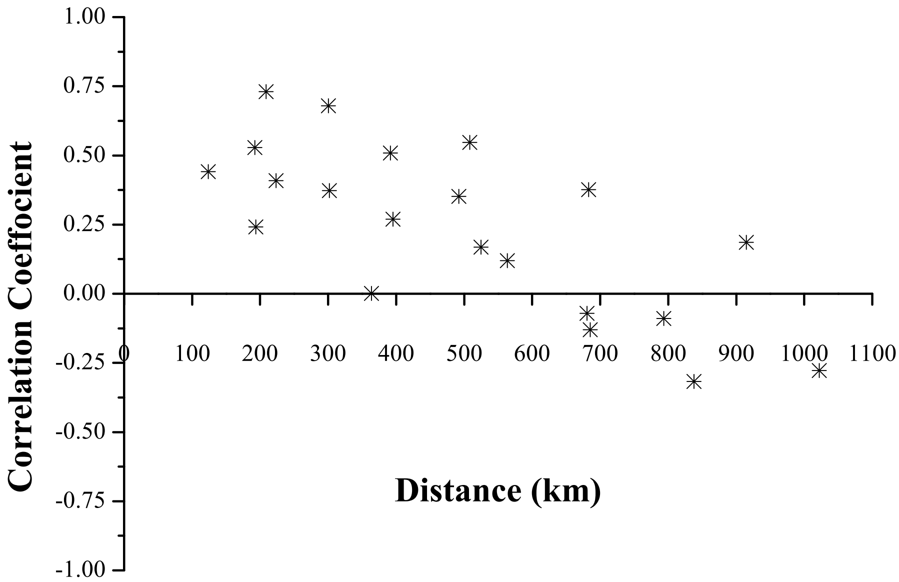


Figure 4.8: *Pearson correlation coefficient calculated using the annual mean wind speeds at the BADC-7 sites as a function of the distance between the different site combinations*

Figure (4.8) confirms an obvious trend with separation distance. What is striking is the high average correlation between the annual mean wind speeds at sites such as Aldergrove and Tiree (0.731) which are 208km apart. This would be extremely useful during the so called *Phase 1* among wind analysts/developers. During Phase 1, a wind analyst has no information about the wind regime (i.e. no actual recorded data) at the prospective site. The usual procedure followed therefore, is to identify similar sites, preferably in the vicinity of the site under investigation, in order to make the initial estimation. However, this may not be feasible and during this phase several assumptions have to be taken which obviously may increase the uncertainty in evaluating a project. The usual reason behind this procedure is that the time required for an application for a mast to be granted or the time until a landowner consents to a developer to install a mast can be crucial from a planning perspective for a future project. One condition is the wind speed distribution to be very alike between the reference sites and the target one. Indeed, as Figure (3.3) illustrates, the predominant wind in Aldergrove is at 210° whereas in Tiree, while it is fairly exposed from all the directional sectors, the predominant direction is 240°. Moreover, the topography between the reference sites and the target station must be the same. This is due to the fact that different terrain will have different local effects which may affect wind speed. Again, when checking the satellite images for the aforementioned sites, no major issue related to the topography of both sites is raised. However, as it is mentioned in Chapter 6, it is highly recommended a study to take place that will assess the impact of the distinct topography of several sites on their between correlation in the long-term wind speeds.

4.5.3 UK Annual Regional Wind Indices

By using the BADC-57 stations, a wind index was calculates by region for the period 1983-2011 using equation (4.1) sub-setting the stations by the six-regions, as denoted in Table (3.2) in section 3.2. Figure (4.9) shows this wind index for the six regions. It can be seen that there is a general trend to decreasing wind speeds in all regions except the South-East which shows slightly increasing trend. The largest decreasing trend is in North-West. When these trends were analysed in more detail, it was found that the greatest declines occurred in the winter months and the smallest in the summer, with the South-East region showing a significant increasing trend during the summer months. This is confirmed by figures (D.1)-(D.4) in Appendix (D). It should be stressed that these trends are tentative given the large degree of inter-annual variation.

4.6 The Effect of Wind Speed Variability on Wind Energy

The theoretical power in the wind varies as the third power of the wind speed, so small changes in wind speed could be expected to translate to rather larger changes in wind power. In fact, a modern MW-sized wind turbine will start to regulate above $\sim 9\text{ms}^{-1}$, so the change in annual energy yield from a turbine will not increase as rapidly as the third power of the wind speed. The capacity factor (CF) of a wind turbine is defined as the average energy produced by a turbine over a representative period, e.g. a year, divided by that which would have been produced had the turbine operated at full output during that entire period. Using a power curve from Vestas

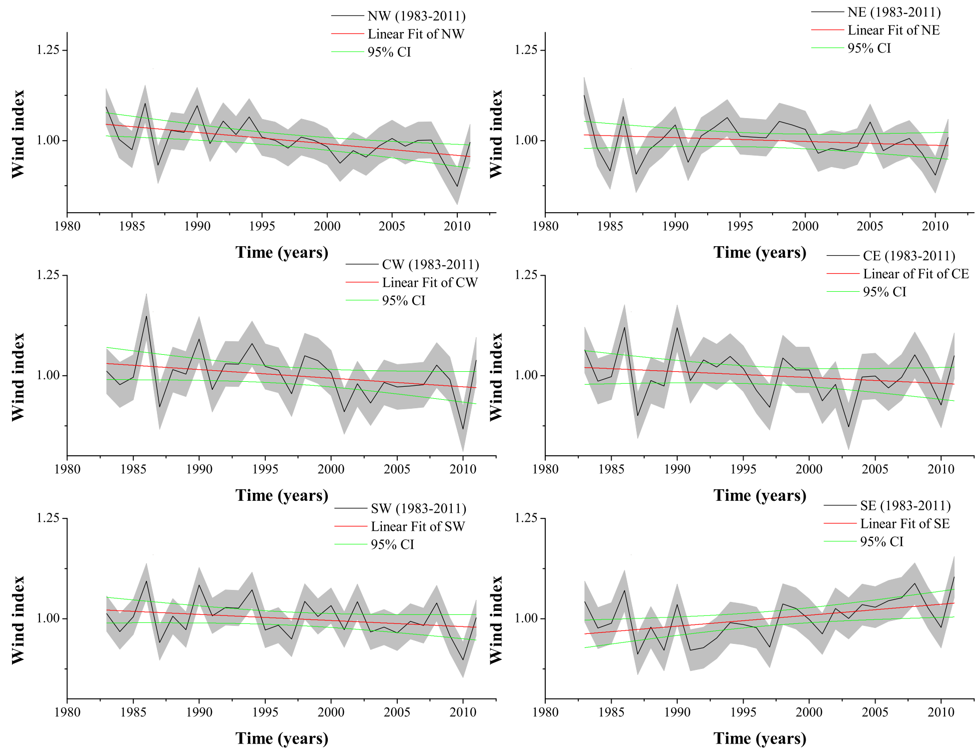


Figure 4.9: *UK wind index by region generated using the BADC-57 stations over the period 1983-2011*

V80 2MW wind turbine, it is possible to calculate the expected CF assuming a given average wind speed at hub height. The CF will also depend on the distribution of wind speeds. As mentioned in Chapter 2, equation (2.2) is used to describe the distribution of wind speeds at a site. Using the Rayleigh distribution (equation 2.3), the CF for the given example of Vestas V80 2MW turbine is calculated and is shown as a function of mean wind speed $\overline{u}$ in Figure (4.10). It can be seen that the CF goes up fairly linearly with mean wind speed until 9ms^{-1} and then reduces more slowly, flattering of near 14ms^{-1}. This is due to a combination of the turbine regulating at 9ms^{-1} reaching rated wind speed at around 14ms^{-1} and shutting down at 25ms^{-1}.

If a mean wind speed of 7 ms^{-1} is assumed at 80m hub height, then a 4% variation (one standard deviation) in wind speed, as seen for the BADC-7 and BADC-57 indices, represents a variation in CF and associated annual energy yield of 7%.

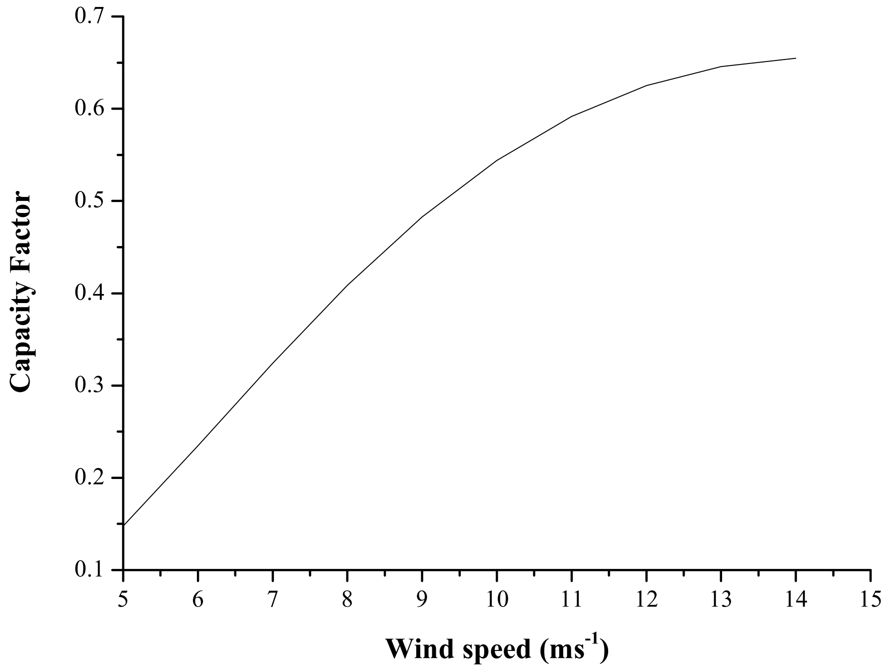


Figure 4.10: *Capacity factor of a Vestas V80 2MW turbine as a function of mean wind speed*

4.7 Chapter Summary

This Chapter has presented two wind indices for the UK based on surface station observations of wind speed: one based on seven stations (BADC-7) over a 55-year period (1957-2011) and a second one based on 57 stations (BADC-57) over a 29-year period (1983-2011). These indices have been compared with indices generated using a gridded dataset of values interpolated from UK stations, three indices calculated using the ERA-40 re-analysis dataset and another UK wind index. The directional wind speed at the BADC-7 stations was examined, and there was found to be no obvious evidence of changes in site exposure or instrument location that could affect the continuity of wind speeds at these stations. There have, however, been notable changes in instrument height at many of the stations included in the two indices, as detailed in Chapter 3, and, as a result, wind speed observations at all stations were corrected to 10 m agl.

The principal findings were:

- The BADC-7 and BADC-57 indices agree reasonably well over the common period 1983-2011 with a correlation coefficient of 0.887;
- The inter-annual variation of the BADC-7 and BADC-57 indices was found to be 4%;

- Given a specific type of wind turbine this variation equates to a standard deviation in wind farm CF of 7%;
- From the late 1950s until the late 1960s, there was an apparent increase in annual mean wind speeds followed by a rapid decrease and another increase throughout the 1970s into the early 1980s. Between the mid-1980s and 2011, there have been some significant low wind speed years, including 1987, 2001 and 2010;
- For the UK as a whole for the BADC-7 index, there is a slight but not significant decline;
- When compared to indices calculated from other sources of data, namely the ERA-40 dataset, the UKCIP Met Office Gridded Dataset and an index created by the renewable energy consultancy GH, similar trends in wind variability are seen, though correlations vary. The most significant correlation is found to be between our index and Garrad Hassan. The correlation between BADC-7 and ERA-40 index is found to be lowest reflecting the difference levels of spatial smoothing. The difference in correlation reflect the different levels of spatial smoothing, different averaging periods and site-specific effects;
- All of the indices show declining trends (between 0.05% and 0.71% per year) with the exception of the ERA-40 index calculated over the UK, which shows an increase (+0.1% per year), although not all of these trends were found to be significant;
- The differences in trends do not suggest that increases in large-area (mesoscale) roughness are a significant factor but that there may be differences in trends between land and sea.
- When the correlation between annual mean wind speeds at the BADC-7 sites is analysed, it is found to be a strong trend with distance;
- Local changes in exposure and changes in measurement height may have had an influence, particularly for the UKCIP index;
- When the BADC-57 index is broken down by region, differences in inter-annual variability are seen across the UK. In addition, there appears to be a general trend to decreasing wind speeds in all regions except the South-East and the largest decreasing trend is observed in the North-West. The greatest declines are seen in the winter months and least in the summer, where the South-East region shows an increasing

trend. However, the uncertainty on these trends is large considering the large degree of inter-annual variation.

Chapter 5
Model Development

This chapter contains parts of three published papers (one journal [99] and two conferences [100, 172])

The organisation of this Chapter reflects chronologically the stages of developing a novel model for predicting monthly wind speed time series. This was achieved by setting the first benchmark on long-term forecasting models. In view of this, several models are presented in this Chapter each of which contributed to the final proposed model. The difference in this research compared to the existing literature is that the model presented herein took also into account external variables apart from the dependent one. Assessing the correlation between different combination of variables indicated that these can be used as additional inputs to the model. As a result, a relationship between monthly mean wind speed over the UK and the independent variables was established. This led to identifying the lagged effects of the independent variables on wind speed. The criteria for fitting the data to the model were critically conducive to the correct setting of its parameters. This guaranteed the ability of the model to capture the characteristics of the time series and produce reliable forecasts. The latter finding was further verified by the low scores in the statistical metrics.

5.1 Data Analysis and Diagnostic Tests

5.1.1 Frequency Spectrum

As discussed in Chapter 1, wind fluctuates and thus, according to the definition given in Chapter 2 (see section 2.1.2.2), wind is considered to be a non stationary random process. This process is expressed by the following equation:

$$u(t) = u_m(t) + u_t(t) \tag{5.1}$$

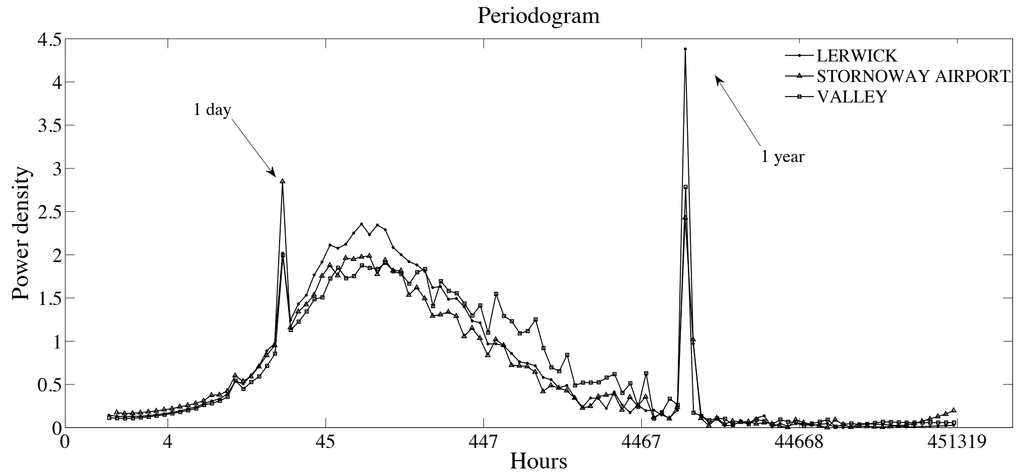
where $u_m(t)$ describes the long-term variations observed in wind speed and $u_t(t)$ is the turbulence component that corresponds to high frequency changes such as seconds. These components and hence these variations were first presented by Van der Hoven [173]. The advantage in using this method is that it can illustrate periodical trends and seasonal characteristics in time series. This was a vital step in this research as the fundamental principle of time series forecasting depends on identifying a pattern in the series and then, based on the history of incidents over time, on forecasting ahead. Generally, a time series pattern discloses sub-patterns such as trends with cyclic or periodic patterns, random or sporadic variations, seasonal and level shifts or, in some cases, a combination of the above sub-patterns. Thus, it was deemed useful to identify any patterns in the wind speed series prior to any analysis.

A Fast Fourier Transform (FFT) was applied to generate a wind speed frequency spectrum. Figures 5.1(a), 5.1(b) and 5.1(c) clearly show the existence of two strong peaks. The first peak appears at 24 hours and the second peak at one year due to diurnal and annual variation in wind speed [173], respectively. In Figures 5.1(a), 5.1(b) and 5.1(c) the stations are split into three cases:

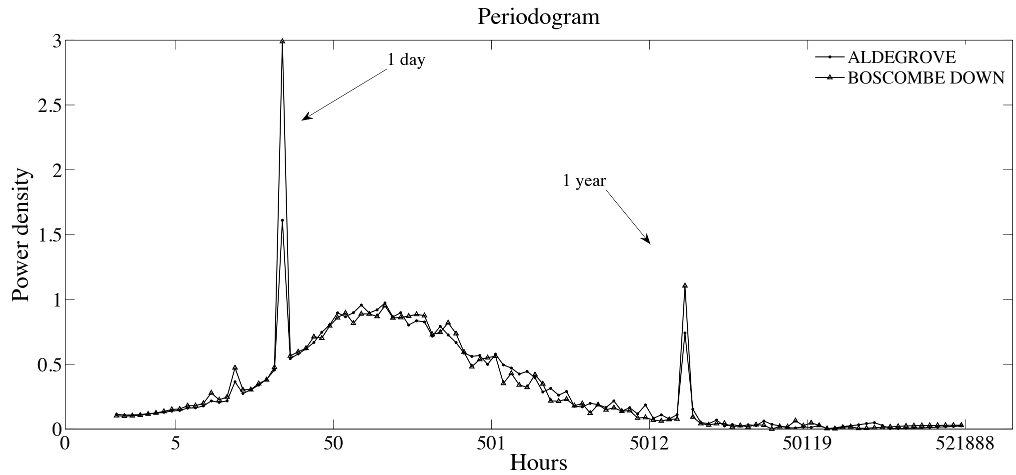
1. Strong diurnal peak and strong annual peak 5.1(a)
2. Strong diurnal peak and weak annual peak 5.1(b), and
3. Weak diurnal peak and strong annual peak 5.1(c).

The results from the FFT analysis showed that all sites exhibit a seasonal component which occurs on an annual basis. This finding was of great importance in order to start conceptualise which model could deliver monthly mean forecasts. From the different classes it was revealed that a model which would consider seasonality would possibly behave well in the testing datasets by producing accurate forecasts.

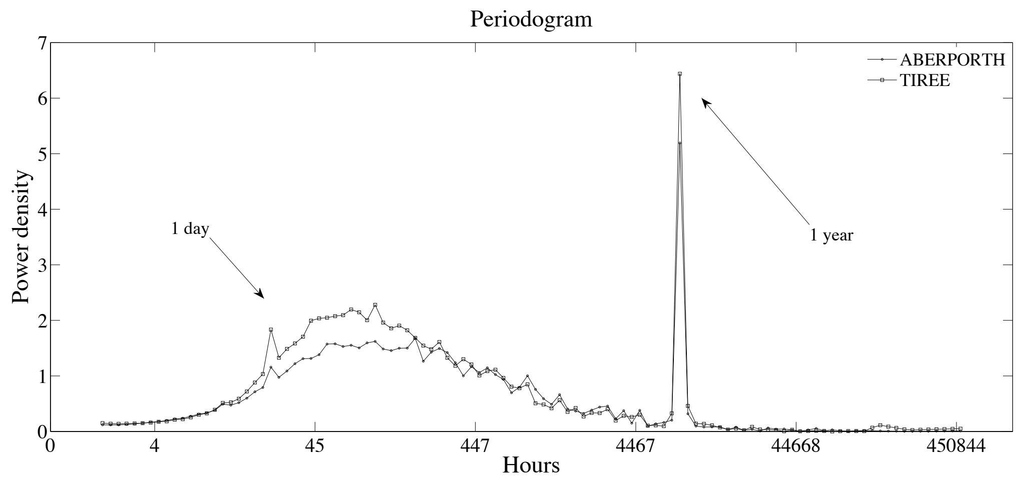
After generating the wind speed spectrum and observing the strong patterns on a daily and yearly basis, a time series analysis was applied on the wind speed data. As Figure 5.2(a) confirms, the



(a) Case 1



(b) Case 2



(c) Case 3

Figure 5.1: *Wind speed power spectrum, at 10-m height agl for BADC-7*

analysis showed that the time series in Case 1 are non stationary. Analysing the monthly mean wind speed averaged over all years per station resulted in the observation of a similar seasonal behaviour for each station (see Figure 5.2(b)). It is clear that during the summer, where the lowest wind speeds occur, and during the winter, where the highest wind speeds have been recorded, all the stations showed similar periodic variations. Similar trends were revealed when the other cases, namely 2 and 3, were examined and thus these figures are omitted.

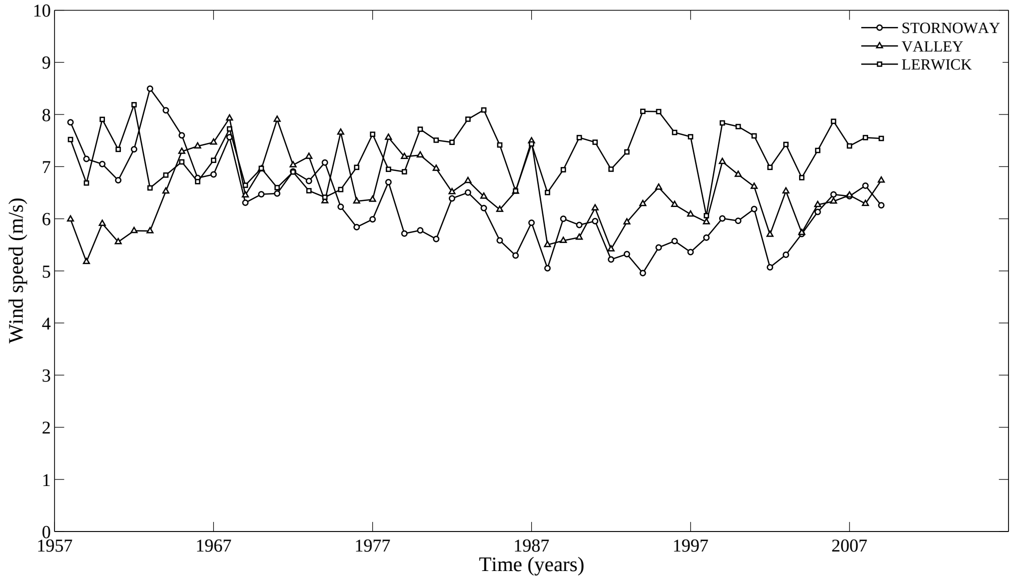
5.1.2 Checking for Stationarity

An important step in forecasting involves the choice of an appropriate model. Plotting the time series as well as their correlograms is informative and provides an initial estimation of the potential model. Nevertheless, prior to the selection of the model, conclusions must be drawn with regard to the process itself. Time series can be stationary or non stationary, and, thus, determining the nature of the process can help simplify some assumptions which in the case of non stationarity cannot be made.

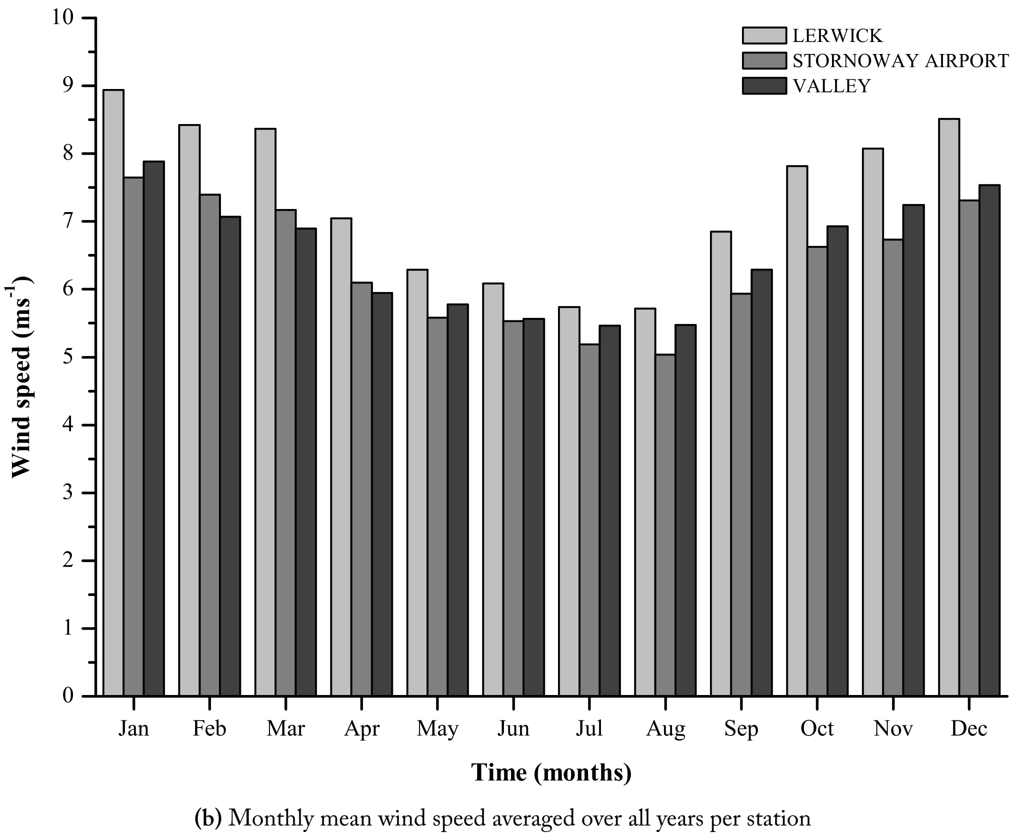
A way to check if time series are stationary is by using the Dickey Fuller (DF) test, as it was published by Dickey and Fuller [174]. In this study all the series were tested for stationarity using the Augmented DF (ADF) test which is a slightly different version of the original test (see [175] for more details). The ADF test is assessed using the "adftest" function from the Econometrics Toolbox™ version 2.0.1 of the commercial package MatLab® (R2011b) [175]. As Table (5.4) indicates, the results from the test showed that all the series are non stationary, since ADF tests the unit root null hypothesis $H_0 : \hat{\psi} = 1$ against the stationary alternative $H_A : \hat{\psi} < 1$.

Stations	$\hat{\psi}$ unit root null hypothesis
Lerwick	1
Stornoway Airport	1
Valley	1
Aldergrove	1
Boscombe Down	1
Aberporth	1
Tiree	1

Table 5.1: *ADF test for stationarity*



(a) Yearly mean wind speed per station



(b) Monthly mean wind speed averaged over all years per station

Figure 5.2: *Seasonal patterns in Case 1 of the BADC-7 stations*

5.2 Model Selection and Fitting

This section presents the various models used in this study in a chronological order. The idea behind this was to use gradually slightly more sophisticated models that while they would reduce the residuals they would avoid overfitting as well. The first model used was the naive predictor, or else the Persistence model, which soon was replaced by a Seasonal Persistence. The reason for replacing the simple Persistence with the seasonal one was based on the findings of the FFT presented in section 5.1.1. Afterwards, the Holt-Winters model was tested and finally models from the SARIMAX family were selected as more suitable. The SARIMAX model firstly used was one that took into account exogenous variables from the same location of the wind data for each station. This was a necessary step as it was vital for the study to assess whether a model that takes into account exogenous meteorological variables can perform well when generating mean monthly wind speed forecasts. However, based on the initial hypothesis it was also required that the exogenous variables would feature spatial and temporal association with wind speed. This led to creating another SARIMAX model by using meteorological variables from the ERA-40 dataset.

5.2.1 Persistence Model

This forecasting model is based on the assumption that the forecasted value at time $t+1$ will be the same as the value at the previous time step t:

$$\hat{y}_{t+1} = y_t \tag{5.2}$$

where $\hat{y}_{t+1}$ is the forecast value at period $t+1$, and
y_t is the actual value at time t.

As mentioned in Chapter 2, this model is required, as it is the benchmark model which every other model must compete with, in order for the forecasters to assess its performance. Figure (5.3) illustrates that the Persistence model fails to capture sudden changes in wind speed variability. The latter is confirmed by the score in the MSE which is found to be 2.197 ms^{-1}.

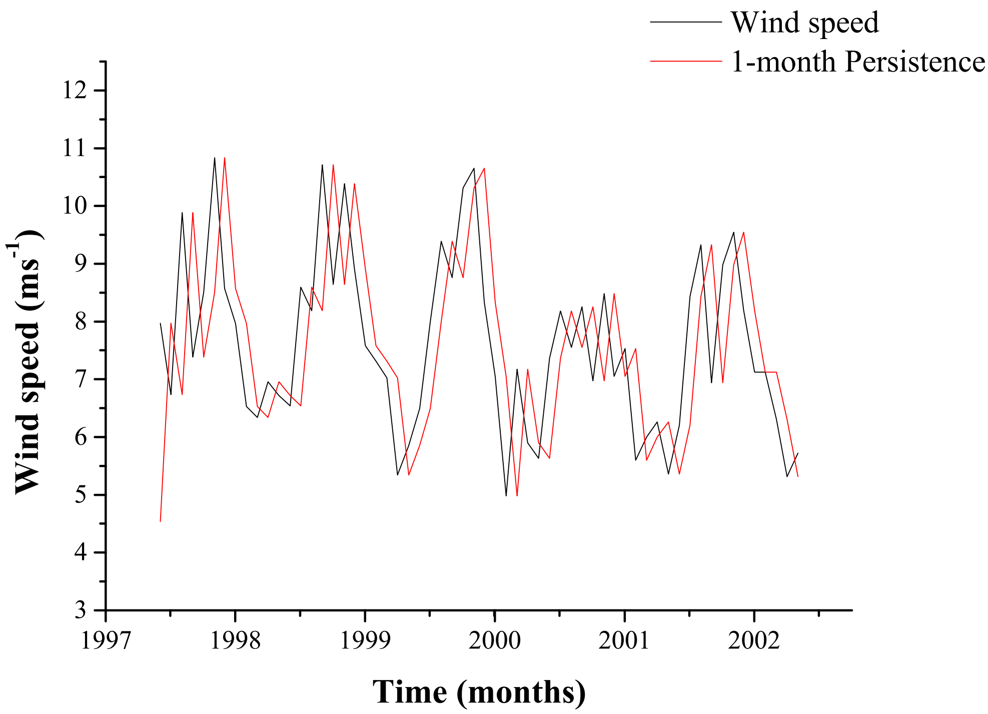


Figure 5.3: *Predictions of the 1-month Persistence model at Lerwick*

At this point, it is noteworthy to mention that one should not confuse the x-axis of the Figure (5.3). The actual time is months as each point on both the actual wind speed and the model is in months. However, since there are 60 points for the actual wind speed and the Persistence it was deemed useful to present the major labels of the graph in years. This also applies to similar graphs regardless of whether they illustrate performance in predictions or statistical errors. The table below contains a summary of the MSE scores for the BADC-7 stations.

Stations	MSE (ms^{-1})
Lerwick	2.197
Stornoway Airport	1.820
Valley	2.260
Aldergrove	0.781
Boscombe Down	0.820
Aberporth	1.844
Tiree	1.930

Table 5.2: *1-month Persistence MSE scores for the BADC-7 stations*

5.2.2 Seasonal Persistence Model

Following the naive predictor and based on the seasonal patterns observed in Figure (5.1) during the FFT analysis, the next model tested was a different version of the Persistence model named Seasonal Persistence. This forecasting model assumes that the forecasted value at time $t+h$ will be the same as the value at time t, where h represents the periods ahead of the forecast ($h = 12$ for monthly data):

$$\hat{y}_{t+h} = y_t \tag{5.3}$$

where $\hat{y}_{t+h}$ is the forecast value at period $t+h$, and
y_t is the actual value at time t.

The following Table (5.3) provides a direct comparison of the MSE scores presented in Table (5.2) for the 1-month Persistence and the MSE scores achieved by the Seasonal Persistence. The results indicate the superiority of the seasonal model. This is due to the model's ability to generate predictions for more periods ahead. On the contrary, the 1-month Persistence resembles a nowcasting model which offers limited predictability. Moreover, in relation to the FFT analysis, Table (5.3) shows that in all cases where strong seasonality was observed the Seasonal Persistence performed better than the simple model. Aberporth and Valley are an exception in that, despite featuring strong seasonal characteristics, they achieved lower MSE under the 1-month Persistence. This discrepancy is appreciated to be attributed to local microclimate as both sites are located in Wales and are in close proximity to the sea.

Stations	**1-month Persistence MSE** (ms^{-1})	**Seasonal Persistence MSE** (ms^{-1})
Lerwick	2.197	1.377
Stornoway Airport	1.820	1.466
Valley	2.260	2.878
Aldergrove	0.781	0.879
Boscombe Down	0.820	0.897
Aberporth	1.844	2.308
Tiree	1.930	1.626

Table 5.3: *1-month and Seasonal Persistence MSE scores for the BADC-7 stations*

5.2.3 Holt-Winters Model

The third model tested in this research was the so-called Holt-Winters model or Single Exponential Smoothing. This forecasting model is based on two factors, the forecast from the previous period and the actual value in the previous period. This model is equivalent to an ARIMA (0,1,1) and its mathematical expression is as follows:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_{t-1} \tag{5.4}$$

where $\hat{y}_{t+1}$ is the forecast value at period $t + 1$,
y_t is the actual value at time t,
$\hat{y}_{t-1}$ is the previous estimate at period $t + 1$, and
α is the smoothing factor, $0 \leq \alpha \leq 1$

Ultimately this model is based on the weighted average of the past values of the series and hence priority and weight is given to the most recent values. The results from this model in terms of MSE are presented in the Table (5.4) below. In all the cases except Tiree, the ARIMA model outperforms the previous naive predictors. Interestingly for the Tiree case, the Seasonal Persistence showed better performance which is very likely to be due to the strong seasonal component observed in the series.

Stations	**ARIMA (0,1,1) MSE** (ms^{-1})
Lerwick	1.336
Stornoway Airport	1.335
Valley	1.776
Aldergrove	0.565
Boscombe Down	0.585
Aberporth	1.682
Tiree	1.876

Table 5.4: *ARIMA (0,1,1) MSE scores for the BADC-7 stations*

5.2.4 ARIMAX/SARIMAX Models with X-input from MIDAS Dataset

One of the main goals derived from the initial hypothesis was to construct a model that would use exogenous inputs along with wind speed. Essentially the exogenous variable would integrate an ordinary regression model using external variables into the autoregressive model. To do so several meteorological variables that were recorded at the same time with wind speed were tested. From those meteorological variables that are stored in the same database within the MIDAS dataset only two were selected to be used as the exogenous inputs. These two variables were:

- Mean sea level pressure (MSL)
- Air temperature

The MSL was collected with precision aneroid barometers with a correction for altitude applied by MIDAS to obtain the pressure at mean sea level measured to the nearest 0.1 hpa. The air temperature was measured by using thermometers on site and records were stored to the nearest 0.1 °C.

These variables were deemed suitable after using an Analysis of Variance (ANOVA) that showed the significance of each of the meteorological factors in relation to wind speed. ANOVA is a statistical test that provides information about the effect of each of two or more variables on another variable. The resulting p values (Probability>F) express the probability of obtaining the data obtained, given that the null hypothesis is true (i.e. no effect). The null hypotheses for the ANOVA test is that the mean is the same for all groups. If the p value is near zero then at least one sample mean is significantly different from the other sample means or that there is a main effect due to this factor.

	Mean	σ	**SE**	**DF**	**Sum of squares**	**Mean squares**	F**-statistic**	p **values**
Wind	7.290	1.615	0.07	2	3.14E8	1.57E8	8.95E6	0
MSL	1010.293	6.584	0.304	2	3.14E8	1.57E8	8.95E6	0
Air	6.542	2.580	0.119	2	3.14E8	1.57E8	8.95E6	0

Table 5.5: *ANOVA for Lerwick station*

The results showed that the p values for air temperature is nearly zero while for the MSL was found to be zero for all the stations. This is also confirmed when two ARIMAX models were tested each of which used one of the aforementioned variables. When the Tukey test was applied a sign equal of 1 was found for all three groups. This suggests that the difference between these groups is significantly different. The Tukey test is similar to the $t - test$ and is expressed mathematically as:

$$q_s = \frac{Y_a - Y_b}{SE} \tag{5.5}$$

where Y_a is the larger of the two means compared, Y_b is the smallest of the two means compared and SE is the standard error.

Table (5.6) below shows the MSE scores for all stations.

Stations	Air Temperature MSE (ms^{-1})	Mean Sea Level Pressure MSE (ms^{-1})
Lerwick	1.277	1.310
Stornoway Airport	1.373	1.344
Valley	1.745	1.784
Aldergrove	0.511	0.564
Boscombe Down	0.582	0.575
Aberporth	1.607	1.654
Tiree	1.745	1.873

Table 5.6: *ARIMAX MSE scores for the BADC-7 stations*

The results for the ARIMAX models corroborate partially the initially hypothesis. It is proven that a model integrating external meteorological variables can produce better mean monthly wind speed forecasts than a model which does not. However, the picture is far from complete as one of the main objectives of this thesis was to test models that, while they use exogenous inputs, they take into consideration the seasonal component in the series. Therefore, following the construction of the above models, seasonal models were also developed. The score in the MSE depicts the superiority of the SARIMAX model against the simple ARIMAX. The table below shows the MSE achieved for all the BADC-7 stations.

Stations	Air Temperature MSE (ms^{-1})	Mean Sea Level Pressure MSE (ms^{-1})
Lerwick	1.04	1.070
Stornoway Airport	0.747	0.762
Valley	1.005	1.089
Aldergrove	0.458	0.467
Boscombe Down	0.350	0.320
Aberporth	0.807	0.864
Tiree	0.842	0.989

Table 5.7: *SARIMAX MSE scores for the BADC-7 stations*

It is clear that using a model that treats wind speed as a dependent variable and other meteorological variables as exogenous inputs shows higher accuracy and performance compared to a model which does not. Nevertheless, the models presented so far assume that the effect the independent variables have on the dependent one occurs at current time. Thus, they fail to consider effects that take place at different time lags.

5.2.5 SARIMAX Model with X-input from the ERA-40 Dataset

This section presents a model that uses wind speed as a dependent variable while it assumes that there is an effect of wind and other meteorological variables at different time lags. Checking the correlation coefficient between two variables helped to identify and measure the strength of their linear relationship (if any existed). Wind speed records covering the size of the geographical area of the UK were averaged and tested for correlation with other meteorological variables such as:

- Wind speed,
- Mean sea level pressure, and
- Sea surface temperature

5.2.5.1 Exogenous Variables

- *Wind speed:*
  Wind speed was chosen to be tested for correlation due to the possibility of existence of a pattern in the circulation of the air masses. Such a pattern was worthy to investigate since this would mean that measuring wind speed in a location other than the UK, e.g in an Atlantic Ocean area, could improve the estimation of the wind speed in the UK. This would be of great importance, especially if it proved that it could occur at different time lags. The latter was considered useful as it would help estimating wind speed in the UK by recording in advance wind speed in different geographical areas.
- *Mean sea level pressure (MSL):*
  Cluster analysis has been employed on sea level pressure data in several studies looking at identifying climatological regimes and weather forecasting [176–178]. Nevertheless, it has also been used specifically in wind studies, as the one presented by Paluticof et

al. [179], where the authors used mean sea level pressure data to reconstruct surface wind speeds. It is also known that air density can contribute to the calculation of wind power (P_w):

$$P_w = \frac{1}{2}\, C_P\, \rho\, A\, u^3 \text{ (W)} \tag{5.6}$$

where ρ is the density of air (1.225 kg/m^3), C_P is the power coefficient and A is the rotor swept area in m^2. Air density, however, is derived from barometric pressure and air temperature.

$$\rho = \frac{p}{RT} \text{ (kg/m}^3\text{)} \tag{5.7}$$

where p is the air pressure in Pa or N/m^2, R is the specific gas constant for air (287 J/kg), and T is the air temperature in degrees Kelvin.

If, however, the air pressure on the site of interest is not available, equation (5.6) can be rewritten as follows:

$$\rho = \left(\frac{p_o}{RT}\right) \exp^{\frac{-g \cdot z}{RT}} \text{ (kg/m}^3\text{)} \tag{5.8}$$

where p_o is the sea level pressure, g is the gravitational constant (9.81 m/s^2), and z the elevation height above sea level in m.

It is clear from the above equations that knowing sea level pressure can aid in estimating wind power. Therefore, in this study, MSL was also employed in an attempt to identify a strong relationship between wind speed over the UK and MSL over different geographical domains.

- *Sea surface temperature (SST):*
  One candidate predictor for the European climate is SST anomalies, as documented by Sutton and Allen [180]. In support of this, further work observed that at a monthly and seasonal timescale the anomalies in SST are linked to the anomalies of other atmospheric variables such as precipitation and air temperature [181, 182]. SST anomalies can also act as useful predictors for monthly rainfall at 1− and 3−month lags, according to Drosdowsky and Chambers [183]. Similar findings indicate the role of SST in wind. Specifically, large-scale persistent wind anomalies associated with El Niño have a strong

relationship with SST anomalies [184], while Chiang et al. [185] demonstrated that winds are influenced by surface pressure gradients and thereby by SST. At this point it should be stated that a similar study that would link specifically SST to wind speeds across the UK is yet to be presented. However, it is clear from this body of work that SST is associated with several weather anomalies. For that reason, SST was chosen to test whether it closely correlates with the wind speed over the UK.

The data used for the estimation of the correlation coefficients were retrieved from the ERA-40 reanalysis dataset. For that reason, the dataset was split in different but equivalent grids such as the one that Figure (5.4) covers. The choice of testing grids of the same size as the UK and not individual points lies in the fact that, in their raw format, the data are distributed by being split by one degree in terms of latitude and longitude. The region for the UK (see Figure (3.2) in Chapter 3), for example, consists of a grid whose latitude ϕ in degrees North, and longitude λ in degrees East are integrals, where:

$$\text{grid}_{\text{uk}} = [\phi, \lambda] = [50°\text{N to } 61°\text{N}, 9°\text{W to } 2°\text{E}]$$

Moreover, the region under investigation in this study consists of a grid whose latitude ϕ and longitude λ are integrals, where:

$$\text{grid}_{\text{study}} = [\phi, \lambda] = [14°\text{N to } 73°\text{N}, 85°\text{W to } 2°\text{E}]$$

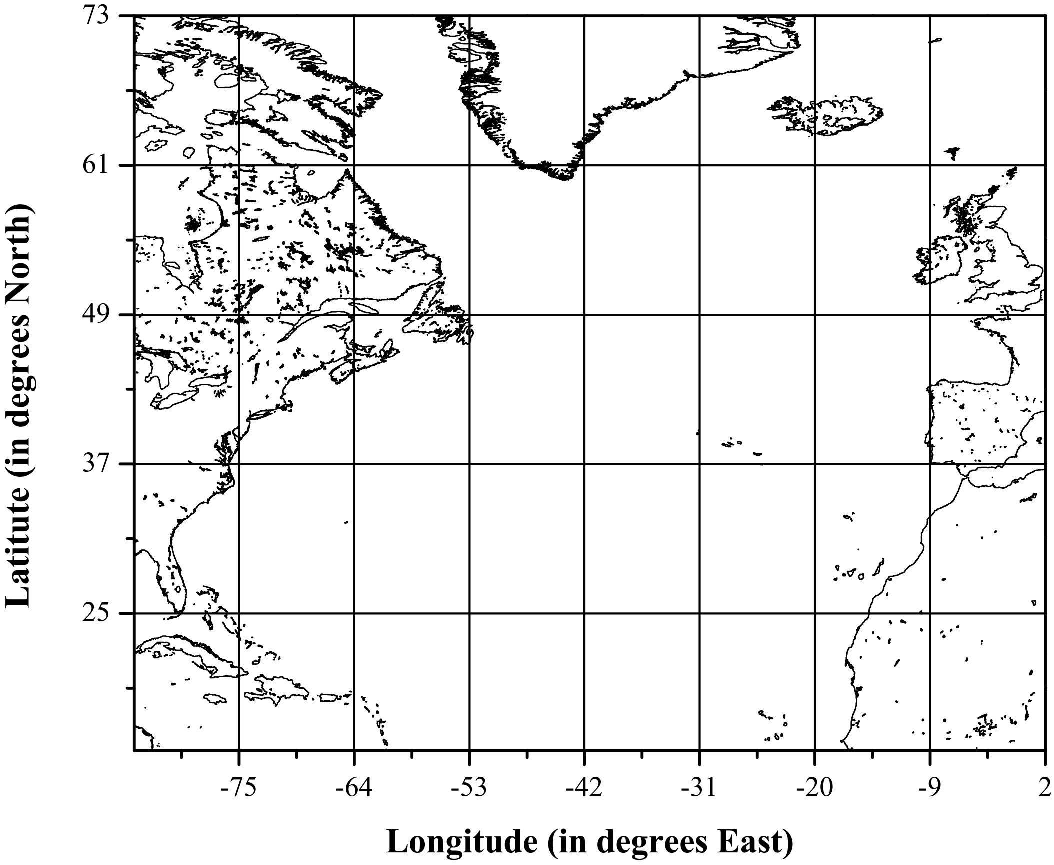


Figure 5.4: *Domain under investigation for correlation between the wind speed over the UK's grid and several meteorological variables over the remaining grids*

It is obvious that the larger grid contained approximately 4,900 individual points each of which corresponded to a set of records taken at different time intervals. Thus, trying to measure the correlation coefficients between one single point with the rest of the points, and also with itself, at varying time lags would be impractical[7] considering the time limits and the scope of this work.

Figure (5.5) shows scatter plot matrices that provide a visual summary of pair-wise comparisons between the wind speed over the UK and (*a*) wind speed (in ms^{-1}), (*b*) SST (in °C), and (*c*) MSL (in hPa), over different geographical grids.

The high correlation coefficients observed from the scatter plot matrices do not necessarily imply that one variable causes the other. However, this information gives an indication of a

[7] Based on combination theory, if k is the number of the time lags and n the total number of records measured at a given time, then the total number of combinations that need to be checked for correlation is equal to: $(n-1)k^2 + \frac{k!}{(k-2)!2!}$. Since, $k! = k(k-1)(k-2)!$, the total number of combinations is equal to $k[(n-1)k + \frac{(k-1)}{2}]$. This means that, in this study, the total number of combinations for just a single point is 828,009 and over 109,000,000 different combinations for all the points (~ 132) that cover the UK's area.

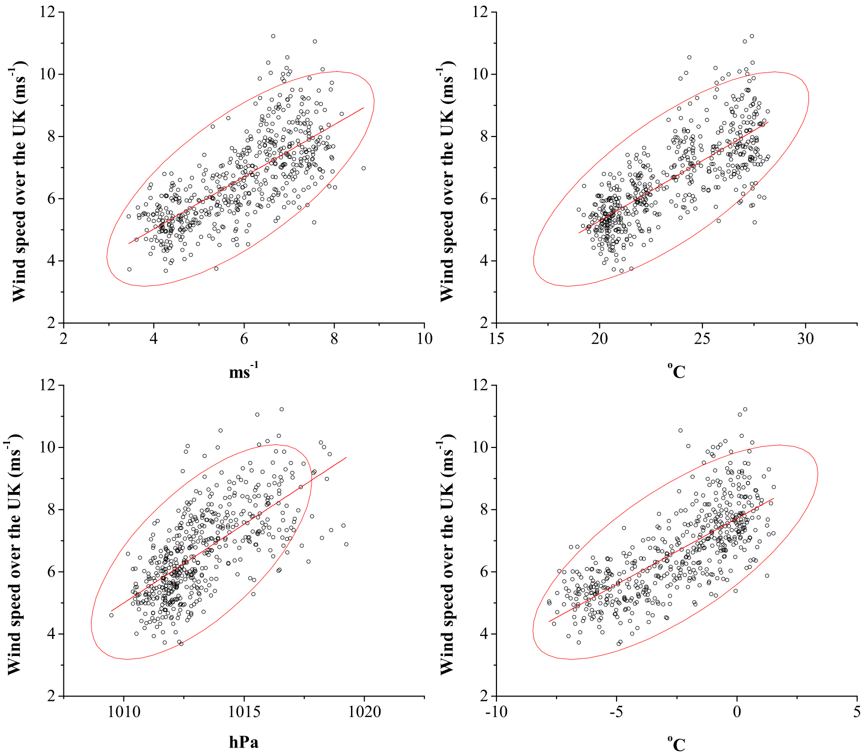


Figure 5.5: *Scatter plot matrices between wind speed over the UK and different independent variables for different time lags*

strong association between these variables. This encouraged the idea of incorporating these independent variables into a model for predicting the dependent variable of interest (i.e. wind speed).

It also exhibits another case (d) which is the SST gradient (in °C). SST gradient is the rate that SST increases or decreases in relation to change in distance. The reason for including the SST gradients will be discussed further below. As shown in Appendix (F.1), all the tested independent variables had high correlation coefficients for different time lags. The backwards lag for each predictor was chosen under the condition that it could improve the forecasts of wind speed over the UK. To explain in more detail, Table (F.1) in Appendix (F.1) demonstrates that wind speed from the ERA-40 dataset was highly correlated with wind speed over the UK during the time of which the observations occurred, as well as for 1−month lag. However, this can not increase the efficacy in forecasts since the wind speed over the UK needs to be estimated beforehand. On the other hand, in the same Table, 8− , 11− and 12−month lags may be deemed useful due to higher correlation coefficients. Figures (5.6) - (5.8) illustrate the Pearson's correlation over the domain under investigation. In Figure (5.6) it is demonstrated that for a 12−month lag wind speeds over the Atlantic Ocean, Iceland, Greenland and the North

Sea were highly correlated with wind speed over the UK.

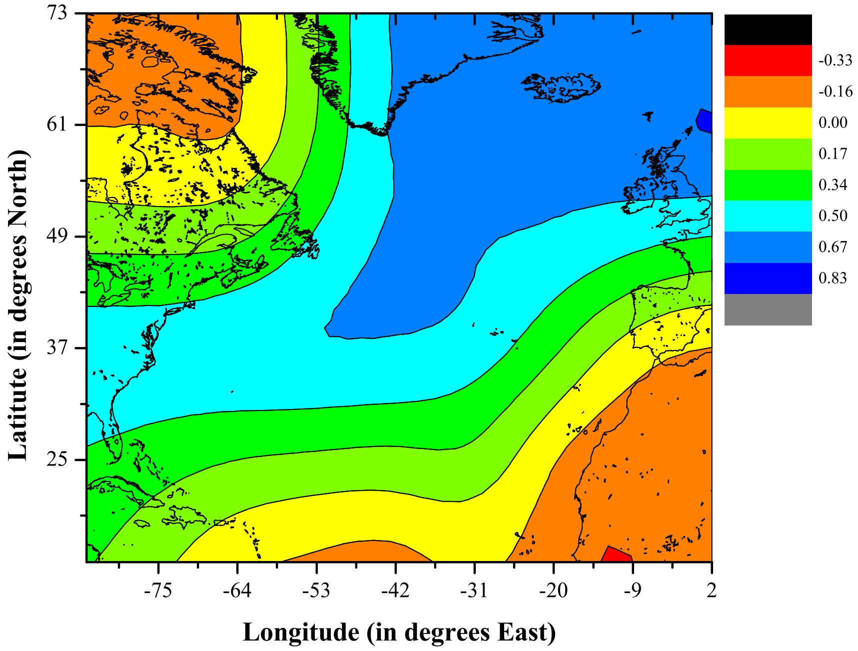


Figure 5.6: *Correlation coefficient between wind speed over the UK and wind speed from ERA-40 for a time lag of 12 months*

Figure (5.6) confirms the findings presented by Qu et al. [186] that the variation in SST lagged three months behind that of wind speed.

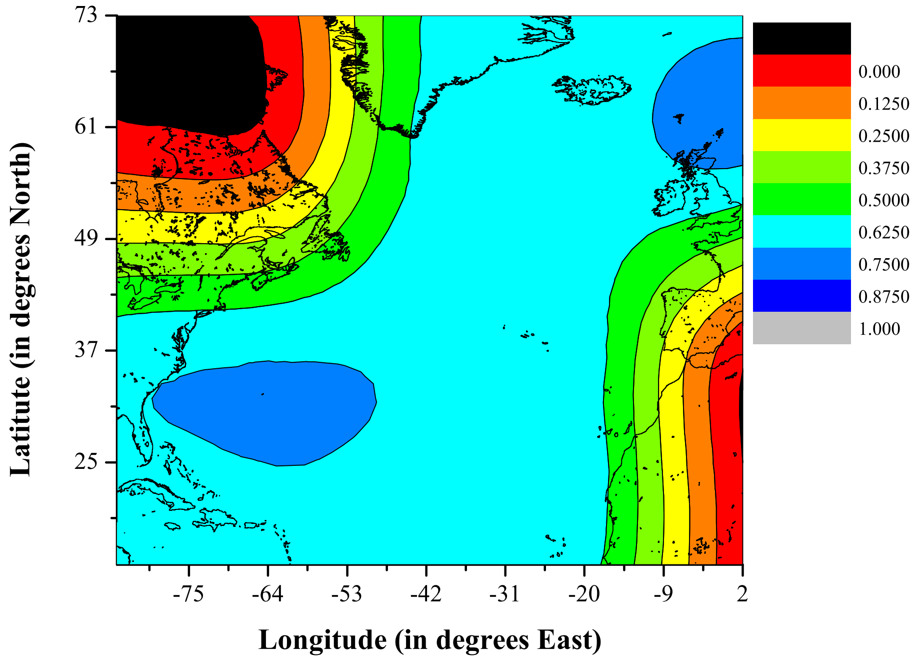


Figure 5.7: *Correlation coefficient between wind speed over the UK and SST form ERA-40 for a time lag of 4 months*

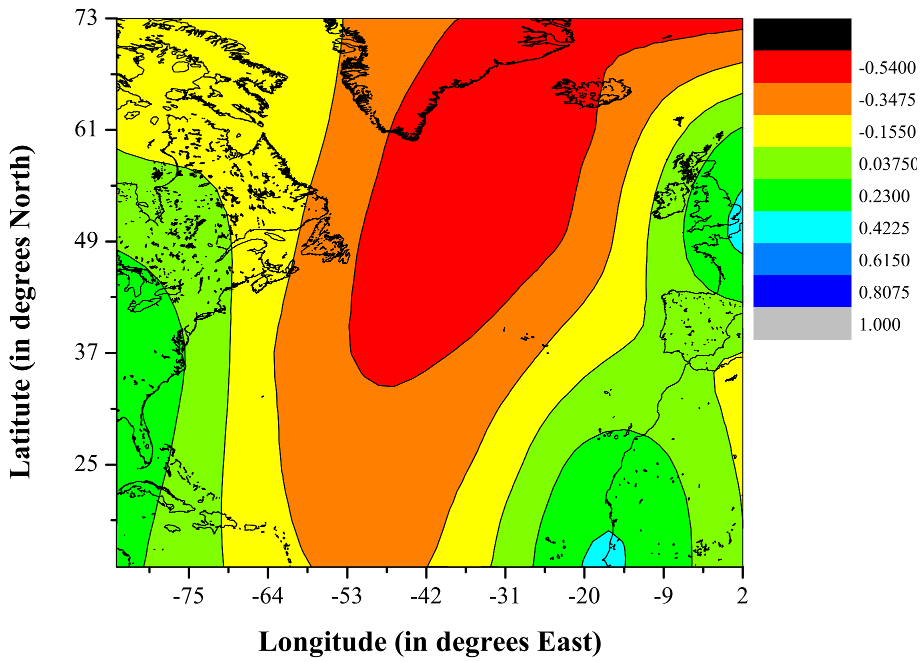


Figure 5.8: *Correlation coefficient between wind speed over the UK and MSL form ERA-40 for a time lag of 12 months*

Figure (5.9) presents the histograms of the residuals from the correlation coefficients after fitting a linear model to each case. All histograms showed symmetry and they were evenly distributed around zero. This means that the assumption of normality is likely to be true and that the observed residuals coincide with the theoretical ones. Thus, the relationship between the dependent and the independent variable(s) can be modelled by using models from the ARIMA family.

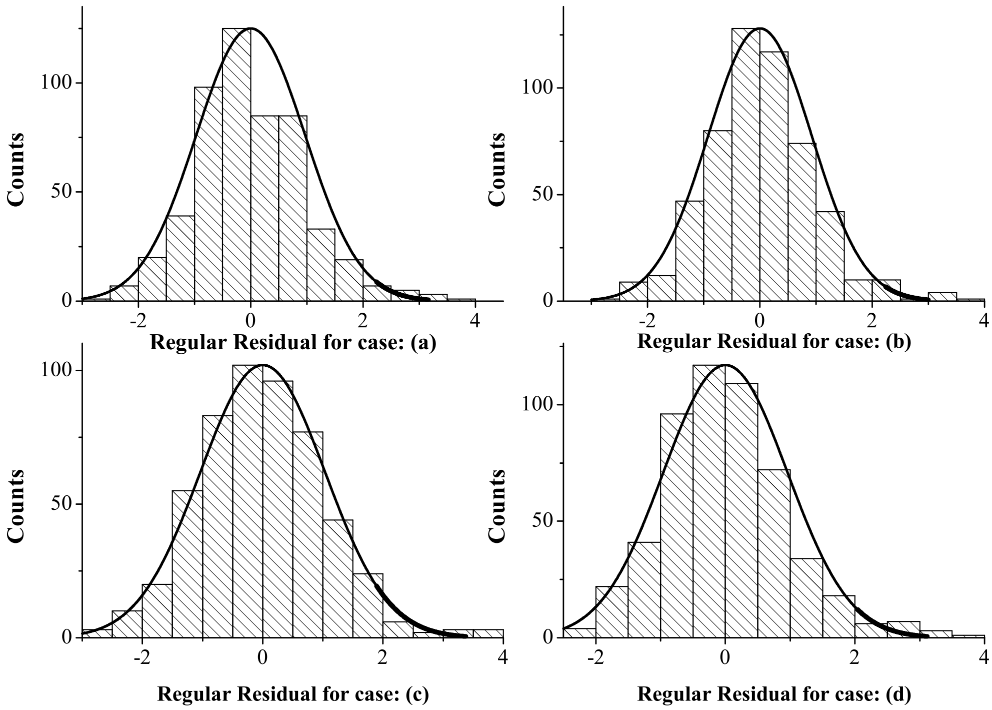


Figure 5.9: *Histograms of regular residual for all cases*

However, seasonality may mask any hidden patterns in the series. For that reason, the seasonal differences of the series were also taken into account as well as a further difference, in case remaining non-stationarity was still present (see Makridakis et al. [187] for further details about differencing time series). The results showed that by removing the seasonal component in the series the correlation coefficients dropped significantly. From this analysis it was found that seasonality plays an important role in both exposing patterns in the series and depicting high correlations between different variables. For brevity reasons, these results are presented in Appendix (F.2).

Several authors have indicated a strong relationship between surface convergence and deep convection which, in turn, may affect surface pressure gradients, winds, and climatological

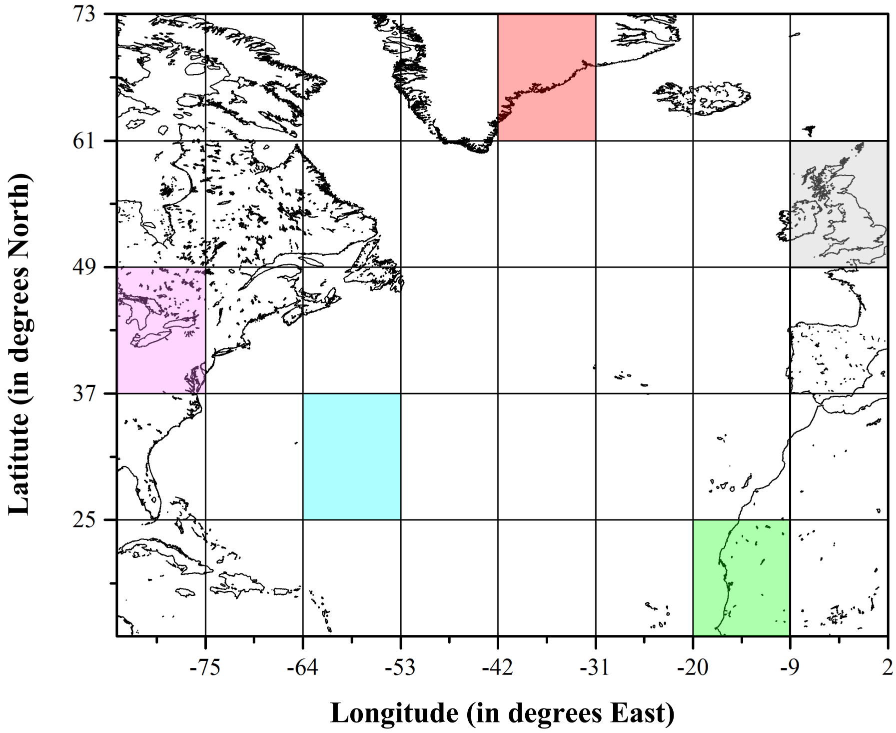


Figure 5.10: *Domains with the highest correlation between the wind speed over the UK's grid and the X-inputs*

convergence patterns. This relationship has been explained by SST gradients and documented in many studies [188–191]. Therefore, similarly to the previous approach, the correlation coefficients between wind speed over the UK and SST gradients (known henceforth as -case-(d)) were also calculated. Following the aforementioned discussion, the results presented in Table (F.7) in Appendix (F.3) encouraged the use of the SST gradients since the correlation coefficients were significantly high. Again, similarly to the cases of wind speed (a), SST (b), and MSL (c), high correlation coefficients were observed for different time lags. At this point, it is noteworthy that differencing the SST gradients series resulted in even lower correlation coefficients. Figure (5.10) shows the geographical location of the cells which are highly correlated with the wind speed over the UK. The red, magenta, green and purple cells illustrate the areas where the wind speed, the SST, the MSL and the SST gradient are highly correlated with the wind speed over the UK.

5.2.5.2 Model Selection and Fitting

In the previous section it was documented that the variable of interest (i.e. wind speed) has strong correlation with other variables. However, this association with wind speed occurred at different time lags depending on the independent variable chosen. For instance, it was found that the wind speed over the UK has strong correlation with the wind speed over Greenland on a time lag equal to 12 months. This means that there is a strong association between measuring wind speed over that region 12 months in advance and wind speed over the UK a year after. Similarly, SST over a region East than Florida has high correlation with wind speed over the UK on a time lag equal to 4 months. MSL over Mauritania is associated with wind speed over the UK on a lagging time of 12 months. Last, in the case of the SST gradient (d) the lag time was found to be 10 months.

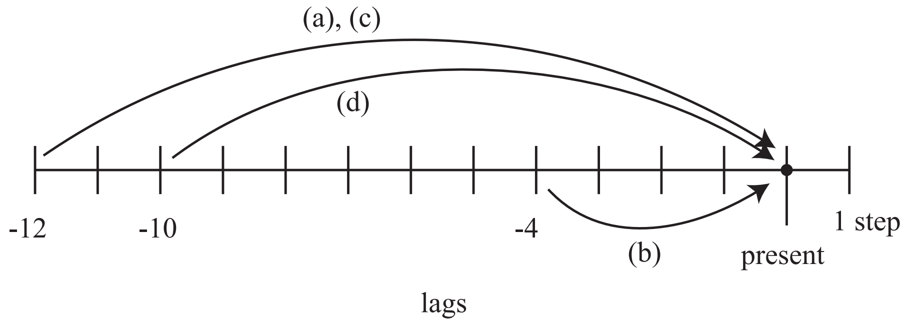


Figure 5.11: *Lags of exogenous variables*

The difference in the time lags resulted in different training periods for each case. Figure (5.12) below, shows the training as well as the estimation periods for each case.

The testing period, on the other hand, was selected to be common for all the different cases. The reason behind this choice was not only to estimate the performance of each model but also to conduct a comparison between those models under the same conditions.

The fitting of all models was tested based on Akaike's information criterion (AIC) [192]. AIC is a metric tool that can identify whether a model fits well to a set of observations, and is expressed as:

$$\mathrm{AIC} = -2\ln L_f + 2p \tag{5.9}$$

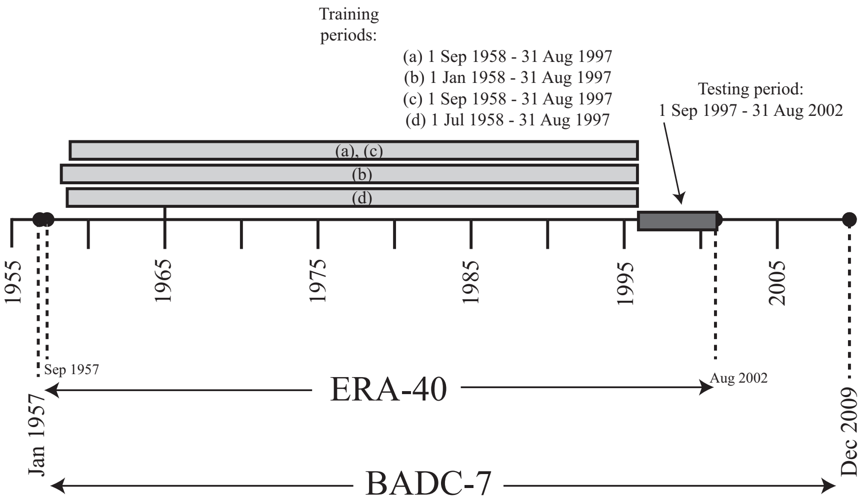


Figure 5.12: *Training and testing period of data*

where p is the number of parameters in the model and $\ln L_f$ is the natural logarithm of the maximum likelihood (L_f). The least score in AIC resulted in the selection of the best model.

The model presented herein, was first published by Kritharas et al. [100] with the difference that this study used atmospheric pressure, atmospheric temperature, and relative humidity data measured at the location of each station as independent variables. The authors proposed a SARIMAX model, as described in section 2.1.2.2 and expressed by equation (2.22).

In contrast to the present work, Kritharas et al. [100] assumed that the effect of the exogenous variable(s) on the predictand occurred at concurrent times. The association between the meteorological variables and wind speed at different time lags that was previously presented was used to test the hypothesis formulated at the beginning of this study (section 1.6). The research demonstrated in this Chapter can be, therefore, considered a sequel of the previous work.

The proposed model was developed using the "arima" function from the System Indetification Toolbox™ version 2.3 of MatLab®. Based on AIC, the scope during the development of the model was to determine the optimum parameters of equation (2.22) that minimised the Final Prediction Error (FPE). Akaike's FPE is defined as:

$$\mathrm{FPE} = V\left(\frac{1 + d/N}{1 - d/N}\right) \tag{5.10}$$

where V is the loss function, d is the number of estimated parameters, and N is the number of values in the estimation data set [193].

According to the user's guide, the toolbox assumes that the final prediction error is asymptotic for $d << N$ and uses the following approximation to compute FPE:

$$\mathrm{FPE} = V\left(1 + 2d/N\right) \tag{5.11}$$

The SARIMAX model was justified from section (5.1.2) when the series were checked for stationarity. Following this, it was also shown that there is a correlation with other meteorological variables and wind speed in the UK. This was also documented by presenting several models that surpassed Persistence. However, as it was amply explained, a model that could use as exogenous inputs variables that lag from wind speed has never been previously published. Using the AIC criterion mentioned above for fitting the models led to the development of several orders of SARIMAX for each case. The best models are presented in Table (5.8).

Cases	Order	FPE	V
Wind speed	$(4,1,1,1)(0,1,1)_{12}$	0.744	0.694
SST	$(1,1,1,1)(0,1,1)_{12}$	0.734	0.693
MSL	$(1,1,1,1)(0,1,1)_{12}$	0.730	0.69
SST gradient	$(1,1,1,1)(0,1,1)_{12}$	0.775	0.73

Table 5.8: *Order of the best SARIMAX for each model*

The term *order*, as mentioned in section (2.1.2.2), refers to the order of the autoregressive process which is expressed as a polynomial that takes into account only the previous terms of the process and the error term. A SARIMAX is essentially a SARIMA with exogenous input with degree p, q, P, Q, is an extension of an ARIMA model, which also takes into account seasonality s, and can be written as a SARIMA $(p, d, q) \times (P, D, Q)_s$.

In order to evaluate the performance of these models, the data were separated into two subsets;

a training and a testing one, as Figure (5.12) illustrates. This study was focused on training the models using a long dataset while at the same time evaluating their performance over a relatively long time horizon. In contrast to the models that employ AI, such as ANNs and SVMs (see Chapter 2), the model used in this study required a long dataset in order to capture the characteristics of the time series. For this reason, 468, 476 and 472 monthly mean values were used for training the models for the cases (a) *and* (c), (b), and, (d), respectively. The next section presents how the models performed as well as their scores in terms of MSE.

5.3 Model Validation and Statistical Errors

Figure (5.13) illustrates the performance of the proposed model when predicting the monthly mean wind speed over the region of the UK.

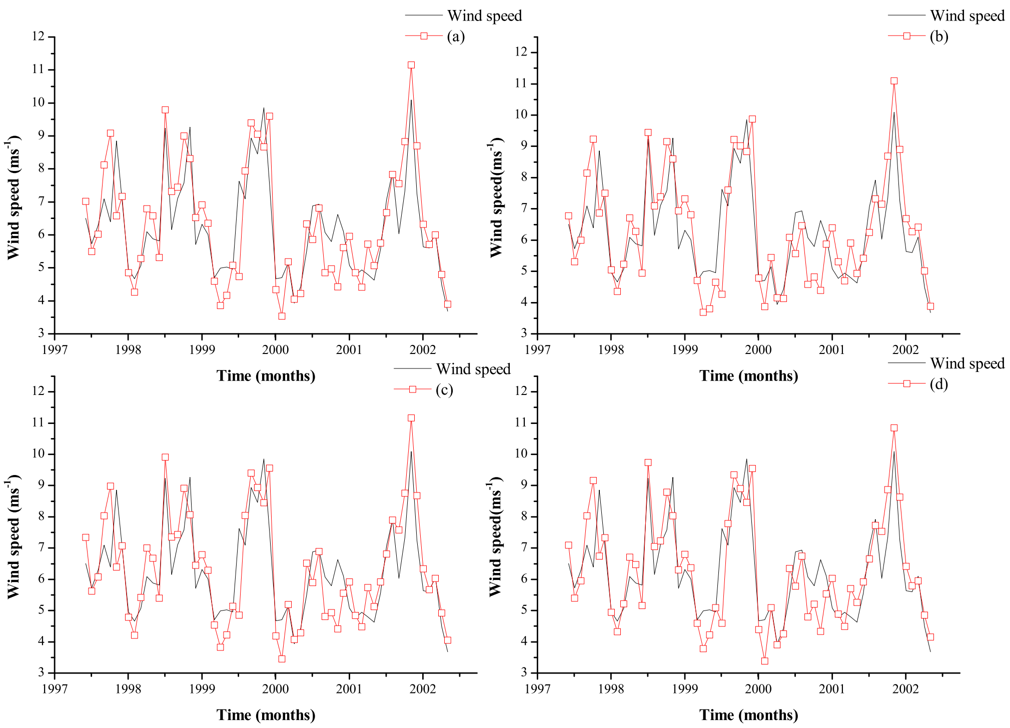


Figure 5.13: *Predictions of SARIMAX models over the whole UK*

From the visual inspection of the plots it is confirmed that the models do capture the changes in monthly mean wind speed on a satisfactory level. However, during individual years where extreme events were observed the model failed to capture instantly the steep changes such as in Summer 1997 and Autumn 2000. At this point, it is noteworthy that similar to Figure (5.3) the actual time is in months as each point on both the actual wind speed and the model is in months. Interestingly, each model after experiencing a sudden change, it slowly regains its ability to capture the wind speed variability. This confirms the benefits of autoregressive moving average models. Nonetheless, which model produced the best result was concluded from a direct comparison of the errors that each model generated. To accomplish this, the statistical metric of MSE was used, as previously described in section 2.1.2.2. Table (5.9) below presents each case and the corresponding value in terms of MSE. From this comparison, it was shown that the best independent variable for generating monthly mean wind speed forecasts is MSL. At this point, it was crucial to determine if the data fitted well to the model. The common diagnostic test for doing so is to plot the ACF and PACF of the residuals. If the plots revealed any autocorrelation in the data, it would mean that there would be still information undetected by the models. As Figure (5.14) depicts, the data fitted well to the model since there was no trace of strong autocorrelation. This is not to be confused with the weak single spike at lag 11. This is not evidence of a strong remaining autocorrelation which would mean that the model's parameters had been specified wrongly. As demonstrated by Makridakis et al. [187], a strong ACF consists of several positive or negative spikes that tend to reach $+1$ and -1, respectively. Indeed, this is not confirmed by the PACF in the same figure which means that there is no remaining information in the series.

Cases	MSE
Wind speed	1.248
SST	1.154
MSL	0.973
SST gradient	0.996

Table 5.9: *Error of each SARIMAX model*

From this evaluation, as it can be seen in Table (5.9), the proposed model which used the MSL input, performed better than the ones that used the wind speed (a), SST (b), and SST gradient (d) by reducing the error of the prediction by 22.1%, 15.68%, and 2.29% respectively.

However, throughout the literature review in Chapter 2 the criticism of the past work lies in the fact that all of the studies have not used a single dataset to evaluate the performance of the proposed models. As a consequence, the body of literature is weak in producing a robust result which would derive from a generic comparison of all the models on the same dataset. For this reason, the proposed approach was tested for the BADC-7 stations. As it was mentioned previously, earlier work [100] used different statistical models for the same dataset

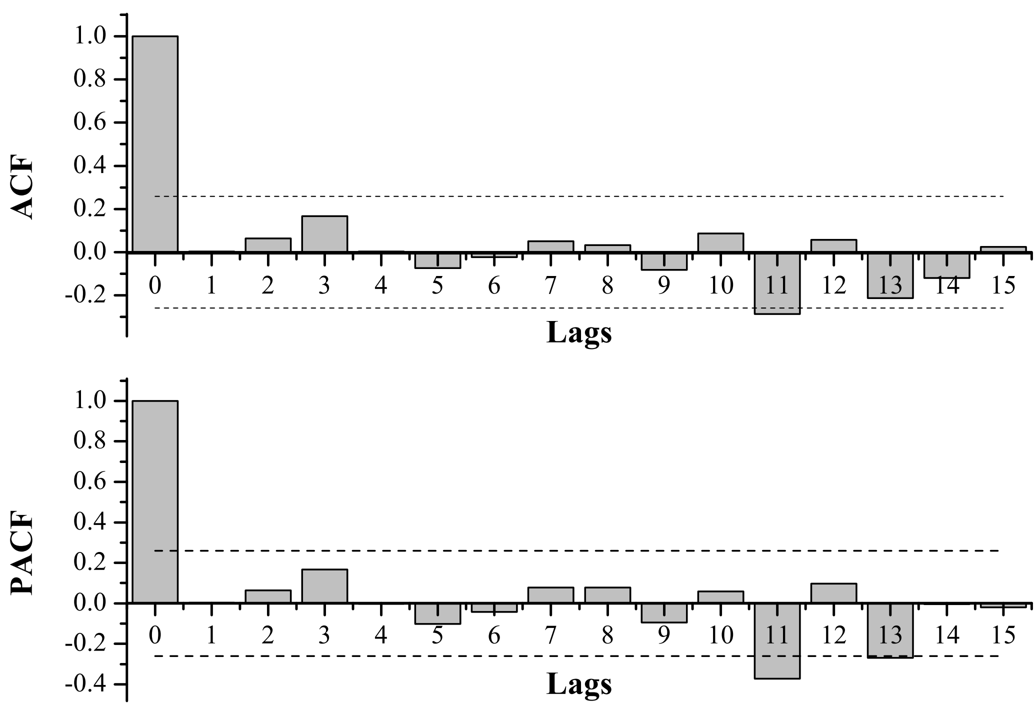


Figure 5.14: *ACF and PACF of the residuals for the SARIMAX over the UK*

employed in this research. The aim of our study was then directed to determine if the proposed model surpasses the previous models presented in Kritharas et al. [100]. This is of paramount importance since the results will serve as a springboard for the continuity of the research in this field. It will also draw useful conclusions and recommendations regardless of whether the proposed model performs better than the previous ones.

5.4 Long-Term Wind Predictions at BADC-7 Stations

The approach described in the previous section was used to generate monthly mean wind speed predictions at each station included in the BADC-7 class. In the following pages the results from Lerwick station are presented though the rest figures and tables can be found in Appendix (G).

When the data at Lerwick station was fitted to the proposed model it was found that different order of SARIMAX produced the minimum FPE than the models presented when the wind speed was averaged over the UK. There were two reasons for this result. Firstly, the targeted time series were different on each occasion; in the first case, data from ERA-40 were averaged over the whole UK whereas in the second case the data was retrieved from a distinct geographical location within the UK. Secondly, the local anomalies at each site were more easily revealed than when the wind speed was averaged over a whole region. This averaging procedure suppressed any effects from local anomalies, and hence it was reasonable to expect differences between the two models.

Cases	Order	FPE	V
Wind speed	$(3,1,1,1)(0,1,1)_{12}$	1.207	1.13
SST	$(4,1,1,1)(0,1,1)_{12}$	1.195	1.114
MSL	$(0,1,1,1)(0,1,1)_{12}$	1.23	1.168
SST gradient	$(8,1,1,1)(0,1,1)_{12}$	1.171	1.074

Table 5.10: *Identifying the order of the best SARIMAX on the training dataset for each model at Lerwick*

Similarly to before, the orders of the models were determined by assessing the MSE in each case. The most promising remarks were drawn when the model developed was compared with the models presented by Kritharas et al. [100]. As Table (5.11) demonstrates the proposed model increased the efficacy in the predictions by 10.04%, 12.8% and 14.6%. In the case of the whole UK (Table 5.9), MSL was the *best* predictor at all stations. However, in the case of each station SST gradient exhibited the lowest errors. The exception was Aberporth achieving the lowest errors when using the SST. That station is located in Wales and therefore it is assumed that these differences may be attributed to each station's microclimate and local anomalies, though this requires further investigation. It is noted that this research is confined to the development of a statistical model without intending to investigate the underlying processes of atmospheric physics. Although the two disciplines (i.e. wind energy and meteorology) are often intertwined, because of the nature of wind and its relationship with atmospheric phenomena, it is beyond the scope and potential of this work to explore this area. However, it would be prudent for follow-up research to focus on this direction incorporating both fields.

Table (5.11) shows the statistical errors achieved by the models when they were introduced in the testing dataset. The best model was proved to be the one that used wind speed as the exogenous input.

The results for the rest of the BADC-7 stations are shown in Appendix (G). Figures (5.15) - (5.16) illustrate the performance of the proposed model. The ACF and PACF plots testify that the data fitted well to the model and there was no remaining autocorrelation in the data.

Cases	Order	MSE	ME ms^{-1}
Wind speed	$(3,1,1,1)(0,1,1)_{12}$	1.120	0.014
SST	$(4,1,1,1)(0,1,1)_{12}$	1.20	−0.044
MSL	$(0,1,1,1)(0,1,1)_{12}$	1.155	0.005
SST gradient	$(8,1,1,1)(0,1,1)_{12}$	1.248	0.056
(atm. pres.)	$(6,1,1,1)(0,1,1)_{12}$	1.245	−0.003
(rh)	$(7,1,1,1)(0,1,1)_{12}$	1.284	−0.011
(atm temp.)	$(7,1,1,1)(0,1,1)_{12}$	1.311	0.072

Table 5.11: *Order of the best SARIMAX at Lerwick and corresponding statistical errors*

Assessing the ME was a way to identify if the forecasts were biased. ME, or else, the *forecast bias* is determined by checking if the residuals show a trend which may reveal consistent differences between actual values and previously generated forecasts. Similar to all the other statistical errors, ME is expressed in ms^{-1} when looking at forecasting errors in wind speed and in kW when looking at the prediction errors in wind power.

As Table (5.11) indicates, the ME is very close to zero which means that the forecasts generated by the models were not biased. In addition to Table (5.11), Figure (5.17) serves as a direct visual comparison of the models used in this study since it shows the Absolute Error (AE) for the validation dataset. Figure (5.17) confirms that the minimum errors occurred when the SARIMAX model used wind speed as an exogenous input.

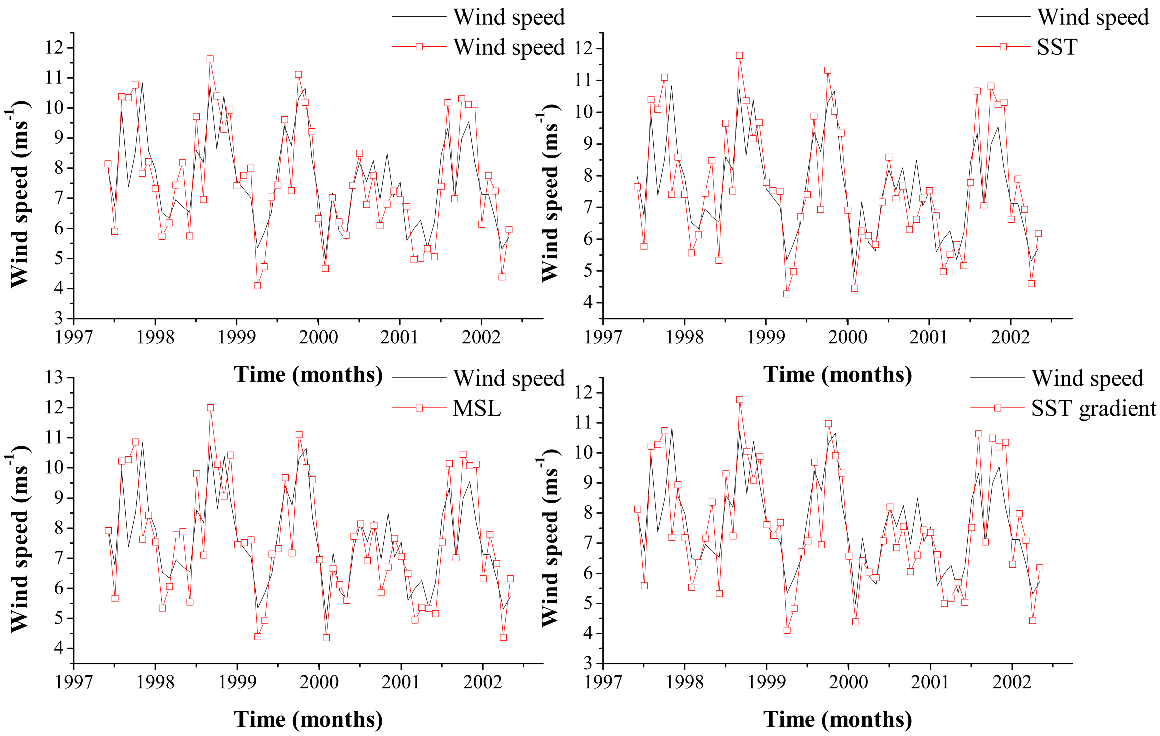


Figure 5.15: *Predictions of SARIMAX models at Lerwick*

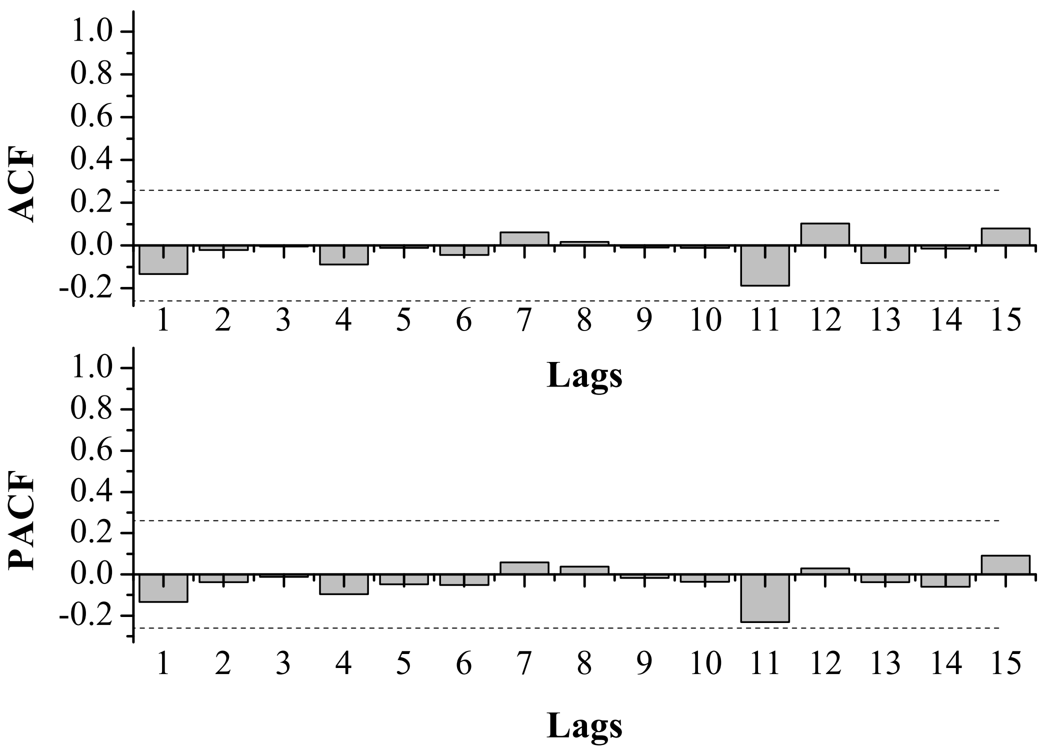


Figure 5.16: *Correlograms of residuals at Lerwick*

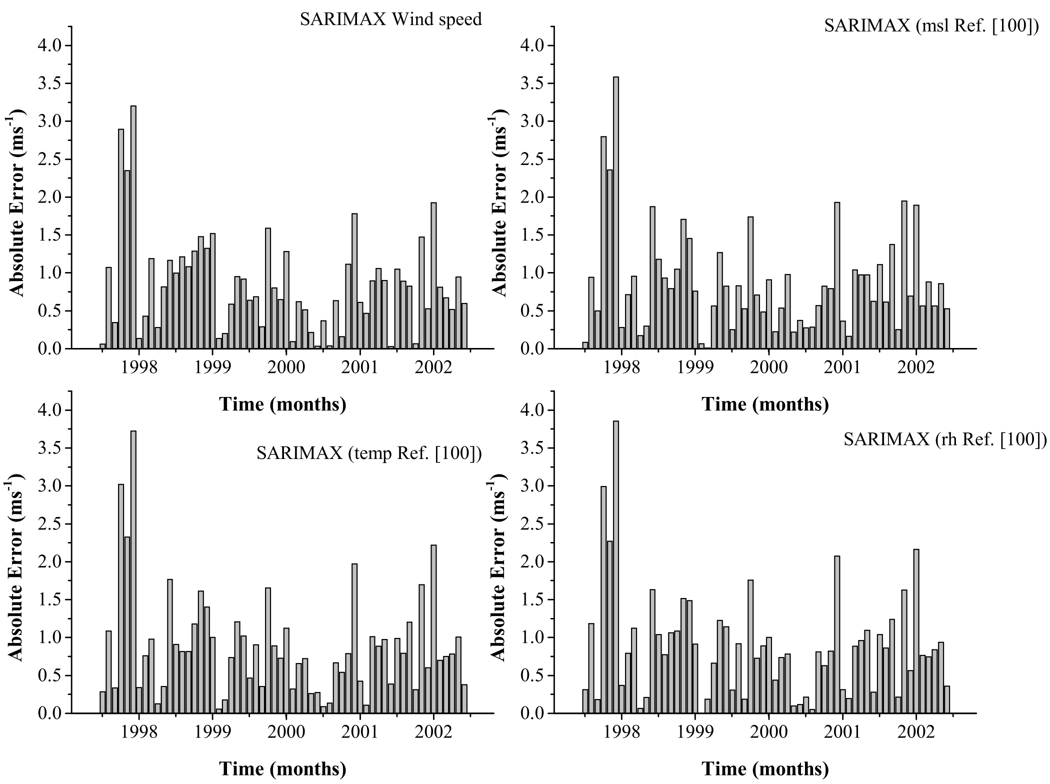


Figure 5.17: *Absolute errors produced by different models at Lerwick*

5.5 Practical Implementation of the SARIMAX

Monthly mean wind speed forecasts can be utilised by different groups from the Wind Industry such as:

- TSOs: Monthly wind speeds could offer insights in monthly power outputs which allows for calculating monthly capacity factors. This information can be used for planning and scheduling power systems in a more effective way.
- Consultants: Typically, wind assessments undertaken by wind consultants are based on monthly mean wind speeds as described by Youm at al. [194]. The majority of the wind assessments require correlating the mean monthly records of wind speed at a location of interest with a long-term historic dataset. By this approach the monthly averages can be used for determining:
  1. the inter annual variability of wind speed,
  2. the long-term wind speed estimate which derives from the MCP between the monthly averages and the historic dataset
- Banks: Setting the terms of concessions and repayment of any loan before granting it is a common practice for bankers and external investors. Wind power projects are not different from any investment which requires financial support. Therefore, it is crucial for the bankers to assess the associated risk and the interest rate of return (IRR) that they will have to apply on the project in order to minimise exposure to overall risk [195, 196]. Thus it is a prerequisite for the developers to demonstrate information that proves the viability of the project. Average monthly wind speed provides details on how wind may vary on an annual basis and most importantly how seasonality could affect the project cash flow. The monthly wind speeds along with the measured shear can lead to monthly power production estimates which in turn can be translated to monthly energy production. The latter is important as it is associated with the revenue of the project and essentially will inform the banks whether the project is likely to cover the debt service.

The usefulness and applicability of the proposed model is underlined as it is able to cover the above needs by generating mean monthly forecasts of wind speed for a given location.

As demonstrated throughout this research, this can be achieved by using information related to wind speed that is recorded on site along with data retrieved from the ERA-40 dataset. Therefore, it is advocated that Wind Industry would benefit from a coordinated utilisation of the proposed SARIMAX as it promises to serve several objectives set on different levels required during the development of a wind project.

5.6 Chapter Summary

This Chapter reflects the actual chronological sequence of testing several statistical models that resulted in the development of the SARIMAX model. This process of model building along with the final proposed model constitute a benchmark in long-term wind speed forecasting. The final model, while incorporating both seasonality and monthly autocorrelation, utilised exogenous variables for different time lags. The novelty in this approach is that the inherent limitation of SARIMAX (i.e. to pre-assume that the effect of the exogenous variable(s) on the predictand occurs at concurrent times) was bypassed. This was achieved by calculating the correlation between the dependent and the independent variables at different time lags. Once the best pair (i.e. correlation between wind speed and exogenous variable) for each case was determined the exogenous time series was "shifted". In this way, the models were *forced* to consider the effect that the independent variable had on wind speed for different time lags. The pairs that were created include the correlation between:

- wind speed over the UK and wind speed for a grid of the same size as the UK;
- wind speed over the UK and SST for a grid of the same size as the UK;
- wind speed over the UK and MSL for a grid of the same size as the UK;
- wind speed over the UK and SST gradients for a grid of the same size as the UK.

The results indicated high correlations up to 0.75 between the tested variables and wind speed which is in line with the body of literature.

The model was trained using monthly wind speed observations for a period of 39 years and was tested for a period of 5 years. The optimum parameters of the model for each case were selected by applying the AIC criterion and using the FPE to minimise the prediction error.

When the four pairs of interest were used to forecast the monthly mean wind speed over the UK the results showed an improvement in the predictions. Comparing results to a previous non exogenous approach revealed an increase in the accuracy by maximum 22.03%, in favour of the lag model. Following this, the proposed model was applied to the BADC-7 stations and its predictions were subsequently compared to the ones presented by Kritharas et al. [100]. The results elucidated that the SARIMAX outperformed the previous model by achieving a minimum and a maximum accuracy of 10.04% and 14.6%, respectively.

The ACF and PACF plots of the residuals for each case/station revealed that the data fitted well to the models and that there was no remaining autocorrelation. The fact that there was not any hidden information left in the series testified that the parameters of the models were selected correctly based on AIC criterion. Finally, assessing the ME of the generated time series showed that the forecasts were not biased. These results verified the objective of this research to identify a model that outperforms in accuracy when incorporating exogenous variables.

Chapter 6
Conclusions and Further Direction of Research

This research aimed to develop a model for monthly wind speed predictions. The study was set out to evaluate the accuracy of this model which, by employing different meteorological variables with wind speed over the UK, shows better performance.

This research sought to answer the following questions;

1. Can we forecast monthly mean wind speed?
2. Is there any explanatory variable that affects wind speed?
3. Is it feasible to develop a model with will take into account both the dependent and independent variables?

Good Energy, which supported this research, typically purchase base load by checking historical seasonal trends and hence underestimates the energy generated from their wind farms. This is not cost effective and thus Good Energy appreciated the need for improving their trading mechanism. Therefore, it was necessary to answer the aforementioned questions in order to provide an industrially viable solution. It was deemed of great importance to develop and evaluate a novel model which while belonging to the family of statistical models would reveal less limitations concerning the inherent characteristics of wind speed time series (i.e. non linear).

Wind is indigenous and non depletable. There is no doubt that the years to come the electricity generated from wind will diversify the energy mix and provide security to the energy supply in the UK as well as globally. However, it is vital to forecast wind speed due to it's variable nature. Seemingly absurd, so far the majority of the literature has focused on short-term forecasting of wind speed/power. Nonetheless, there is a need for forecasting wind speed at longer timescales since this would contribute to the evaluation of the technical feasibility as well as the financial viability of wind projects for their expected operational life span.

From the literature on long-term forecasting, it was concluded that the studies overall have employed several models with an attempt to validate them taking into account the weaknesses of each model. Moreover, it is evident that each case study has used different datasets. As a result, taken together, the findings of these studies do not yield robust and indisputable agreement as to which approach produces better results. It would be reasonably expected, after 15 years of research purely on long-term prediction, that there would have been a generic research framework, within which there would have been a comparison of models as well as of algorithms over the same testing datasets for the same predicted ones. This would not necessarily mean that the training dataset must be similar in terms of the timescale; as extracted from the literature, some models are in need of using a larger amount of data to capture the characteristics of time series while others do not. The strand of long-term wind speed/power forecasting should engage in the conduction of research drawing on a large range of work similar to the one presented by Kariniotakis et al. [197] for ANEMOS project.

In this context, developing a statistical model which would improve the accuracy in monthly wind speed forecasts was considered to be critical in establishing a viable industrial tool for planning, operating, and maintenance scheduling tasks.

To accomplish this, the key objectives were set as follows:

1. To investigate the correlation between wind speed and other meteorological variables,
2. To identify a spatial association between the two variables of interest,
3. To take into account different lags in the association between the variables for the development of the model,
4. To compare the proposed model with one that does not take into account exogenous variables

This Chapter presents a synthesis of the findings from the study with respect to the research questions. A discussion follows on the contribution and impact of this work on the existing research both at theoretical and policy levels. Finally, the thesis concludes with a series of recommendations for further direction of research.

6.1 Findings

The main findings which are summarised in the corresponding chapter (see Chapter 5) were:

- Wind speed, MSL and SST reveal a strong correlation with wind speed over the region of the UK up to 0.72.
- Depending on the exogenous input used, this relationship occurs at different time, varying from 4– up to 12–month lags.
- There is a maximum improvement of 22.03% in the accuracy of the model when the exogenous variables are employed into the model as opposed to the lower performance when this is neglected.

The initial hypothesis that the proposed model would improve the monthly predictions is attested by decreasing the error of the predictions by maximum 22.03% and by minimum 2.29% for the whole UK. Equally, when the model was tested for a specific station against another model which does not incorporate exogenous inputs, it decreased the error by a minimum 10.04% and maximum 14.6%. Determining the correlation coefficients between wind speed and the independent variables revealed a strong association. Given the high correlation it was decided to choose the best case to be tested for each variable. It was then found that the wind speed over the UK has strong correlation with the wind speed over Greenland on a time lag of 12 months. This means that measuring wind speed over that region 12 months in advance has strong association with wind speed over the UK. Similarly, SST over a region East of Florida was highly correlated with wind speed over the UK on a time lag equal to 4 months. MSL over Mauritania is associated with wind speed over the UK on a lagging time of 12 months. Incorporating the above variables in the proposed model was found to be the key answer in generating accurate monthly wind speed predictions at the region of the UK.

The rationale behind these findings is in line with previous work that shows an inter-relationship between the meteorological variables [176–186, 188–191]. The converging evidence from these studies was that the aforementioned variables are linked to the anomalies of other atmospheric ones. The expectation that certain meteorological variables would affect wind speed over the UK led to a close look at the relationship between each of these variables and wind speed. According to Qu et al. [186] there are trends in the region of 75° N to 80° N between wind speed and SST. Qu et al. determined the correlation between the variables under investigation and concluded that there is a 3- month lag between SST and wind speed at the examined area of that study. Based on this evidence and drawing on the present data the same rationale was adopted in the procedure of determining the correlation between the different variables until a high correlation was recorded. Moreover, a correlation was identified between the variables used in this thesis at different geographical locations as opposed to the results presented by Qu el al. [186] which showed high correlation between variables for the same location. This decision making was also grounded in the body of work which indicates an inter-relationship between atmospheric variables from different geographical regions. At the same time, there are limited studies employing exogenous inputs for long-term forecasting whereas there are strong findings from relevant work on short-term predictions. In the research reported herein, this played a fundamental role in drawing on the theoretical framework of autoregressive-moving average models with exogenous inputs.

The steps for identifying the relationship between variables included a systematic and sequential testing of different pairs of "independent-dependent" variables covering different combinations both on spatial and temporal scale. This revealed different correlations depending on the combination of the following:

- A region-based pairing between each independent variable over the same region and the dependent variable over the UK, producing four pairs for each geographical region in the Atlantic ocean (i.e wind speed over the UK and wind speed (a), wind speed over the UK and SST (b) , wind speed over the UK and MSL (c), wind speed over the UK and SST gradient (d);
- The time lagged series of the independent variables (0-lag denotes present time, 1-lag denotes the value of the independent variable for $t - 1$, etc) and the wind speed over the UK at time t.

The results from the correlation coefficients determined which is the optimum combination on both spatial and temporal scale for each independent variable. To date, this is a novel approach for assessing the relationship between wind speed over the UK and other variables at different time lags and different geographical regions. The restriction from Good Energy to use a statistical model was in line with the histograms of the residuals from the above correlations dictating that such models can describe the relationship between the dependent and the independent variables. In parallel, the limitation of SARIMAX models in pre-assuming that the effect of the exogenous variable(s) on the predictand occurs at concurrent times was bypassed by shifting lags to the exogenous time series, as depicted from the correlation coefficients. These actions eventually begot the development of four different models each of which employed wind speed in the UK as endogenous input and correspondingly one of the four exogenous variables, as described above. The ability of the tested models to employ the effect of the independent variables on wind speed resulted in capturing better the characteristics of wind speed time series and hence to produce low statistical errors. During an inter-comparison of the four models it was revealed that the best predictor is MSL which converges with previous findings concerning other models [100, 111]. This corroborates the initial hypothesis supporting the feasibility of incorporating exogenous variables into a case of a SARIMA model to forecast monthly mean wind speed.

The principal contribution of this particular work lies in the use of exogenous variables into a specific case of SARIMA models for predicting monthly mean wind speed in the UK. These models in combination with the sub-objective of this study to create an RDBS with a long-term historic onshore data can be influential on our understanding and possible research trends. Specifically:

1. This approach can aid in developing different models and learning algorithms for generating monthly mean wind speed predictions. As mentioned previously, the drive for an evaluation of the optimum model would provide useful insights into different methodologies, their pros and cons, as well as contributing to the better understanding of how wind speed evolves through time and different locations. The limitation of such an attempt is clearly the fact that the results would be useful only for the UK.

2. The present findings improve understanding of the nature of this group of models in that they appear to be versatile since they have a great potential when using additional input

information.

3. There can be an influence on further understanding and application of this knowledge by adopting the specified independent variables to a different family of models such as ANN or SVM. This model can also be used in conjunction with other models in order to formulate a hybrid approach.

4. The strong association between the tested atmospheric variables and wind speed at different time lags constitutes a baseline for investigating the potential role of other variables in long-term wind speed forecasting as well.

Overall, the resultant focus on creating a database of historic wind speed records successfully served the objectives of this work. This database will lay the foundations for uniformity and homogeneity to occur in future work, by rendering it accessible to the research community. With the ability to test the various models using the same data and to develop different topologies and learning algorithms, concluding remarks will be drawn. Despite the importance of such work it is characterised by engaging in a complex multiprocess approach of model testing, which can be time and resource consuming. Moreover, the particular data is specific to the UK's climatic conditions and therefore one should interpret the findings with care. However, it is appreciated that such a future work will establish new routes for a new body of research, which will bring a considerable development in the general understanding of which models can capture the characteristics of wind speed time series. The community will then have greater potential to decide, grounded in a consensus, which model is optimum for long-term wind speed forecasting in the UK.

6.2 Limitations and Recommendations

Certain limitations are found across the research encountered mainly due to time constraints. A first limitation is that the size of the areas under investigation is equal to the geographical size of the UK. However, identifying different points across these areas would produce numerous pairs of variables and combinations, a procedure which was not attainable given the nature of the framework of this research.

One can also observe that this study did not experiment on increasing the number of the exogenous variables in each model, most importantly due to the initial intention of this work to provide a baseline for this type of model.

The boundaries of this research did not allow further and in-depth exploration of the meteorological variables' role in the correlation. However, the choice for using these variables was not arbitrary since it was based on existing meteorological research.

At this point, the necessity of exploring the following recommendations for further research is stressed:

- Convert wind speed prediction into wind power forecasts

  Information is available from Good Energy about Delabole wind farm which is located in the South West of England (50.635975, −4.705904). It consists of 10 Vestas WD34 (400kW) turbines. The scope in this recommendation is to reconstruct the wind speeds at Delabole wind farm by taking into account the best neighbours from the BADC-7 dataset. Wind speed series should be extrapolated to the hub height which for the case of Delabole farm is 32 m agl. Having extrapolated the wind speed at hub height the proposed model will generate the monthly wind speeds. Afterwards, by employing the power curve for the specific wind turbine the wind speed predictions would be converted to wind power. Such a study would serve as an evaluation of the model presented in this thesis and simultaneously would assess whether the proposed model is cost effective for turning into an operational tool.

- Hybrid model

  The proposed model of this thesis can be used in conjunction with an ANN or an SVM model in order to employ a hybrid approach which will aim to improve the forecasts.

- Identifying the best model

  The already built database can be used for testing different statistical models and different training algorithms. Readers should pay attention to section 2.2 in order to identify which models/methods have not yet been used on long-term wind speed forecasting.

6.3 Conclusion

The findings of the present work affirm the two level hypothesis that:

- there is a strong association between independent meteorological variables and wind speed for different time lags, and
- when both the dependent and independent variables are fed into a seasonal model it performs better than a model which does incorporate the same approach.

The present work intends to offer evaluative perspective on using a seasonal autoregressive moving average model with exogenous inputs for long-term wind speed forecasts. The major finding, that such a model achieved higher accuracy in the forecasts, allows this research to provide valid insights into the intriguing topic of long-term wind speed forecasting.

Appendix A
Mean Annual Wind Speed per Direction

This Appendix includes the graphs of mean annual wind speed by 30° direction for the remaining BADC-7 stations that was mentioned in Chapter 3.

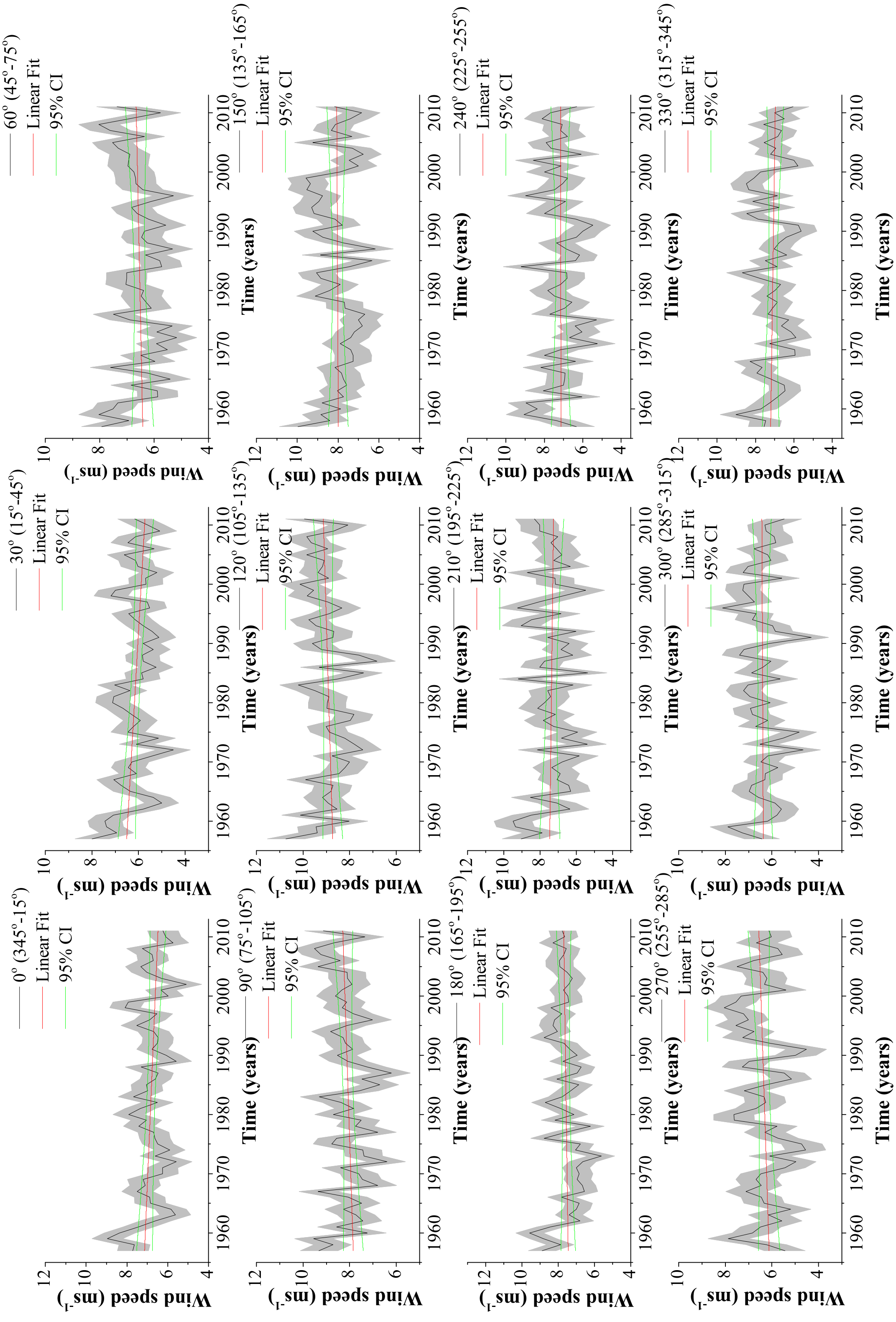


Figure A.1: *Mean annual wind speed by 30° direction for Lerwick*

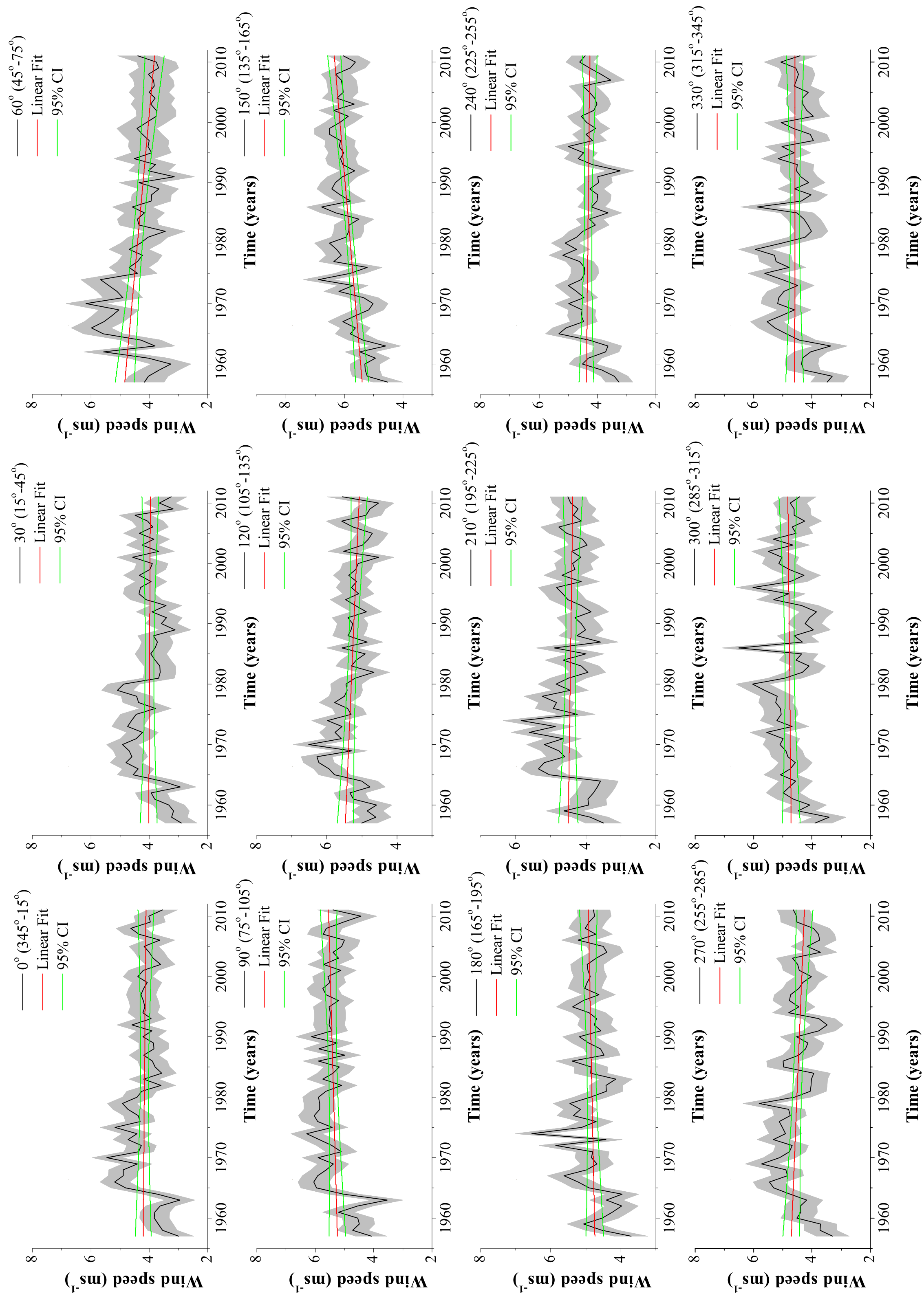


Figure A.2: *Mean annual wind speed by 30° direction for Boscombe Down*

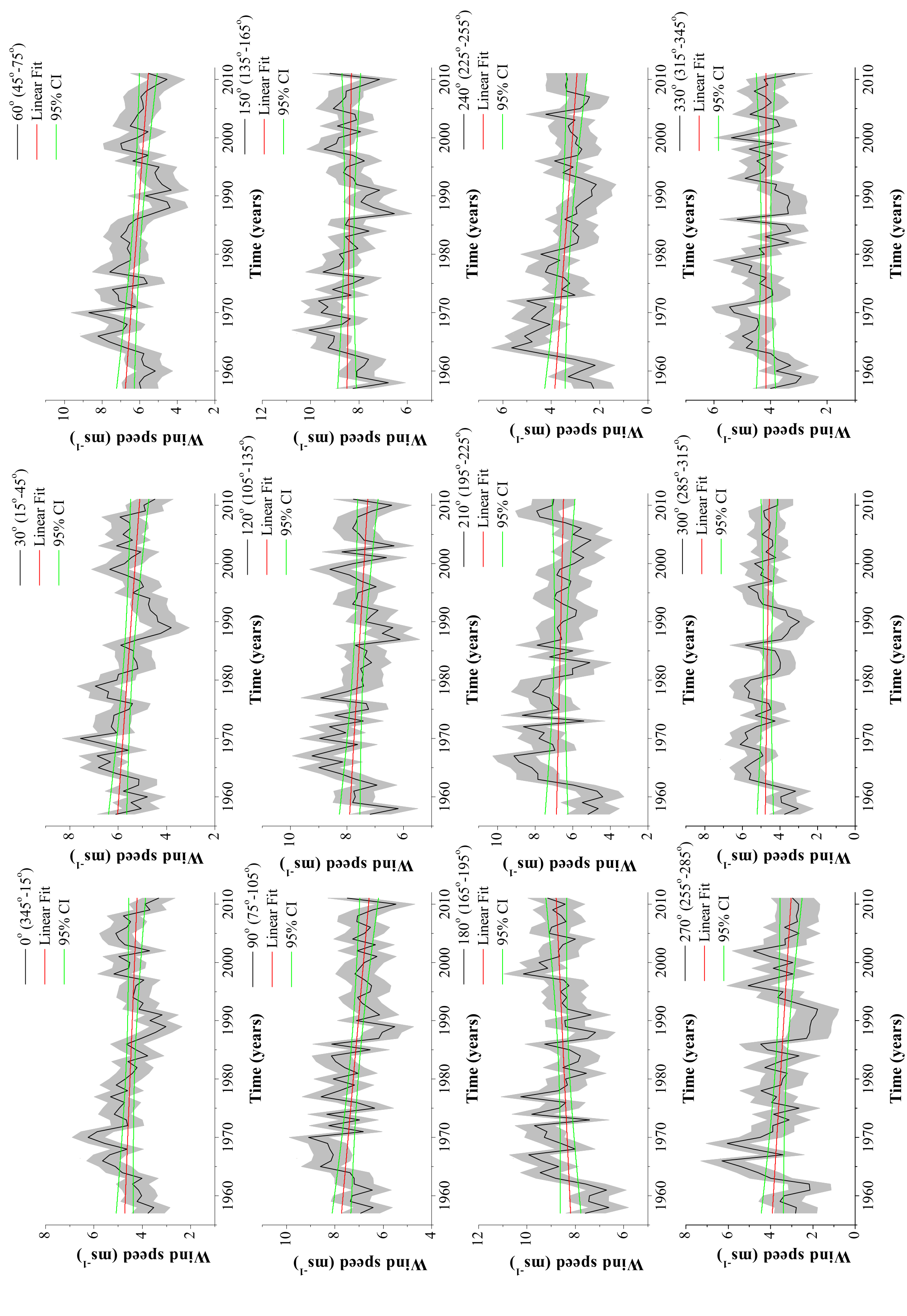


Figure A.3: *Mean annual wind speed by 30° direction for Valley*

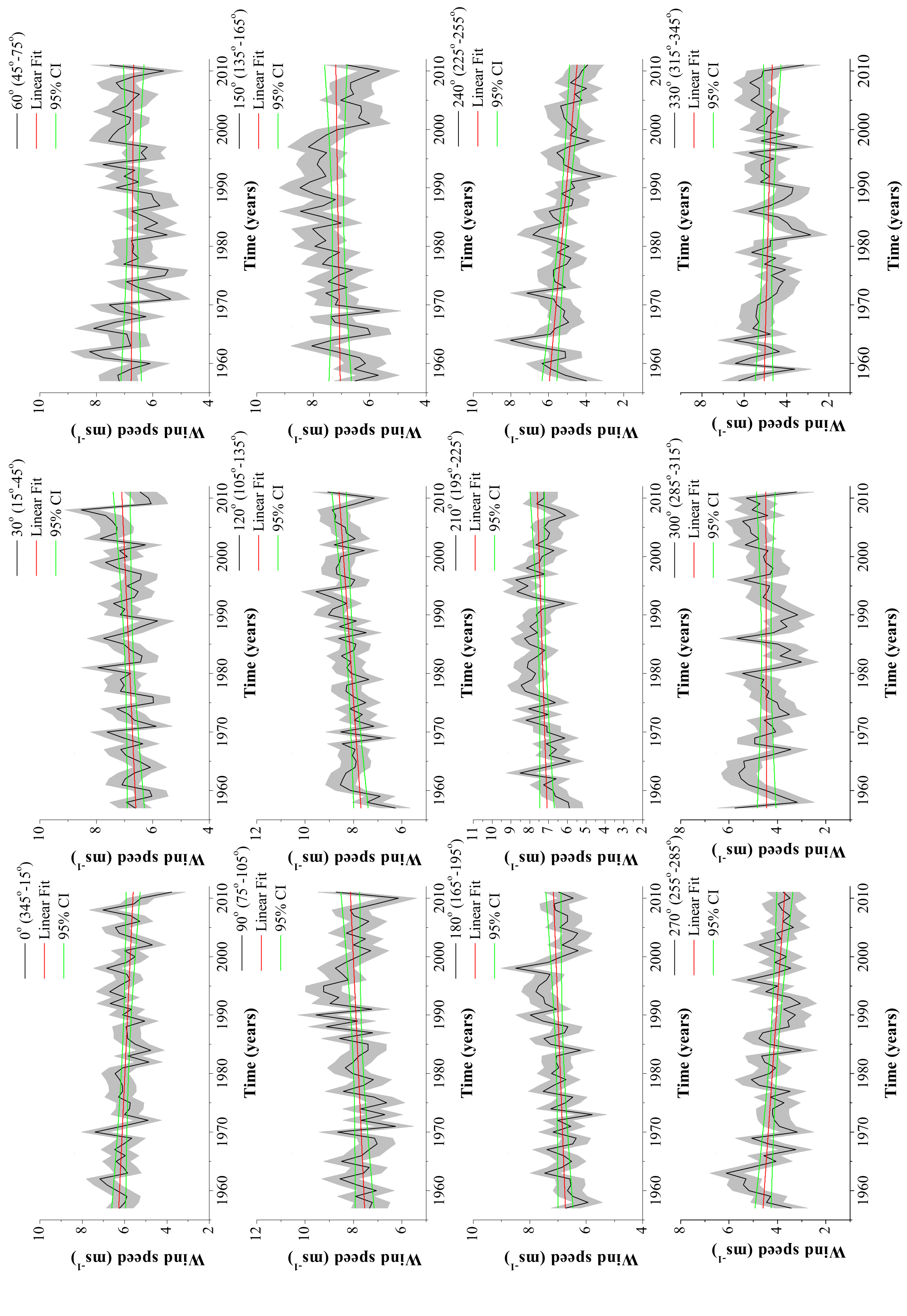


Figure A.4: *Mean annual wind speed by 30° direction for Aberporth*

Appendix B
Instruments and Height Measurement Changes

This Chapter includes information related to instrument changes for the BADC-57 stations that was mentioned in Chapter 3 (section 3.3.2).

Sites	Dates	Instrument	Effective height
Lerwick	01/01/1983 - 01/06/1984	Munro Mk4 anemometer	10
	01/06/1984 - 01/10/1999	Munro Mk4 anemometer	10
	01/10/1999 - 31/12/2009	Munro Mk6 anemometer	10
Sella Ness	01/01/1983 - 21/08/2009	Munro Mk4 anemometer	10
	21/08/2009 - 31/12/2009	Vector Mk6 anemometer	10
Wick Airport	01/01/1983 - 21/10/1999	Munro Mk4 anemometer	10
	21/10/1999 - 31/12/2009	Vector Mk6 anemometer	10
Stornoway Airport	01/01/1983 - 19/08/2002	Munro Mk5 anemometer	10
	19/08/2002 - 31/01/2009	Munro Mk6 anemometer	10
Aviemore	01/01/1983 - 20/09/1999	Munro Mk4 anemometer	10
	20/09/1999 - 31/01/2009	Munro Mk6 anemometer	10
Kinloss	01/01/1983 - 31/12/1992	Anemometer	10
	31/12/1992 - 24/05/1993	Assman Mk2 anemometer	10
	24/05/1993 - 01/12/1996	Munro Mk4 anemometer	10
	01/12/1996 - 15/03/2001	Munro Mk5 anemometer	10
	15/03/2001 - 31/12/2009	Vector Mk6 anemometer	10
Lossiemouth	01/01/1983 - 01/05/1992	Munro Mk4 anemometer	10
	01/05/1992 - 13/10/2006	Munro Mk5 anemometer	10
	13/10/2006 - 31/12/2009	Vector Mk6 anemometer	10
Dyce	01/01/1983 - 31/05/2000	Munro Mk4 anemometer	10
	31/05/2000 - 31/01/2009	Vector Mk6 anemometer	10
Peterhead Harbour	01/01/1983 - 17/12/1996	Assman Mk2 anemometer	15

continued on next page

continued from previous page			
Sites	Dates	Instrument	Effective height
	17/12/1996 - 11/09/2009	Assman Mk2 anemometer	10
	11/09/2009 - 31/12/2009	Vector Mk6 anemometer	10
Leuchars	01/01/1983 - 01/10/1990	Munro Mk5 anemometer	10
	01/10/1990 - 15/11/1994	Munro Mk5 anemometer	10
	15/11/1994 - 16/12/2000	Munro Mk5 anemometer	10
	16/12/2000 - 31/12/2009	Vector Mk6 anemometer	10
Boulmer	01/01/1983 - 01/01/1997	Munro Mk4 anemometer	10
	01/01/1997 - 03/05/2000	Munro Mk4 anemometer	10
	03/05/2000 - 31/12/2009	Vector Mk6 anemometer	10
Linton on Ouse	01/01/1983 - 01/12/1984	Munro Mk4 anemometer	10
	01/12/1984 - 31/12/2009	Munro Mk4 anemometer	10
Waddington	01/01/1983 - 27/01/1983	Cup Mk1 anemometer	10
	27/01/1983 - 27/03/2009	Munro Mk4 anemometer	10
	27/03/2009 - 31/12/2009	Vector Mk6 anemometer	10
Cranwell	01/01/1983 - 27/10/2009	Munro Mk4 anemometer	12
	27/10/2009 - 31/12/2009	Vector Mk6 anemometer	12
Coningsby	01/01/1983 - 31/01/2005	Munro Mk4 anemometer	10
	31/01/2005 - 31/12/2009	Vector Mk6 anemometer	10
Marham	01/01/1983 - 28/06/2000	Munro Mk4 anemometer	10
	28/06/2000 - 31/12/2009	Vector Mk6 anemometer	10
Gorleston	01/01/1983 - 31/12/2009	Munro Mk4 anemometer	10
Wattisham	01/01/1983 - 08/08/2000	Munro Mk4 anemometer	10
	08/08/2000 - 31/12/2009	Vector Mk6 anemometer	10
Bedford	01/01/1983 - 30/11/1999	Munro Mk4 anemometer	10
	30/11/1999 - 25/11/2008	Munro Mk4 anemometer	10
	25/11/2008 - 31/12/2009	Vector Mk6 anemometer	10
Bingley no2	01/01/1983 - 20/04/2000	Munro Mk4 anemometer	10
	20/04/2000 - 31/12/2009	Vector Mk6 anemometer	10
High Bradfield	01/01/1983 - 29/05/2009	Assman Mk2 anemometer	10
	29/05/2009 - 31/12/2009	Vector Mk6 anemometer	10
Church Fenton	01/01/1983 - 01/10/1983	Munro Mk4 anemometer	10
	01/10/1983 - 27/03/2001	Munro Mk4 anemometer	10
	27/03/2001 - 31/12/2009	Vector Mk6 anemometer	10
			continued on next page

continued from previous page

Sites	Dates	Instrument	Effective height
Nottingham, Watnall	01/01/1983 - 18/08/1999	Munro Mk4 anemometer	10
	18/08/199 - 31/12/2009	Vector Mk6 anemometer	10
Wittering	01/01/1983 - 07/06/2001	Munro Mk4 anemometer	10
	07/06/2001 - 31/12/2009	Vector Mk6 anemometer	10
Brize Norton	01/01/1983 - 23/07/1986	Munro Mk4 anemometer	10
	23/07/1986 - 11/03/2001	Munro Mk4 anemometer	10
	11/03/2001 - 31/12/2009	Vector Mk6 anemometer	10
Benson	01/01/1983 - 01/07/2000	Munro Mk4 anemometer	10
	01/07/2000 - 20/03/2007	Vector Mk6 anemometer	10
	20/03/2007 - 31/12/2009	Vector Mk6 anemometer	10
Shawbury	01/01/1983 - 01/03/1988	Munro Mk4 anemometer	10
	01/03/1988 - 28/11/2001	Munro Mk4 anemometer	10
	28/11/2001 - 31/12/2009	Vector Mk6 anemometer	10
Avonmouth	01/01/1983 - 15/04/1989	Munro Mk4 anemometer	10
	15/04/1989 - 17/06/2009	Munro Mk4 anemometer	10
	17/06/2009 - 31/12/2009	Vector Mk6 anemometer	10
Heathrow	01/01/1983 - 11/09/2002	Munro Mk5 anemometer	10
	11/09/2002 - 01/12/2003	Vector Mk6 anemometer	10
	01/12/2003 - 31/12/2009	Vector Mk6 anemometer	10
East Malling	01/01/1983 - 01/07/1998	Anemometer	10
	01/07/1998 - 31/12/2009	Vector Mk6 anemometer	10
Manston	01/01/1983 - 01/10/1988	Munro Mk4 anemometer	10
	01/10/1988 - 01/01/1996	Munro Mk4 anemometer	10
	01/01/1996 - 01/05/1999	Munro Mk4 anemometer	10
	01/05/1999 - 28/03/2003	Vector Mk6 anemometer	10
	28/03/2003 - 31/12/2009	Vector Mk6 anemometer	10
Hurn	N/A	N/A	N/A
Middle Wallop	01/01/1983 - 01/10/1987	Munro Mk4 anemometer	10
	01/10/1987 - 04/01/2001	Munro Mk4 anemomete	10
	04/01/2001 - 31/12/2009	Vector Mk6 anemometer	10
Lyneham	01/01/1983 - 01/06/1983	Munro Mk4 anemometer	13
	01/06/1983 - 22/07/2000	Munro Mk4 anemometer	10
	22/07/2000 - 31/12/2009	Vector Mk6 anemometer	10

continued on next page

continued from previous page

Sites	Dates	Instrument	Effective height
Larkhill	01/01/1983 - 26/09/1984	Munro Mk4 anemometer	10
	26/09/1984 - 29/02/2000	Munro Mk4 anemometer	10
	29/02/2000 - 27/04/2001	Vector Mk6 anemometer	10
	22/07/2000 - 31/12/2009	Vector Mk6 anemometer	10
Boscombe Down	01/01/1983 - 19/11/2009	Munro Mk4 anemometer	10
	19/11/2009 - 31/12/2009	Vector Mk6 anemometer	10
Machrihanish	01/01/1983 - 30/10/1992	Munro Mk4 anemometer	10
	30/10/1992 - 01/01/1996	Munro Mk4 anemometer	10
	01/01/1996 - 29/03/2001	Munro Mk4 anemometer	10
	29/03/2001 - 31/12/2009	Vector Mk6 anemometer	10
Dunstaffnage	01/01/1983 - 11/05/2000	Munro Mk4 anemometer	10
	11/05/2000 - 31/12/2009	Vector Mk6 anemometer	10
Salsburgh	01/01/1983 - 13/01/2000	Munro Mk4 anemometer	11
	13/01/2000 - 09/07/2000	Munro Mk4 anemometer	10
	09/07/2000 - 31/12/2009	Vector Mk6 anemometer	10
Eskdalemuir	01/01/1983 - 28/10/1999	Munro Mk4 anemometer	10
	28/10/1999 - 31/12/2009	Vector Mk6 anemometer	10
West Freugh	01/01/1983 - 01/12/1995	Munro Mk4 anemometer	10
	01/12/1995 - 26/02/2009	Munro Mk4 anemometer	10
	26/02/2009 - 31/12/2009	Vector Mk6 anemometer	10
Ronaldsway	01/01/1983 - 25/09/1999	Munro Mk4 anemometer	10
	25/09/1999 - 31/12/2009	AGI - Mk6 anemometer	10
Blackpool, Squires Gate	01/01/1983 - 15/11/2005	Munro Mk4 anemometer	11
	15/11/2005 - 31/12/2009	Munro Mk4 anemometer	11
Valley	01/01/1983 - 01/04/1988	Munro Mk4 anemometer	12
	01/04/1988 - 01/04/1995	Munro Mk4 anemometer	10
	01/04/1995 - 26/07/2007	Munro Mk5 anemometer	10
	26/07/2007 - 31/12/2009	Munro Mk5 anemometer	10
Bala	01/01/1983 - 10/02/2009	Munro Mk4 anemometer	10
	10/02/2009 - 31/12/2009	Vector Mk6 anemometer	10
Aberporth	01/01/1983 - 05/04/2000	Munro Mk4 anemometer	10
	05/04/2000 - 31/12/2009	Vector Mk6 anemometer	10
Milford Haven	01/01/1983 - 01/01/1985	Munro Mk4 anemometer	10

continued on next page

continued from previous page			
Sites	Dates	Instrument	Effective height
	01/01/1985 - 01/05/1995	Munro Mk4 anemometer	10
	01/05/1995 - 16/12/1999	Munro Mk4 anemometer	12
	16/12/1999 - 17/07/2002	Vector Mk6 anemometer	12
	17/07/2002 - 31/12/2009	Vector Mk6 anemometer	10
Yeovilton	01/01/1983 - 15/05/1986	Munro Mk4 anemometer	10
	15/05/1986 - 21/02/2008	Munro Mk5 anemometer	10
	21/02/2008 - 31/12/2009	Vector Mk6 anemometer	10
Plymouth, Mountbatten	01/01/1983 - 13/12/1999	Munro Mk4 anemometer	10
	13/12/1999 - 31/12/2009	Vector Mk6 anemometer	10
Chivenor	01/01/1983 - 15/12/1999	Munro Mk4 anemometer	10
	15/12/1999 - 31/12/2009	Vector Mk6 anemometer	10
Camborne	01/01/1983 - 01/10/1999	Munro Mk4 anemometer	10
	01/10/1999 - 31/12/2009	Vector Mk6 anemometer	10
Aldergrove	01/01/1983 - 24/01/2003	Munro Mk4 anemometer	10
	24/01/2003 - 31/12/2009	Vector Mk6 anemometer	10
Ballypatrick Forest	01/01/1983 - 01/04/1984	Munro Mk4 anemometer	10
	01/04/1984 - 01/01/1992	Munro Mk4 anemometer	10
	01/01/1992 - 10/02/2009	Munro Mk4 anemometer	10
	10/02/2009 - 31/12/2009	Vector Mk6 anemometer	10
Orlock Head	01/01/1983 - 01/07/1993	Munro Mk4 anemometer	12
	01/07/1993 - 31/12/2009	Vector Mk6 anemometer	12
Leeming	01/01/1983 - 25/04/1991	Munro Mk4 anemometer	10
	25/04/1991 - 13/05/2008	Munro Mk5 anemometer	10
	13/05/2008 - 31/12/2009	Vector Mk6 anemometer	10
Tiree	01/01/1983 - 11/07/2001	Munro Mk4 anemometer	10
	11/07/2001 - 31/12/2009	Vector Mk6 anemometer	10

Table B.1: *List of of BADC-57 stations*

Appendix C
The BADC Surface Wind Speed Indices

Time (years)	BADC-7 index normalised to 55-year period	BADC-7 index normalised to 29-year period	BADC-57 index normalised to 29-year period
1957	0.987	0.999	-
1958	0.915	0.927	-
1959	0.987	1.000	-
1960	0.967	0.979	-
1961	1.040	1.054	-
1962	1.017	1.031	-
1963	1.031	1.044	-
1964	1.048	1.061	-
1965	1.017	1.030	-
1966	1.047	1.060	-
1967	1.109	1.124	-
1968	0.958	0.970	-
1969	0.977	0.989	-
1970	1.046	1.060	-
1971	0.981	0.994	-
1972	1.018	1.031	-
1973	0.984	0.997	-
1974	1.059	1.073	-
1975	0.978	0.991	-
1976	0.995	1.008	-
1977	1.065	1.079	-
1978	1.019	1.033	-
1979	1.037	1.051	-
1980	1.025	1.038	-

continued on next page

continued from previous page

Time (years)	BADC-7 index normalised to 55-year period	BADC-7 index normalised to 29-year period	BADC-57 index normalised to 29-year period
1981	1.041	1.055	-
1982	1.037	1.050	-
1983	1.026	1.039	1.050
1984	0.956	0.969	0.981
1985	0.945	0.958	0.978
1986	1.062	1.076	1.100
1987	0.891	0.903	0.920
1988	0.955	0.968	0.998
1989	0.966	0.979	0.984
1990	1.028	1.042	1.077
1991	0.933	0.945	0.971
1992	0.971	0.984	1.016
1993	0.988	1.000	1.018
1994	1.044	1.058	1.056
1995	0.991	1.004	1.002
1996	0.988	1.001	0.989
1997	0.970	0.983	0.958
1998	1.052	1.066	1.040
1999	1.019	1.032	1.022
2000	1.002	1.015	1.007
2001	0.924	0.936	0.944
2002	1.006	1.019	0.998
2003	0.957	0.969	0.953
2004	0.984	0.997	0.992
2005	1.011	1.024	1.003
2006	0.999	1.012	0.989
2007	0.999	1.012	0.998
2008	1.025	1.038	1.035
2009	0.993	1.006	0.985
2010	0.931	0.944	0.907
2011	1.010	1.023	1.032

Table C.1: *The BADC Surface Wind Speed Indices*

Appendix D
Annual Regional Wind Indices by Season

This Chapter includes information related to annual wind indices using the BADC-57 stations over the period 1983-2011 that was mentioned in Chapter 4. The wind speed indices are broken down by season:

- winter: December - January - February
- spring: March - April - May
- summer: June - July - August
- autumn: September - October - November

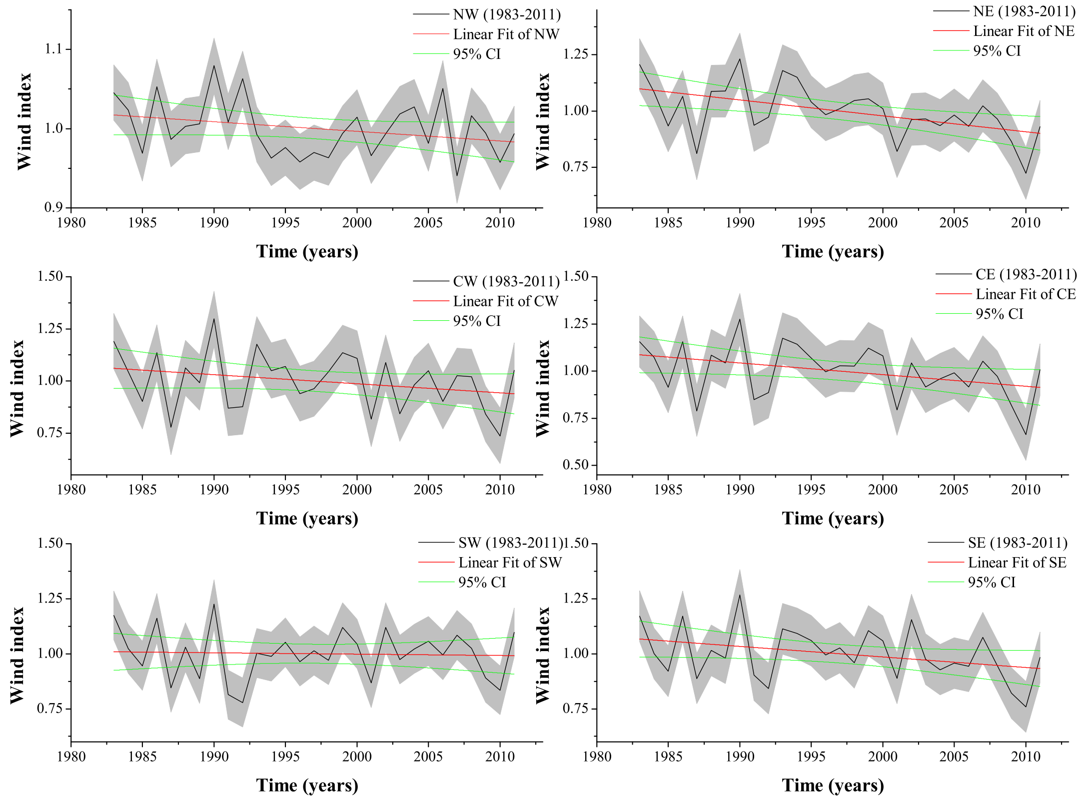


Figure D.1: *Regional winter wind index*

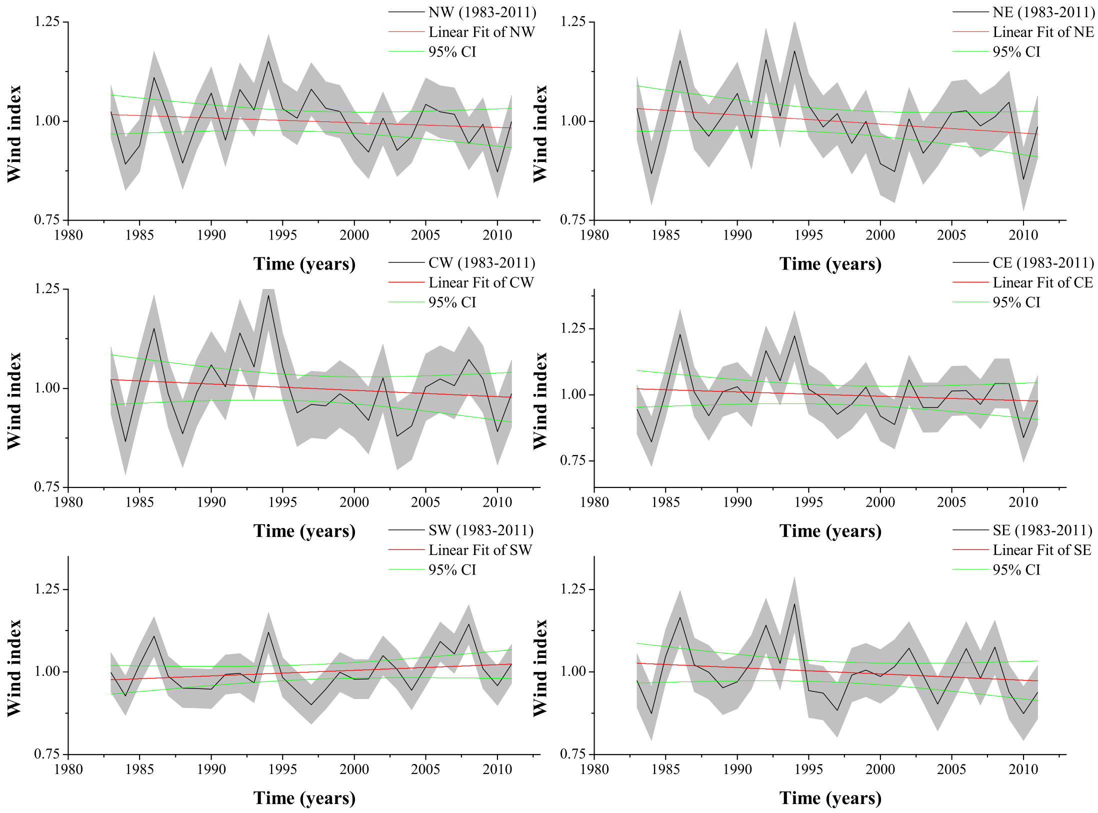


Figure D.2: *Regional spring wind index*

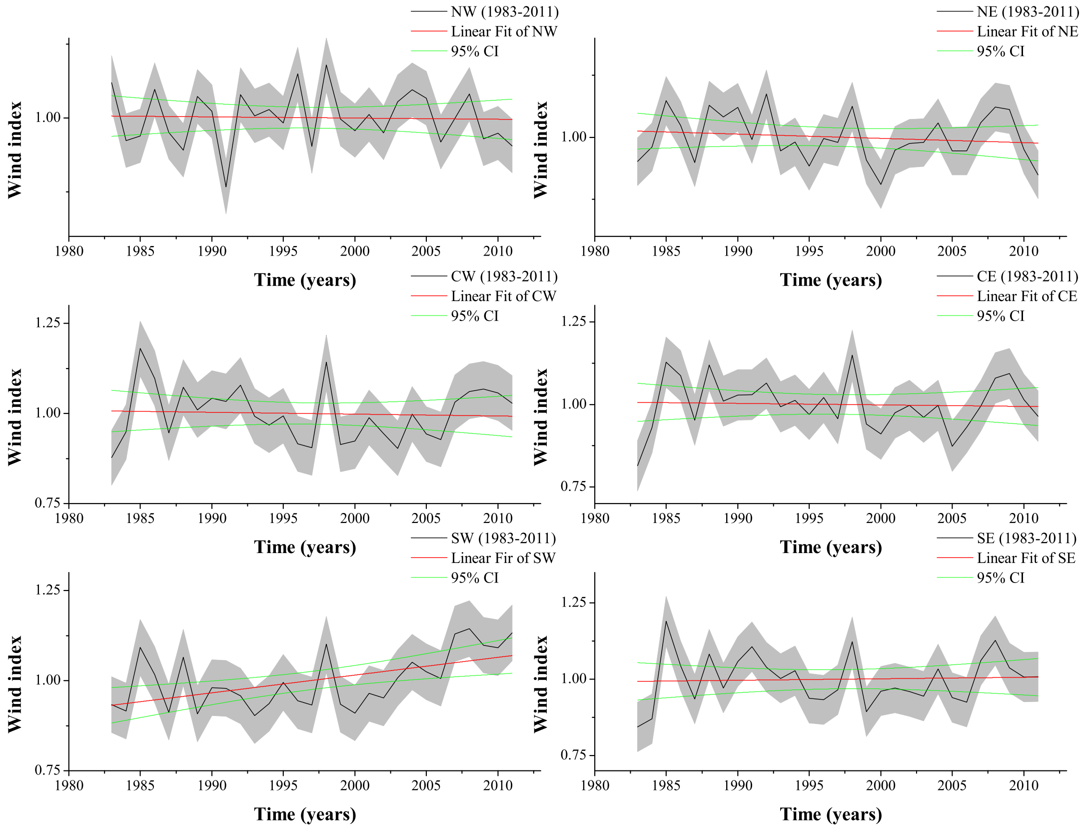


Figure D.3: *Regional summer wind index*

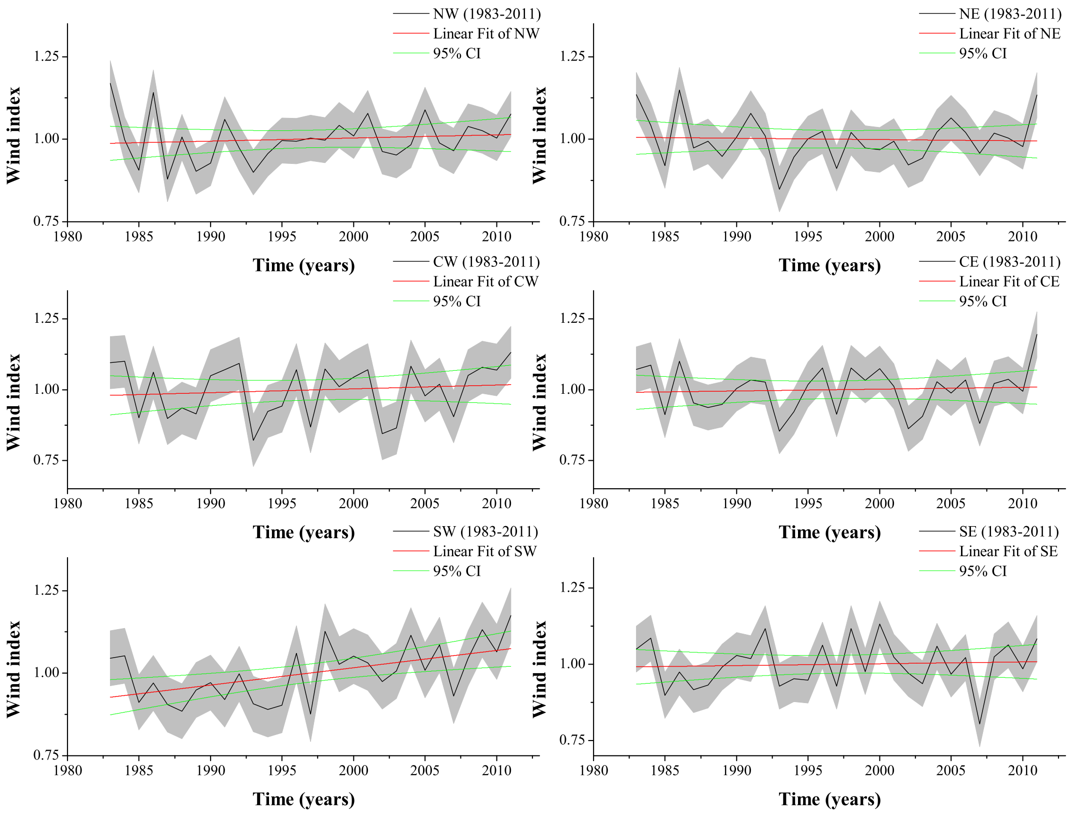


Figure D.4: *Regional autumn wind index*

Appendix E
Augmented Dickey Fuller Test

This Chapter includes information related to Augmented Dickey Fuller Test that was mentioned in Chapter 5 (section 5.1.2).
Consider a stochastic process of the form:

$$y_t = \rho y_{t-1} + \epsilon_t, \quad \forall t \in \mathbb{N} \tag{E.1}$$

Where y_t is the endogenous variable of the process, t is the time index, ρ is a parameter of the model and ϵ_t is the irregular component, or else *white noise*.

Solving for ϵ_t, equation (E.1) $\Longrightarrow$

$$(1 - \rho L)y_t = \epsilon_t \tag{E.2}$$

Then the process is stationary if the root of the characteristic equation, $(1 - \rho L) = 0 > 1$. Solving the characteristic equation for L we get that $L = 1/\rho > 1$ for $0 < \rho < 1$.

Testing if the root is greater than 1 is not a straightforward task since prior to the test, ρ has to be estimated ($\hat{\rho}$). This can be done by estimating equation (E.1) and by using the hypothesis test:

$$H_0 : \hat{\rho} = 1 \;\; \text{non-stationary} \tag{E.3}$$

Against the alternative:

$$H_1 : \hat{\rho} < 1 \ \text{ stationary} \tag{E.4}$$

Thus, for equation (E.2), $\hat{\rho}$ is estimated as in Ref. [174]:

$$\hat{\rho_r} = \sum y_t y_{t-1} \left(\sum y_{t-1}{}^2 \right)^{-1} \tag{E.5}$$

However, the problem in the standard unit root hypothesis testing, lies to the fact that, under the null hypothesis, the test does not follow the usual t-distribution. Thus, instead of the standard test, we use the DF test or its *augmented* version ADF (if more than one lag is included). In DF test, rather than estimating the stochastic process as in equation E.1, the same regression model is estimated by incorporating the differenced series $y_t - y_{t-1}$:

$$y_t - y_{t-1} = \rho y_{t-1} + \epsilon_t - y_{t-1} \overset{(??)}{\Longrightarrow} \nabla y_t = \hat{\tau} y_{t-1} + \epsilon_t \tag{E.6}$$

Where $\hat{\tau} = \rho - 1$. Then the equation (E.6) is tested for the hypothesis:

$$H_0 : \hat{\tau} = 0 \ \text{ non-stationary} \tag{E.7}$$

Against the alternative:

$$H_1 : \hat{\tau} < 0 \ \text{ stationary} \tag{E.8}$$

The DF test is valid only for first order autocorrelation of y due to its principal condition which needs ϵ_t to be uncorrelated. For higher orders of y we employ the ADF test. The general form of an AR(p) model is defined as in equation 2.6.

According to equation (2.6), an AR(2) model is defined as:

$$
\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \\
&\xRightarrow{\text{(simultaneously adding and subtracting } \phi_2 y_{t-1} \text{ in the right side)}} \\
&= c + (\phi_1 + \phi_2) y_{t-1} - \phi_2 (y_{t-1} - y_{t-2}) + \epsilon_t \\
&\xRightarrow{\text{(subtracting } y_{t-1} \text{ in both sides)}} \\
y_t - y_{t-1} &= c + (\phi_1 + \phi_2) y_{t-1} - \phi_2 (y_{t-1} - y_{t-2}) + \epsilon_t - y_{t-1} \\
&\xRightarrow{(??)} \\
\nabla y_t &= c + (\phi_1 + \phi_2 - 1) y_{t-1} - \phi_2 (\nabla y_{t-1}) + \epsilon_t \\
&\xRightarrow{\text{(substituting } \phi_1 + \phi_2 - 1 = \psi \wedge \phi_2 = \alpha_1)} \\
\nabla y_t &= c + \psi y_{t-1} - \alpha_1 \nabla y_{t-1} + \epsilon_t
\end{aligned}
\tag{E.9}
$$

Thus, equation (E.9) can be extended to its generic form when a model is of the order of p:

$$
\nabla y_t = c + \psi y_{t-1} - \sum_{i=0}^{p} \alpha_i \nabla y_{t-i} + \epsilon_t \tag{E.10}
$$

Similarly to the DF test, the augmented unit root test is also carried out under the hypothesis:

$$
H_0 : \hat{\psi} = 0 \;\; \text{non-stationary} \tag{E.11}
$$

Against the alternative:

$$
H_1 : \hat{\psi} < 0 \;\; \text{stationary} \tag{E.12}
$$

The value for the test statistic is computed according to the equation (E.13) and compared afterwards against the critical values (C.V.) provided in [198]. If the value of the DF_τ < C.V., then the null hypothesis H_0 of $\psi = 0$ is rejected and no unit root is present.

$$\mathrm{DF}_{\tau} = \frac{\hat{\psi}}{\mathrm{SE}(\hat{\psi})} \tag{E.13}$$

Appendix F
Correlation Coefficients

This Chapter includes information related to correlation coefficient between wind speed over the UK and variables from the ERA-40 dataset that was mentioned in Chapter 5.

F.1 Tables for Correlation Coefficients

	cor	**p**	**Lat**	**Lon**
same month	0.893	0	50 to 61	−19 to −9
1-month lag	0.605	0	50 to 61	−19 to −9
2-month lag	0.323	0	62 to 73	−52 to −42
3-month lag	–	–	–	–
4-month lag	–	–	–	–
5-month lag	0.083	0.054	14 to 25	−19 to −9
6-month lag	0.350	0	14 to 25	−19 to −9
7-month lag	0.515	0	14 to 25	−19 to −9
8-month lag	0.572	0	14 to 25	−19 to −9
9-month lag	0.440	0	14 to 25	−19 to −9
10-month lag	0.541	0	26 to 37	−63 to −53
11-month lag	0.702	0	38 to 49	−63 to −53
12-month lag	0.720	0	62 to 73	−41 to −31

Table F.1: *Correlation Coefficients for Actual Wind Speed at the Reference Grid and Wind Speed from the ERA-40 Dataset*

	cor	**p**	**Lat**	**Lon**
same month	–	–	–	–
1-month lag	0.194	0	14 to 25	−41 to 31
2-month lag	0.516	0	14 to 25	−41 to −31
3-month lag	0.693	0	14 to 25	−41 to −31
4-month lag	0.753	0	26 to 37	−63 to −53
5-month lag	0.749	0	26 to 37	−85 to −75
6-month lag	0.589	0	26 to 37	−85 to −75
7-month lag	0.261	0	26 to 37	−85 to −75
8-month lag	–	–	–	–
9-month lag	–	–	–	–
10-month lag	–	–	–	–
11-month lag	–	–	–	–
12-month lag	–	–	–	–

Table F.2: *Correlation Coefficients for Actual Wind Speed at the Reference Grid and SST from the ERA-40 Dataset*

	cor	p	Lat	Lon
same month	0.701	0	14 to 25	−19 to −9
1-month lag	0.575	0	14 to 25	−19 to −92
2-month lag	0.318	0	14 to 25	−19 to −9
3-month lag	0.204	0	38 to 49	−74 to −64
4-month lag	0.385	0	38 to 49	−63 to −53
5-month lag	0.445	0	50 to 61	−52 to −42
6-month lag	0.49	0	50 to 61	−52 to −42
7-month lag	0.44	0	62 to 73	−8 to 2
8-month lag	0.511	0	14 to 25	−52 to −42
9-month lag	0.48	0	14 to 25	−41 to −31
10-month lag	0.432	0	14 to 25	−30 to −20
11-month lag	0.549	0	14 to 25	−19 to −9
12-month lag	0.672	0	14 to 25	−19 to −9

Table F.3: *Correlation Coefficients for Actual Wind Speed at the Reference Grid and MSL from the ERA-40 Dataset*

F.2 Tables of Differenced Series for Correlation Coefficients

	cor	p	Lat	Lon
same month	0.704	0	50 to 61	−19 to −9
1-month lag	0.177	0	50 to 61	−19 to −9
2-month lag	0.112	0.009	26 to 37	−30 to −20
3-month lag	0.106	0.014	26 to 37	−41 to −31
4-month lag	0.156	0	26 to 37	−30 to −20
5-month lag	0.086	0.048	26 to 37	−19 to −9
6-month lag	0.081	0.065	62 to 73	−8 to 2
7-month lag	0.084	0.055	38 to 49	−52 to −42
8-month lag	0.041	0.354	26 to 37	−63 to −53
9-month lag	0.108	0.014	14 to 25	−30 to −20
10-month lag	0.087	0.047	26 to 37	−19 to −9
11-month lag	0.136	0.002	38 to 49	−52 to −42
12-month lag	0.082	0.064	38 to 49	−52 to −42

Table F.4: *Seasonal Differences of Table F.1*

	cor	p	Lat	Lon
same month	0.113	0.009	26 to 37	−74 to −64
1-month lag	0.106	0.014	62 to 73	−52 to −42
2-month lag	0.143	0.001	62 to 73	−52 to −42
3-month lag	0.141	0.001	62 to 73	−52 to −42
4-month lag	0.126	0.003	26 to 37	−52 to −42
5-month lag	0.133	0.002	26 to 37	−52 to −42
6-month lag	0.079	0.069	14 to 25	−63 to −53
7-month lag	0.14	0.001	50 to 61	−52 to −42
8-month lag	0.132	0.002	62 to 73	−52 to −42
9-month lag	0.118	0.007	62 to 73	−52 to −42
10-month lag	0.159	0	62 to 73	−52 to −42
11-month lag	0.149	0	62 to 73	−52 to −42
12-month lag	0.119	0.006	50 to 61	−52 to −42

Table F.5: *Seasonal Differences of Table F.2*

	cor	p	Lat	Lon
same month	0.448	0	38 to 49	−19 to −9
1-month lag	0.133	0.002	26 to 37	−19 to −9
2-month lag	0.077	0.079	26 to 37	−74 to −64
3-month lag	0.06	0.172	50 to 61	−19 to −9
4-month lag	0.127	0.004	26 to 37	−52 to −42
5-month lag	0.083	0.058	38 to 49	−30 to −20
6-month lag	0.056	0.201	38 to 49	−85 to −75
7-month lag	0.051	0.247	14 to 25	−63 to −53
8-month lag	0.054	0.221	14 to 25	−63 to −53
9-month lag	0.092	0.036	14 to 25	−30 to −20
10-month lag	0.092	0.036	62 to 73	−41 to −31
11-month lag	0.132	0.003	14 to 25	−63 to −53
12-month lag	0.274	0	62 to 73	−8 to 2

Table F.6: *Seasonal Differences of Table F.3*

F.3 Tables for Correlation Coefficients between Wind Speed and SST Gradients

	cor	p	Lat	Lon
same month	0.508	0	38 to 49	−85 to −75
1-month lag	0.416	0	26 to 37	−85 to −75
2-month lag	0.306	0	50 to 61	−19 to −9
3-month lag	0.59	0	50 to 61	−19 to −9
4-month lag	0.696	0	50 to 61	−19 to −9
5-month lag	0.664	0	14 to 25	−52 to −42
6-month lag	0.599	0	14 to 25	−30 to −20
7-month lag	0.47	0	14 to 25	−19 to −9
8-month lag	0.391	0	26 to 37	−52 to −42
9-month lag	0.607	0	38 to 49	−63 to −53
10-month lag	0.73	0	38 to 49	−63 to −53
11-month lag	0.71	0	38 to 49	−85 to −75
12-month lag	0.517	0	38 to 49	−74 to −64

Table F.7: *Correlation Coefficients between Wind Speed at Reference Grid and SST Gradients from the ERA-40 Dataset*

Appendix G
SARIMAX Forecasts for the BADC-7 Stations

This Chapter includes the graphs of the monthly mean wind speed predictions for the remaining BADC-7 stations that were mentioned in Chapter 5.

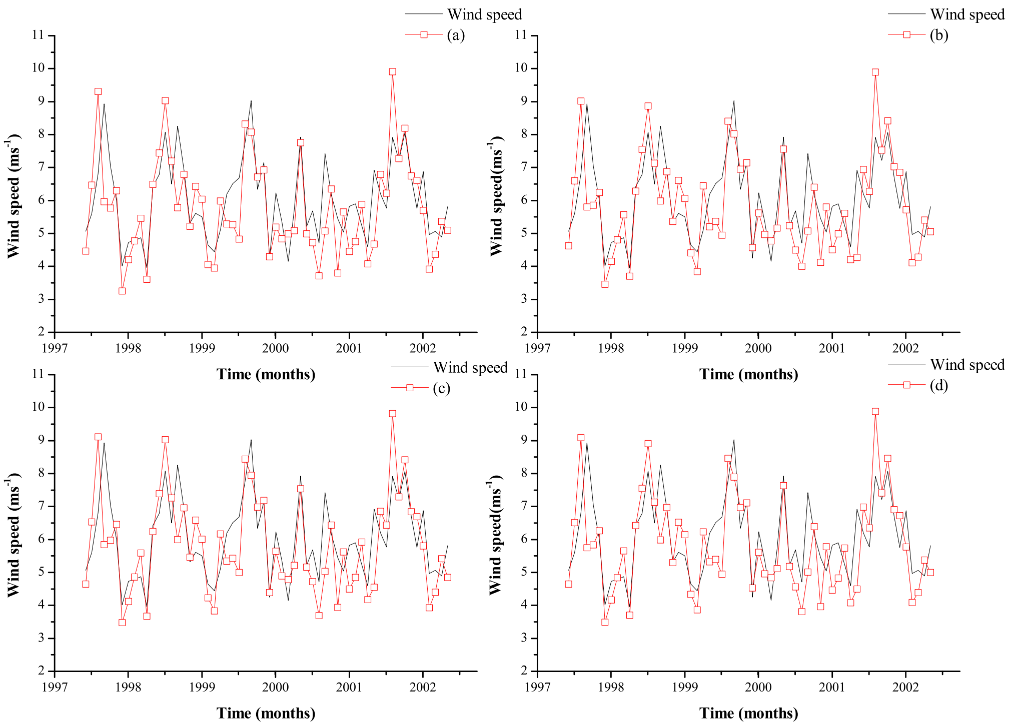


Figure G.1: *Predictions of SARIMAX models at Stornoway Airport*

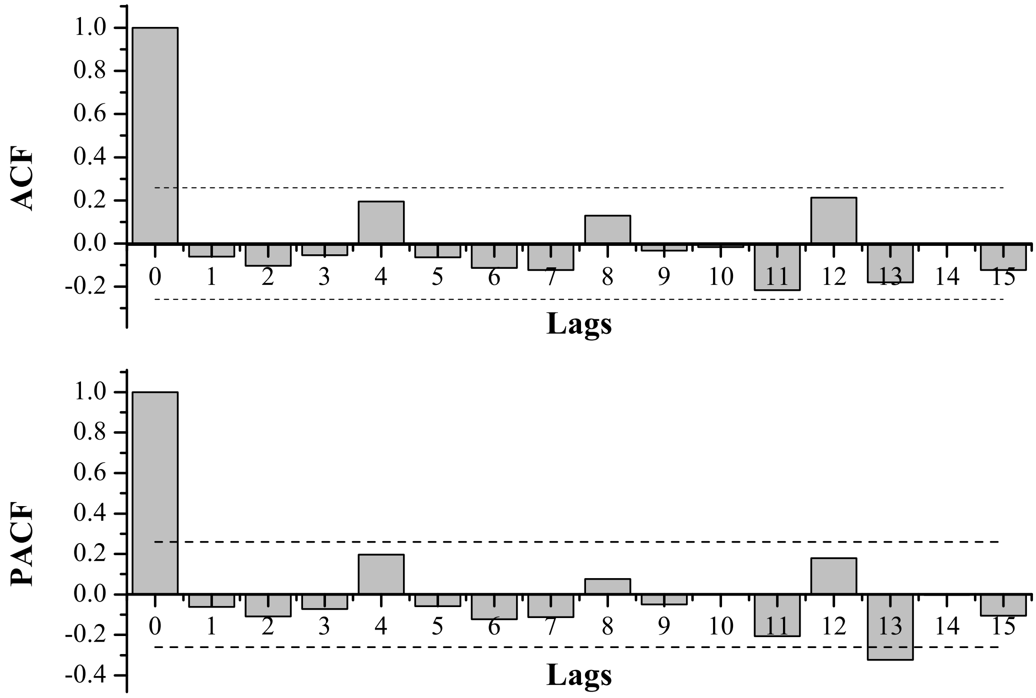


Figure G.2: *Correlograms of residuals at Stornoway Airport*

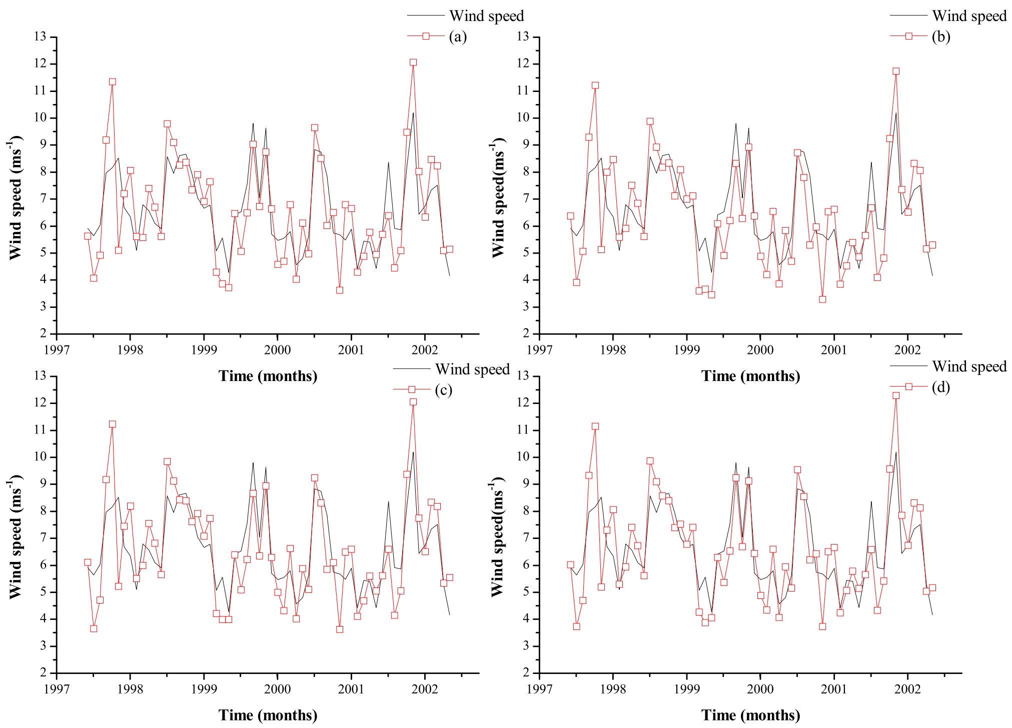


Figure G.3: *Predictions of SARIMAX models at Valley*

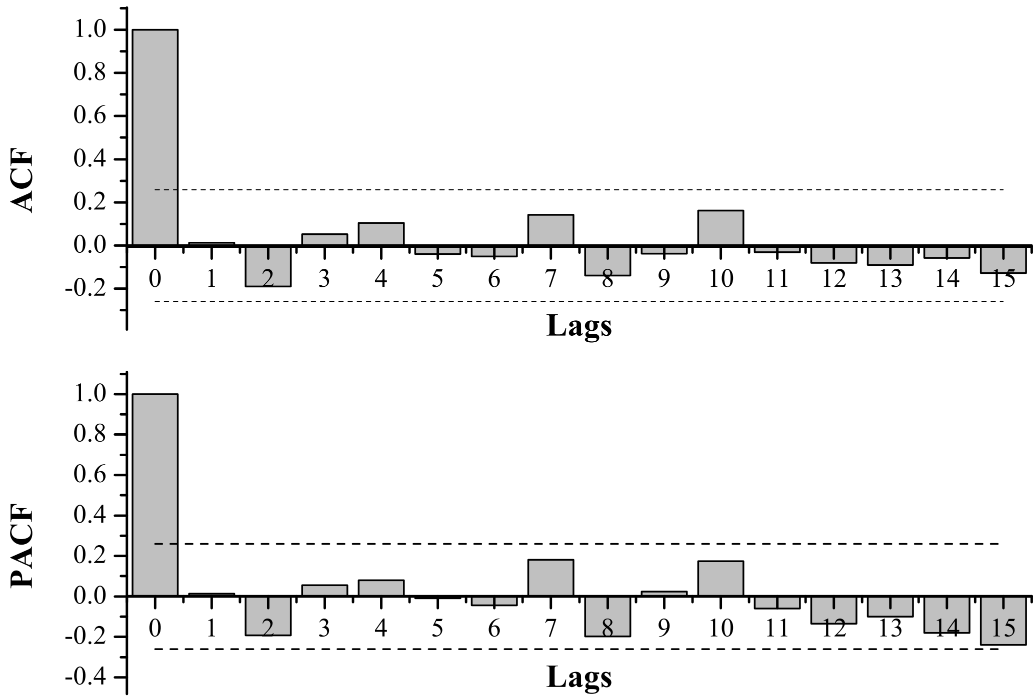


Figure G.4: *Correlograms of residuals at Valley*

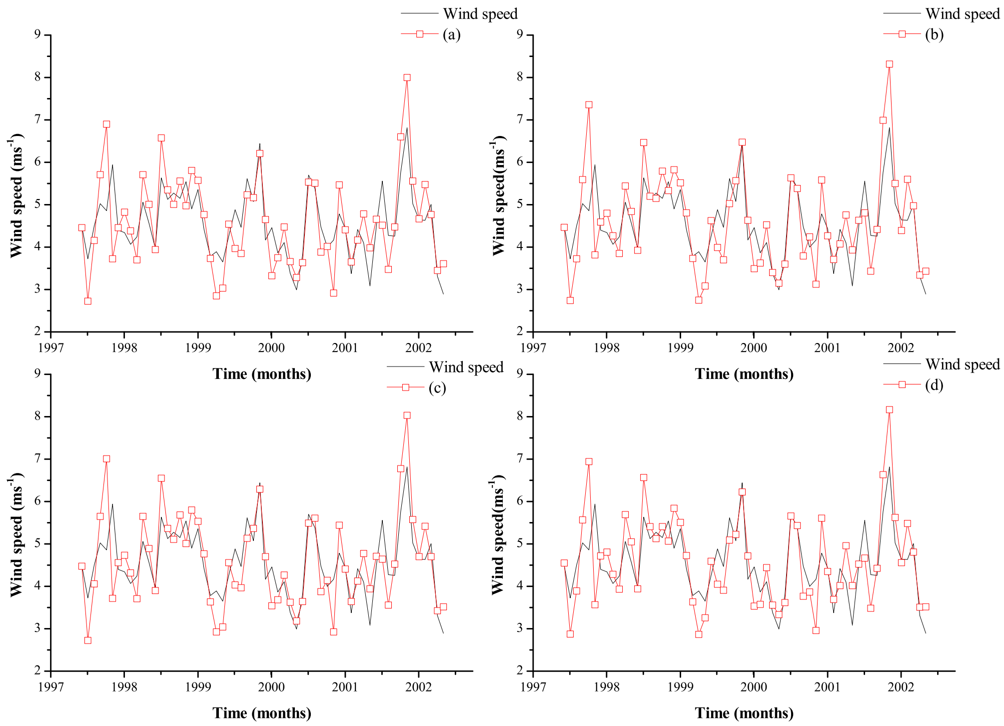


Figure G.5: *Predictions of SARIMAX models at Aldergrove*

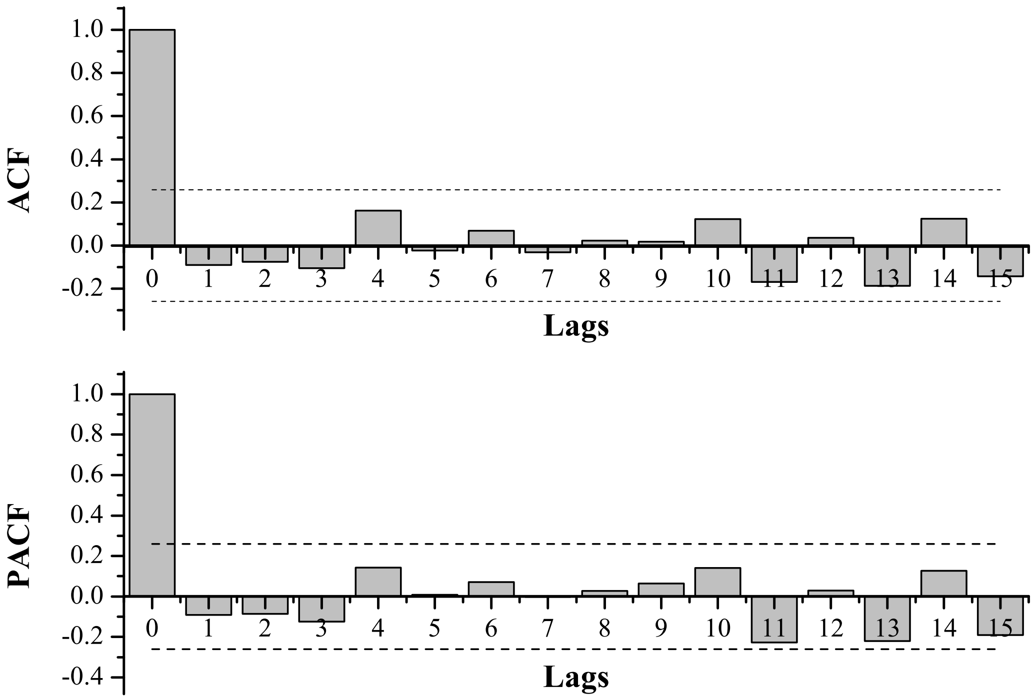


Figure G.6: *Correlograms of residuals at Aldergrove*

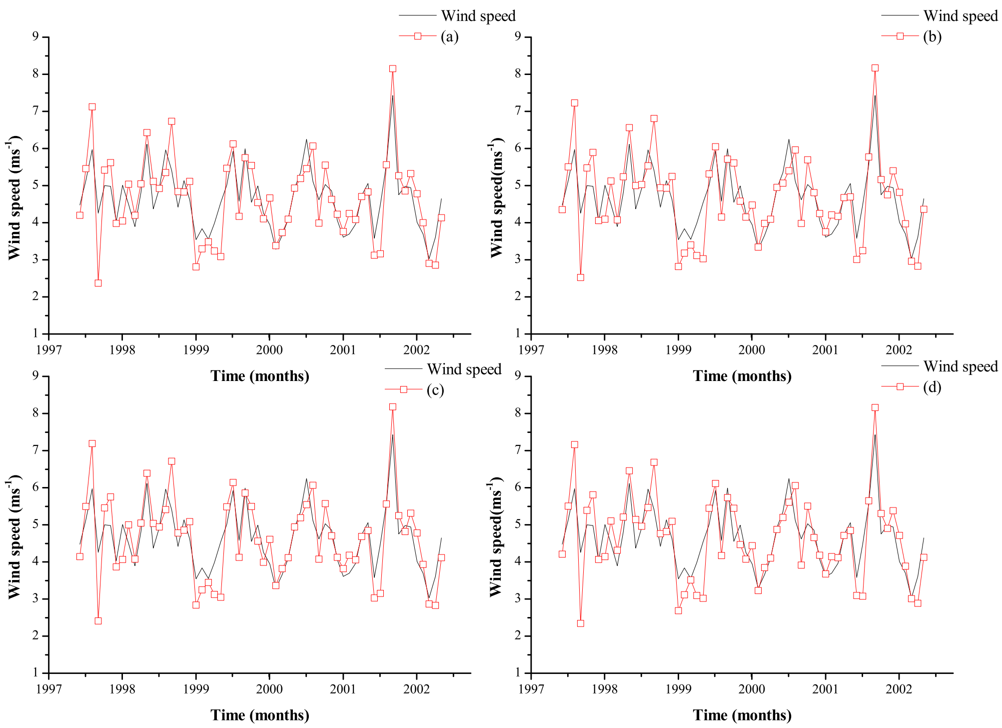


Figure G.7: *Predictions of SARIMAX models at Boscombe Down*

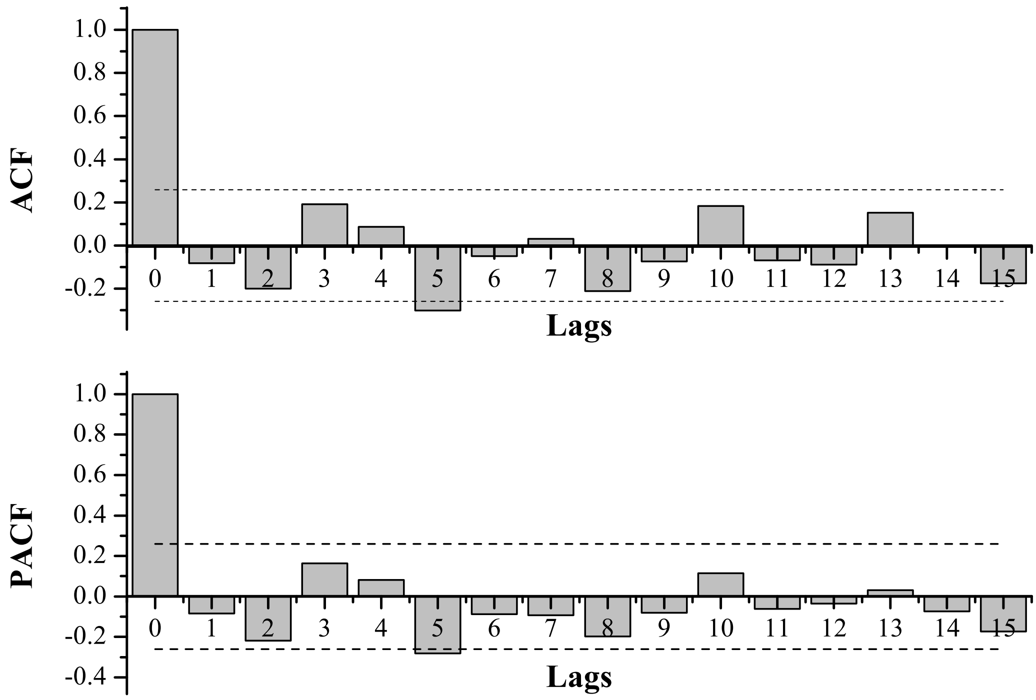


Figure G.8: *Correlograms of residuals at Boscombe Down*

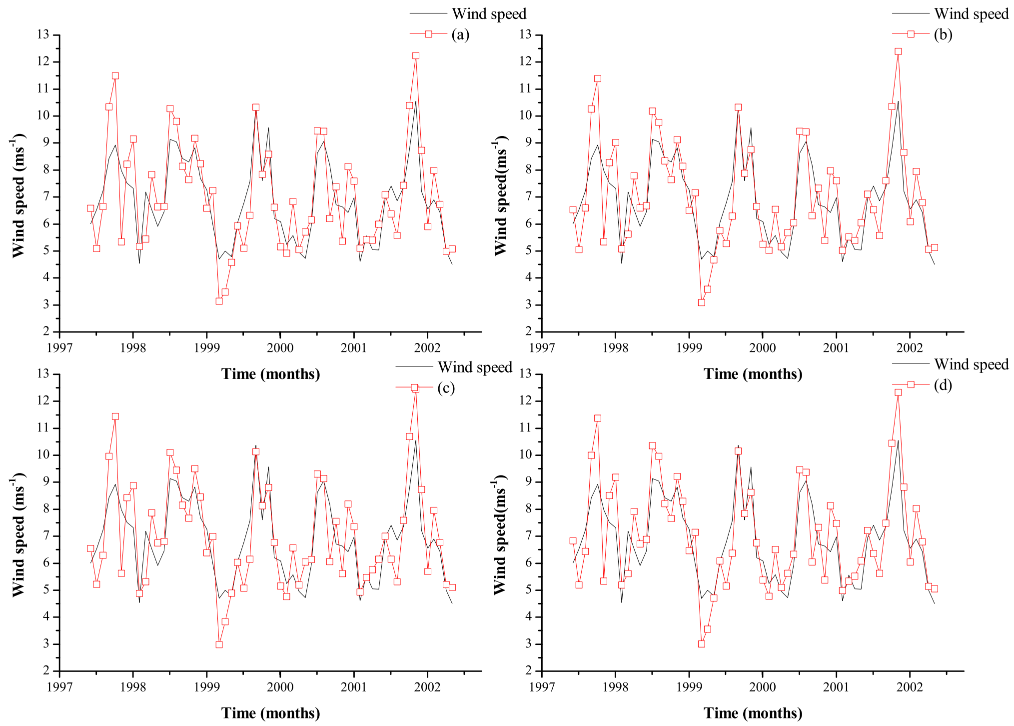


Figure G.9: *Predictions of SARIMAX models at Aberporth*

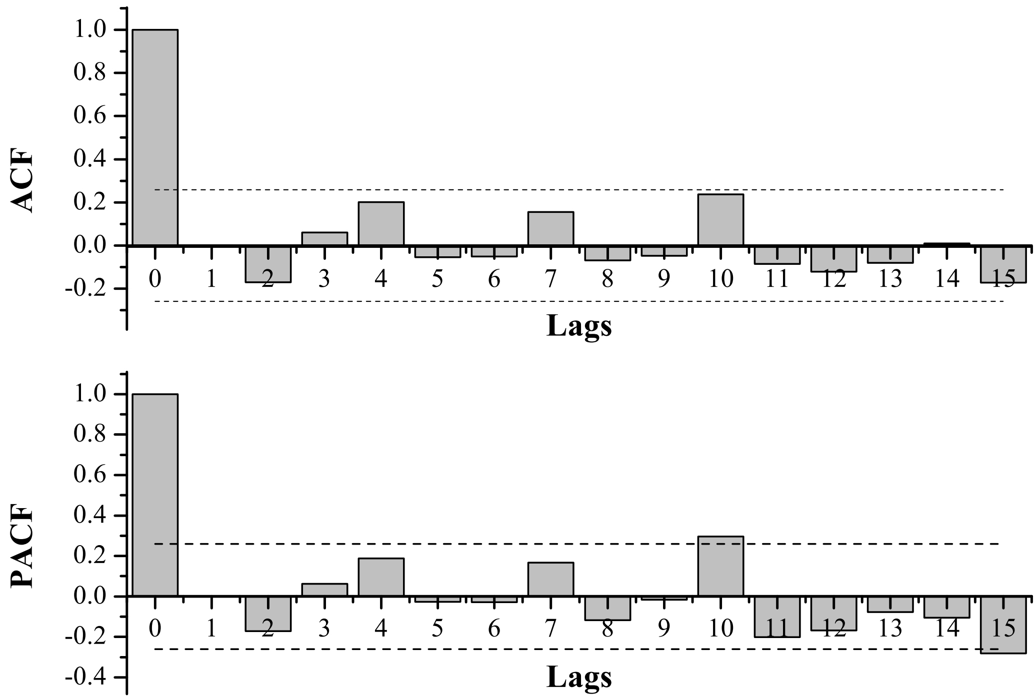


Figure G.10: *Correlograms of residuals at Aberporth*

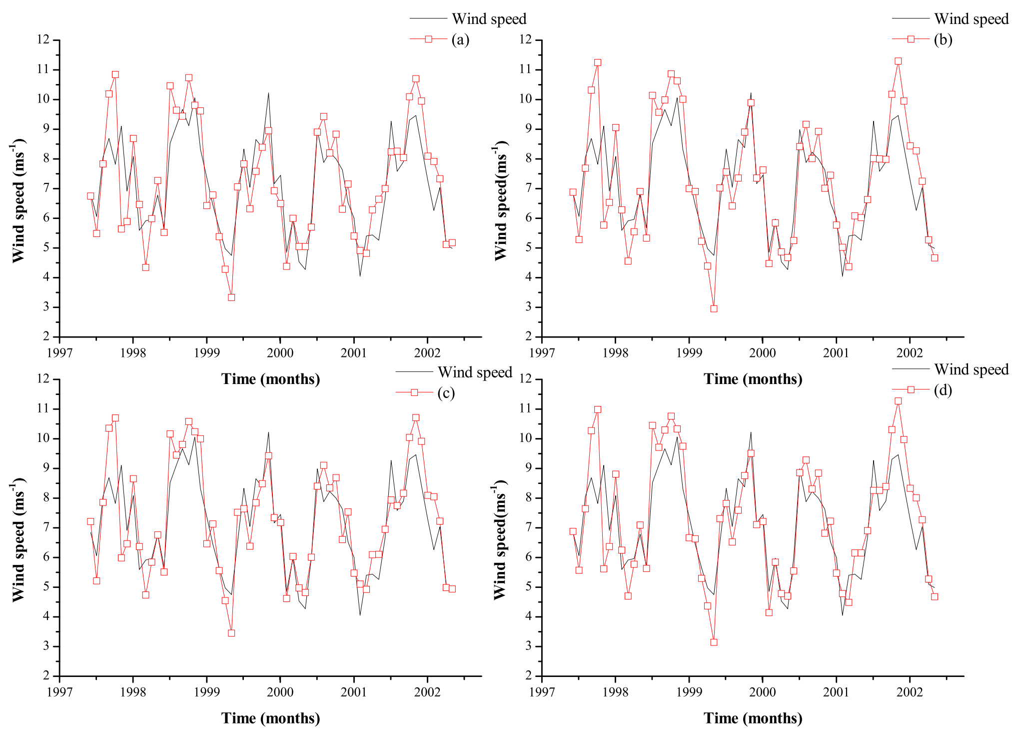


Figure G.11: *Predictions of SARIMAX models at Tiree*

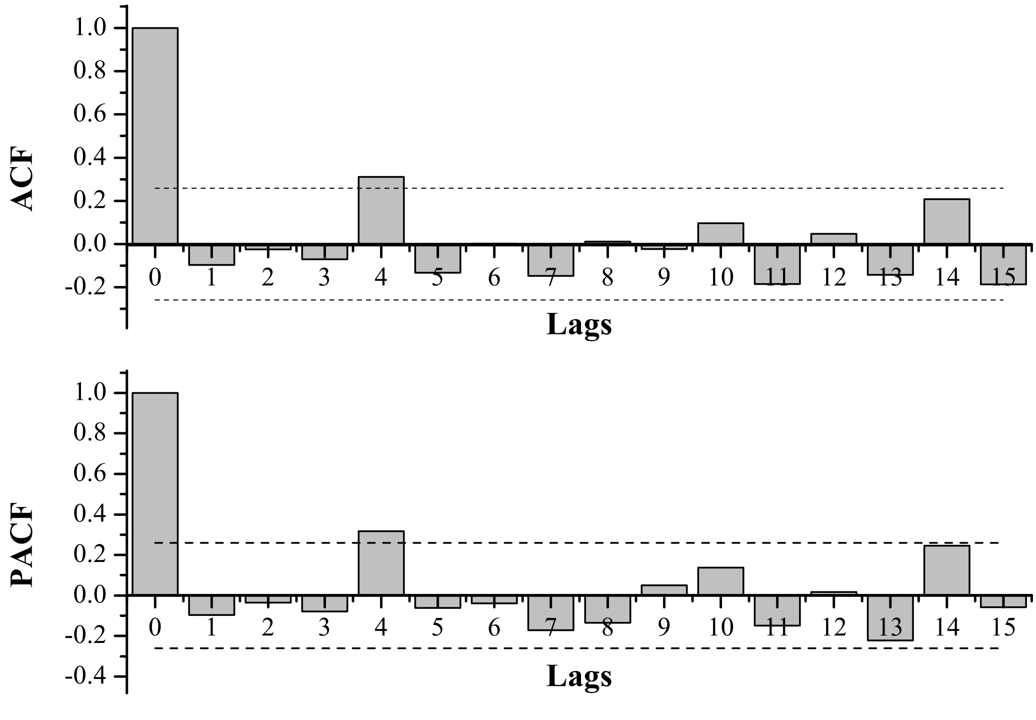


Figure G.12: *Correlograms of residuals at Tiree*

Appendix H
SARIMAX Forecasts for the BADC-7 Stations

This Chapter includes the statistical measures of the monthly mean wind speed predictions for the remaining BADC-7 stations that were mentioned in Chapter 5 (section 5.4).

H.1 Stornoway Airport

Cases	Order	MSE	ME
(a)	$(5,1,1,1)(0,1,1)_{12}$	1.127	0.272
(b)	$(9,1,1,1)(0,1,1)_{12}$	1.113	0.2
(c)	$(4,1,1,1)(0,1,1)_{12}$	1.089	0.21
(d)	$(7,1,1,1)(0,1,1)_{12}$	1.118	0.206

Table H.1: *Order of the best SARIMAX at Stornoway Airport and corresponding MSE*

H.2 Valley

Cases	Order	MSE	ME
(a)	$(10,1,1,1)(0,1,1)_{12}$	1.336	0.017
(b)	$(7,1,1,1)(0,1,1)_{12}$	1.516	0.206
(c)	$(7,1,1,1)(0,1,1)_{12}$	1.352	0.206
(d)	$(3,1,1,1)(0,1,1)_{12}$	1.244	0.206

Table H.2: *Order of the best SARIMAX at Valley and corresponding MSE*

H.3 Aldergrove

Cases	Order	MSE	ME
(a)	$(1,1,1,1)(0,1,1)_{12}$	0.475	0.206
(b)	$(4,1,1,1)(0,1,1)_{12}$	0.515	0.206
(c)	$(1,1,1,1)(0,1,1)_{12}$	0.464	−0.046
(d)	$(5,1,1,1)(0,1,1)_{12}$	1.485	−0.046

Table H.3: *Order of the best SARIMAX at Aldergrove and corresponding MSE*

H.4 Boscombe Down

Cases	Order	MSE	ME
(a)	$(1,1,1,1)(0,1,1)_{12}$	0.387	−0.007
(b)	$(9,1,1,1)(0,1,1)_{12}$	0.4	−0.033
(c)	$(1,1,1,1)(0,1,1)_{12}$	0.386	0.001
(d)	$(1,1,1,1)(0,1,1)_{12}$	0.397	−0.002

Table H.4: *Order of the best SARIMAX at Boscombe Down and corresponding MSE*

H.5 Aberporth

Cases	Order	MSE	ME
(a)	$(6,1,1,1)(0,1,1)_{12}$	1.204	−0.128
(b)	$(3,1,1,1)(0,1,1)_{12}$	1.111	−0.133
(c)	$(8,1,1,1)(0,1,1)_{12}$	1.221	−0.13
(d)	$(5,1,1,1)(0,1,1)_{12}$	1.203	−0.153

Table H.5: *Order of the best SARIMAX at Aberporth and corresponding MSE*

H.6 Tiree

Cases	**Order**	**MSE**	ME
(a)	$(3,1,1,1)(0,1,1)_{12}$	1.096	−0.131
(b)	$(8,1,1,1)(0,1,1)_{12}$	1.146	−0.183
(c)	$(0,1,1,1)(0,1,1)_{12}$	0.943	−0.182
(d)	$(7,1,1,1)(0,1,1)_{12}$	1.076	−0.194

Table H.6: *Order of the best SARIMAX at Tiree and corresponding MSE*

References

[1] K. E. Boulding, "The Economics of the Coming Spaceship Earth," *Environmental Quality in a Growing Economy*, pp. 3–14, 1966.

[2] J. Hansen, R. Ruedy, M. Sato, and K. Lo, "Global Surface Temperature Change," *Reviews of Geophysics*, vol. 48, no. 4, p. RG4004, 2010.

[3] J. Hansen, D. Johnson, A. Lacis, S. Lebedeff, P. Lee, D. Rind, and G. Russell, "Climate Impact of Increasing Atmospheric Carbon Dioxide," *Science*, vol. 213, pp. 957–966, 1981.

[4] D. Ren, "Effects of global warming on wind energy availability," *Journal of Renewable and Sustainable Energy*, vol. 2, no. 5, p. 052301, 2010.

[5] A. P. Sokolov, P. H. Stone, C. E. Forest, R. Prinn, M. C. Sarofim, M. Webster, S. Paltsev, C. A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. Reilly, C. Wang, B. Felzer, J. M. Melillo, and H. D. Jacoby, "Probabilistic Forecast for Twenty-First-Century Climate Based on Uncertainties in Emissions (Without Policy) and Climate Parameters," *Journal of Climate*, vol. 22, no. 19, pp. 5175–5204, 2009.

[6] A. P. Sokolov, P. H. Stone, C. E. Forest, R. Prinn, M. C. Sarofim, M. Webster, S. Paltsev, C. A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. Reilly, C. Wang, B. Felzer, J. M. Melillo, and H. D. Jacoby, "CORRIGENDUM," *Journal of Climate*, vol. 23, no. 8, pp. 2230–2231, 2010.

[7] MIT, "Climate change odds much worse than thought." http://web.mit.edu/newsoffice/2009/roulette-0519.html, May 19 2009. Last checked: November 2011.

[8] European Parliament and the Council of the European Union, "DIRECTIVE 2001/77/EC on the promotion of electricity produced from renewable energy sources in the internal electricity market," *Official Journal of the European Communities*, 2001.

[9] Department of Trade and Industry (DTI), "Meeting the Energy Challenge: A White Paper on Energy." http://www.berr.gov.uk/, May 2007. Last checked: November 2011.

[10] European Parliament and the Council of the European Union, "DIRECTIVE 2009/28/EC on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC," *Official Journal of the European Union*, 2009.

[11] The Guardian, "Greenhouse gases rise by record amount." http://www.guardian.co.uk/, November 4 2011. Last checked: November 2011.

[12] European Wind Energy Association (EWEA), "Wind in power. 2012 European statistics." http://ewea.org, February 2013. Last checked: July 2013.

[13] Global Wind Energy Council (GWEC), "Global Wind Report - Annual Market Update 2010." http://www.gwec.net, 2010. Last checked: December 2012.

[14] Global Wind Energy Council (GWEC), "Global Wind Report - Annual Market Update 2012." http://www.gwec.net, 2013. Last checked: July 2013.

[15] Smith, M., "Crucial role of wind highlighted on eve of Chancellor's Autumn Statement." RenewableUK, December 2012. Last checked: December 2012.

[16] Casey, Z., "What is the cost of energy?." European Wind Energy Association (EWEA), December 2012. Last checked: December 2012.

[17] T. S. Schmidt, R. Born, and M. Schneider, "Assessing the costs of photovoltaic and wind power in six developing countries," *Nature Climate Change*, vol. 2, pp. 1–6, April 15 2012.

[18] European Wind Energy Association (EWEA), "Green Growth - The impact of wind energy on jobs and the economy." http://www.ewea.org, April 2012. Last checked: December 2012.

[19] US Energy Department, "Offshore Wind Market and Economic Analysis." http://www1.eere.energy.gov, November 2012. Last checked: December 2012.

[20] Cambridge Econometrics, "A study into the Economics of Gas and Offshore Wind." http://www.wwf.org.uk, November 2012. Last checked: December 2012.

[21] Electricity Network Association, "ENA ER G59 Generator Connection Application Form." http://www.energynetworks.org/electricity/engineering/distributed-generation/distributed-generation.html. Last checked: June 2013.

[22] National Grid, "British Electricity Trading and Transmission Arrangements (BETTA)." http://www.nationalgrid.com/uk. Last checked: December 2010.

[23] Private communication with Good Energy, March 6 2009. Discussion with Mr. Nick Haines.

[24] G. N. Bathurst, J. Weatherill, and G. Strbac, "Trading wind generation in short term energy markets," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 782–789, 2002.

[25] I. Martí, G. Kariniotakis, P. Pinson, I. Sanchez, T. S. Nielsen, H. Madsen, G. Giebel, J. Usaola, A. M. Palomares, R. Brownsword, J. Tambke, U. Focken, M. Lange, G. Sideratos, and G. Descombes, "Evaluation of Advanced Wind Power Forecasting Models - Results of the Anemos Project," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (Athens, Greece), February 27 - March 2 2006.

[26] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and H. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.

[27] G. Giebel, L. Landberg, G. Kariniotakis, and R. Brownsword, "State-of-the-Art on Methods and Software Tools for Short-Term Prediction of Wind Energy Production," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (Madrid, Spain), June 16-19 2003.

[28] L. Landberg, G. Giebel, H. A. Nielsen, T. S. Nielsen, and H. Madsen, "Short-Term Prediction - an Overview," *Wind Energy*, vol. 6, no. 3, pp. 273–280, 2003.

[29] R. Vautard, J. Cattiaux, P. Yiou, J.-N. Thépaut, and P. Ciais, "Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness," *Nature Geosciences*, vol. 3, no. 11, pp. 756–761, 2010.

[30] P. Kritharas and S. J. Watson, "Long Term Forecasting of Wind Speed Using Historical Patterns," in *European Wind Energy Conference and Exhibition (EWEC)*, (Marseille, France), March 16–19 2009.

[31] D. Weisser and T. J. Foxon, "Implications of seasonal and diurnal variations of wind velocity for power output estimation of a turbine: a case study of Grenada," *International Journal of Energy Research*, vol. 27, no. 13, pp. 1165–1179, 2003.

[32] D. J. Brayshaw, A. Troccoli, R. Fordham, and J. Methven, "The impact of large scale atmospheric circulation patterns on wind power generation and its potential predictability: A case study over the UK," *Renewable Energy*, vol. 36, pp. 2087 – 2096, August 2011.

[33] D. KEMA, "Reducing Uncertainty in Wind Project Energy Estimates," white paper, 2012.

[34] B. Hodgetts, "A Statistical Review of Recent Wind Speed Trends in the UK," in *Proceedings of RenewableUK*, (Manchester, UK), October 25–27 2011.

[35] N. Atkinson, K. Harman, M. Lynn, A. Schwarz, and A. Tindal, "Long-Term Wind Speed Trends in NorthWestern Europe," in *BWEA28 Conference Proceedings*, (Glasgow, UK), October 10-12 2006.

[36] R. E. Luna and W. H. Church, "Estimation of Long-Term Concentrations Using a "Universal" Wind Speed Distribution," *Journal of Applied Meteorology*, vol. 13, no. 8, pp. 910–916, 1974.

[37] J. P. Hennessey Jr., "Some Aspects of Wind Power Statistics," *Journal of Applied Meteorology*, vol. 16, no. 2, pp. 119–128, 1977.

[38] W. E. Bardsley, "Note on the Use of the Inverse Gaussian Distribution for Wind Energy Applications," *Journal of Applied Meteorology*, vol. 19, no. 9, pp. 1126–1130, 1980.

[39] C. G. Justus, W. R. Hargraves, A. Mikhail, and D. Graber, "Methods for estimating wind speed frequency distributions," *Journal of Applied Meteorology and Climatology*, vol. 17, no. 3, pp. 350–353, 1978.

[40] J. A. Carta, P. Ramírez, and S. Velázquez, "A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 5, pp. 933–955, 2009.

[41] K. Pearson and A. Lee, "On the generalised probable error in multiple normal correlation," *Biometrika*, vol. 6, no. 1, pp. 59–68, 1908.

[42] G. J. Connor, "Statistical short-term forecasting technique for wind and temperature in a coastal tropical location (Townsville)," *Australian Meteorological Magazine*, vol. 46, no. 4, pp. 257–265, 1997.

[43] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden-Day, 1970.

[44] M. G. Kendall, "On Autoregressive Time Series," *Biometrika*, vol. 33, no. 2, pp. 105–122, 1944.

[45] M. S. Miranda and R. W. Dunn, "One-hour-ahead wind speed prediction using a Bayesian methodology," in *IEEE Power Engineering Society General Meeting, 2006.*, pp. 6 pp.–, 2006.

[46] H. M. Geerts, "Short range prediction of wind speeds: a system-theoretic approach," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (Hamburg, Germany), pp. 594–599, October 22-26 1984.

[47] E. A. Bossanyi, "Short-Term Stochastic Wind Prediction and Possible Control Applications," in *Proceedings of the Delphi Workshop on "Wind Energy-Applications"*, (Delphi, Greece), May 20-22 1985.

[48] E. A. Bossanyi, "Stochastic Wind Prediction for Wind Turbine System Control," in *Proceedings of the BWEA Wind Energy Conference (British Wind Energy Association)* (A. D. Garrad, ed.), (Oxford, UK), pp. 219–226, Mechanical Engineering Publ Ltd, London, England, March 27-29 1985.

[49] A. Balouksis, D. Tsanakas, and G. Vachtsevanos, "Stochastic simulation of hourly and daily average wind speed sequences," *Wind Engineering*, vol. 10, no. 1, pp. 1–11, 1986.

[50] M. Milligan, M. Schwartz, and Y. Wan, "Statistical wind power forecasting models: Results for U.S. wind farms," in *WINDPOWER*, May 18–21 2003.

[51] Durán, M. and Cros, D. and Riquelme, J., "Short-Term Wind Power Forecast Based on ARX Models," *Journal of Energy Engineering*, vol. 133, no. 3, pp. 172–180, 2007.

[52] U. S. Jensen, E. Pelgrum, and H. Madsen, "The development of a forecasting model for the prediction of wind power production to be used in central dispatch centers.," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (Thessaloniki, Greece), pp. 353–356, October 10-14 1994.

[53] H. Madsen, T. S. Nielsen, and H. A. Nielsen, "Prediction of wind power," in *Proceedings of the Second IEA Joint Action Symposium on Wind Forecasting Techniques, 2^{nd} IEA Symp.*, (Lyngby, Denmark), pp. 99–113, June 9-10 2004.

[54] H. Madsen, "Forecasting of wind power.," in *Meeting at Danish Offshore Center, Meet. Offshore Cent.*, (Esbjerg, Denmark), March 2006.

[55] T. S. Nielsen and H. Madsen, "Experiences with statistical methods for wind power prediction," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (Nice, France), pp. 1066–1069, March 1-5 1999.

[56] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 5, no. 115-133, pp. 18–27, 1943.

[57] L. C. Jain and L. R. Medsker, *Recurrent Neural Networks: Design and Applications*. Boca Raton, FL, USA: CRC Press, Inc., 1st ed., 1999.

[58] B. J. Taylor, *Methods and Procedures for the Verification and Validation of Artificial Neural Networks*. Springer-Verlag New York Inc., 2005.

[59] S. Kalogirou, *Artificial intelligence in energy and renewable energy systems*. Nova Publishers, 2006.

[60] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2nd ed., 1998.

[61] H. G. Beyer, T. Degner, J. Hausmann, M. Hoffmann, and P. Rujan, "Short-Term Prediction of Wind Speed and Power Output of a Wind Turbine with Neural Networks," in *Proceedings of the* 2^{nd} *European Congresson Intelligent Techniques and Soft Computing*, (Aachen, Germany), September 20–23 1994.

[62] G. N. Kariniotakis, G. S. Stavrakakis, and E. F. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE Transactions on Energy Conversion*, vol. 11, no. 4, pp. 762–767, 1996.

[63] K. Sheela and S. Deepa, "A Review on Neural Network Models for Wind Speed Prediction," *Wind Engineering*, vol. 37, pp. 111–124, 04 2013.

[64] R. P. Lippmann, "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine (Acoustics, Speech, and Signal Processing)*, vol. 4, no. 2, pp. 4–22, 1987.

[65] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the* 5^{th} *Annual ACM Workshop on Computational Learning Theory*, (Pittsburgh, PA, USA), pp. 144–152, July 27-29 1992.

[66] V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, and T. Vetter, "Comparison of view-based object recognition algorithms using realistic 3D models," in *Proceedings of Artificial Neural Networks (ICANN'96)*, (Bochum, Germany), pp. 251–256, Springer, July 16-19 1996.

[67] H. Drucker, C. J. C. Burges, K. L., A. J. Smola, and V. Vapnik, "Support Vector Regression Machines," in *Proceedings of Advances in Neural Information Processing Systems 9 (NIPS'96)*, (Denver, CO, USA), pp. 155–161, MIT Press, December 2-5 1996.

[68] S. Haykin, *Neural Networks And Learning Machines*. Prentice Hall PTR, 2008.

[69] N. Cristianini and J. Shave-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[70] S. R. Gunn, "Support vector machines for classification and regression," tech. rep., Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, May 1998.

[71] Schölkopf, B. and Burges, C. J. C. and Smola, A. J., *Advances in Kernel Methods - Support Vector Learning*. Cambridge, MA: MIT Press, 1999.

[72] S. Gunn, M. Brown, and K. Bossley, "Network performance assessment for Neurofuzzy data modelling," in *Advances in Intelligent Data Analysis Reasoning about Data* (X. Liu, P. Cohen, and M. Berthold, eds.), vol. 1280 of *Lecture Notes in Computer Science*, pp. 313–323, Springer Berlin / Heidelberg, 1997.

[73] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, no. 6, pp. 939–947, 2004.

[74] G.-R. Ji, P. Han, and Y.-J. Zhai, "Wind speed forecasting based on support Vector Machine with forecasting error estimation," in *Proceedings of the* 6^{th} *International Conference on Machine Learning and Cybernetics*, (Hong Kong, China), pp. 2735–2739, August 19-22 2007.

[75] P. Zhao, J. Xia, Y. Dai, and J. He, "Wind speed prediction using support vector regression," in *Proceedings of the 2010* 5^{th} *IEEE Conference on Industrial Electronics and Applications, (ICIEA)*, (Taichung, Taiwan), pp. 882–886, June 15-17 2010.

[76] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[77] H. W. Sorenson, "Least-squares estimation: from Gauss to Kalman," *IEEE Spectrum*, vol. 7, no. 7, pp. 63–68, 1970.

[78] P. S. Maybeck, *Stochastic models, estimation, and control*, vol. 141 of *Mathematics in Science and Engineering*. Academic Press Inc, 1979.

[79] R. G. Brown and P. Y. C. Hwang, *Introduction to random signals and applied Kalman filtering*. New York: J. Wiley, 1992. Rev. ed. of: Introduction to random signal analysis and Kalman filtering. 1983.

[80] O. L. R. Jacobs, *Introduction to Control Theory,* 2^{nd} *Edition*. OUP Oxford, 1993.

[81] E. A. Bossanyi, "Short-term wind prediction using Kalman filters," *Wind Engineering*, vol. 9, no. 1, pp. 1–8, 1985.

[82] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Combining neural networks and ARIMA models for hourly temperature forecast," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'2000)*, vol. 4, (Como, Italy), pp. 414–419, July 24-27 2000.

[83] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[84] Y. He, Y. Zhu, and D. Duan, "Research on Hybrid ARIMA and Support Vector Machine Model in Short Term Load Forecasting," in *Proceedings of the* 6^{th} *International Conference on Intelligent Systems Design and Applications (ISDA '06)*, vol. 1, (Jinan, Shandong, China), pp. 804–809, October 16-18 2006.

[85] M. P. Horton, "Real-time identification of missile aerodynamics using a linearised Kalman filter aided by an artificial neural network," in *IEE Proceedings in Control Theory and Applications*, vol. 144, pp. 299–308, July 1997.

[86] S.-C. Huang and T.-K. Wu, "A Hybrid Unscented Kalman Filter and Support Vector Machine Model in Option Price Forecasting," in *Advances in Natural Computation* (L. Jiao, L. Wang, X. Gao, J. Liu, and F. Wu, eds.), vol. 4221 of *Lecture Notes in Computer Science*, pp. 303–312, Springer Berlin / Heidelberg, 2006.

[87] V. Mitra, C.-J. Wang, and S. Banerjee, "Lidar Detection of Underwater Objects Using a Neuro-SVM-Based Architecture," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 717–731, 2006.

[88] H. Liu, D. Xie, and W. Wu, "Soil water content forecasting by ANN and SVM hybrid architecture," *Environmental Monitoring and Assessment*, vol. 143, pp. 187–193, 2008.

[89] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model," *Renewable Energy*, vol. 35, no. 12, pp. 2732–2738, 2010.

[90] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," *Renewable Energy*, vol. 36, no. 4, pp. 1245–1251, 2011.

[91] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2348–2362, 2008.

[92] S. Salcedo-Sanz, E. G. Ortiz-García, Á. M. Pérez-Bellido, A. Portilla-Figueras, L. Prieto, D. Paredes, and F. Corroeso, "Performance Comparison of Multilayer Perceptrons and Support Vector Machines in a Short-term Wind Speed Prediction Problem," *Neural Network World*, vol. 19, no. 1, pp. 37–51, 2009.

[93] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, pp. 273–284, 2006.

[94] G. Kariniotakis, T. Ranchin, J. Moussafir, D. Buty, J. Usaola, I. Sanchez, I. Martí, H. Madsen, T. S. Nielsen, C. Lac, P. Frayssinet, H. P. Waldl, F. Dierich, J. A. Halliday, R. Brownsword, G. Giebel, L. Landberg, G. Kallos, P. Louka, J. Ottavi, U. Focken, M. Lange, J. Kintxo, J. Toefting, D. McCoy, E. Fagan, M. Collmann, E. Thalassinakis, A. Gigandidou, G. Gonzales, Morales, C. Barquero, P. De Arriba, I. Cruz, N. Hatziargyriou, and G. Sideratos, "Towards Next Generation Short-Term Forecasting of Wind Power - The ANEMOS project," in *Proceedings of Global Windpower Conference* & *Exhibition*, (Chicago, USA), March 28–31 2004.

[95] K. Lynch, D. J. Brayshaw, A. Charlton-Perez, C. Durman, and J. Richards, "Monthly/Seasonal Forecasting of Wind Energy," in *Royal Meteorological Society Meeting (RMSM): Renewables and the future of energy meteorology - Poster session*, 2012.

[96] E. García-Bustamante, J. F. González-Rouco, P. A. Jiménez, J. Navarro, and J. P. Montávez, "A Comparison of Methodologies for Monthly Wind Energy Estimation," *Wind Energy*, vol. 12, no. 7, pp. 640–659, 2009.

[97] E. Cadenas and W. Rivera, "Wind speed forecasting in the South Coast of Oaxaca, Mexico," *Renewable Energy*, vol. 32, no. 12, pp. 2116–2128, 2007.

[98] S. Kennedy and P. Rogers, "A Probabilistic Model for Simulating Long-Term Wind-Power Output," *Wind Engineering*, vol. 27, no. 3, pp. 167–181, 2003.

[99] P. Kritharas and S. J. Watson, "A comparison of long-term wind speed forecasting models," *Journal of Solar Energy Engineering*, vol. 132, no. 4, p. 041008, 2010.

[100] P. Kritharas, C. M. Murphy, and S. J. Watson, "Improved long term wind speed prognosis using advanced statistical methods," in *European Wind Energy Conference and Exhibition (EWEC) - Poster session*, (Warsaw, Poland), April 20–23 2010.

[101] A. More and M. C. Deo, "Forecasting wind with neural networks," *Marine Structures*, vol. 16, no. 1, pp. 35–49, 2003.

[102] Gao, S. and Dong, L. and Tian, C. and Liao, X., "Wind power prediction using time-series analysis base on rough sets," in *Electric Information and Control Engineering (ICEICE), 2011 International Conference on*, pp. 2847–2852, 2011.

[103] D. A. Fadare, "The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria," *Applied Energy*, vol. 87, no. 3, pp. 934–942, 2010.

[104] S. Kalogirou, C. Neocleous, S. Paschiardis, and C. Schizas, "Wind Speed Prediction Using Artificial Neural Networks," in *Proceedings of the European Symposium on Intelligent Techniques (ESIT'99)*, (Crete, Greece), June 3–4 1999.

[105] P. R. J. Campbell and K. Adamson, "A Novel Approach to Wind Forecasting in the United Kingdom and Ireland," *International Journal of Simulation: Systems, Science and Technology*, vol. 6, no. 12–13, pp. 1–10, 2005.

[106] K. Huang, L. Dai, and S. Huang, "Wind Prediction Based on Improved BP Artificial Neural Network in Wind Farm," in *Proceedings of the International Conference on Electrical and Control Engineering (ICECE)*, (Wuhan, China), pp. 2548–2551, June 25–27 2010.

[107] M. Bilgili, B. Sahin, and A. Yasar, "Application of artificial neural networks for the wind speed prediction of target station using reference stations data," *Renewable Energy*, vol. 32, no. 14, pp. 2350 – 2360, 2007.

[108] P. López, R. Velo, and F. Maseda, "Effect of direction on wind speed estimation in complex terrain using neural networks," *Renewable Energy*, vol. 33, no. 10, pp. 2266–2272, 2008.

[109] M. Bilgili and B. Sahin, "Wind Speed Prediction of Target Station from Reference Stations Data," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 35, no. 5, pp. 455–466, 2013.

[110] M. A. Mohandes, S. Rehman, and T. O. Halawani, "A neural networks approach for wind speed prediction," *Renewable Energy*, vol. 13, no. 3, pp. 345–354, 1998.

[111] M. Bilgili and B. Sahin, "Comparative analysis of regression and artificial neural network models for wind speed prediction," *Meteorology and Atmospheric Physics*, vol. 109, pp. 61–72, 2010.

[112] Z. Guo, W. Zhao, H. Lu, and J. Wang, "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model," *Renewable Energy*, vol. 37, pp. 241–249, January 2012.

[113] T. Chengwei, D. Lei, G. Shuang, and L. Xiaozhong, "A study of grey theory used in prediction of annual wind power generation," in *Proceedings of the International Conference on Electrical and Control Engineering (ICECE)*, (Wuhan, China), pp. 1952–1955, April 15–17 2011.

[114] Y. Gao, L. Liu, G. Liang, and S. Ma, "Study on the medium-term and long-term forecast technology of wind farm power generation," in *Communication Systems and Information Technology*, vol. 100, pp. 601–607, Springer Berlin / Heidelberg, 2011.

[115] J. Shi, J. Guo, and S. Zheng, "Evaluation of hybrid forecasting approaches for wind speed and power generation time series," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 3471–3480, June 2012.

[116] W. Zhang, J. Wu, J. Wang, J. Zhao, and L. Shen, "Performance analysis of four modified approaches for wind speed forecasting," *Applied Energy*, vol. 99, pp. 324–333, November 2012.

[117] Z. Guo, J. Zhao, W. Zhang, and J. Wang, "A corrected hybrid approach for wind speed prediction in Hexi Corridor of China," *Energy*, vol. 36, pp. 1668–1679, March 2011.

[118] S. J. Watson and P. Kritharas, "Long Term Wind Speed Variability in the UK," in *European Wind Energy Association Conference (EWEA)*, (Copenhagen, Denmark), April 16–19 2012.

[119] S. J. Watson, P. Kritharas, and G. Hodgson, "Wind Speed Variability across the UK between 1957 and 2009." In Review, March 11 2012.

[120] UK Met Office, "Midas Land Surface Stations Data (1853-Current)." http://badc.nerc.ac.uk/data/ukmo-midas. Last checked: December 2010.

[121] British Atmospheric Data Centre, "Met Office Surface Data Users Guide." http://badc.nerc.ac.uk/data/surface/ukmo_guide.html, April 2007. Last checked: December 2012.

[122] S. M. Uppala, P. W. Kållberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. Mcnally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch,

D. Vasiljevic, P. Viterbo, and J. Woollen, "The ERA-40 re-analysis," *International Journal of Climatology*, vol. 131, no. 612, pp. 2961–3012, 2005.

[123] British Atmospheric Data Centre, "Met Office Synoptic network." http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__obs_11694808474912639, December 2012. Last checked: January 2013.

[124] Duplicate values on MIDAS dataset ISSUE=165 PROJ=19, June 22 2012. Private communication with MIDAS team.

[125] A. Lapworth and J. McGregor, "Seasonal variation of the prevailing wind direction in Britain," *Weather*, vol. 63, no. 12, pp. 365–368, 2008.

[126] A. G. Dawson, K. Hickey, T. Holt, L. Elliott, D. Dawson, I. D. L. Foster, P. Wadhams, I. Jonsdottir, J. Wilkinson, J. McKenna, N. R. Davis, and D. E. Smith, "Complex North Atlantic Oscillation (NAO) Index signal of historic North Atlantic storm-track changes," *The Holocene*, vol. 12, no. 3, pp. 363–369, 2002.

[127] Garrad Hassan, "UK Wind Speed Index." http://web.archive.org/web/20080223103132/www.garradhassan.com/services/ghindex/. Last checked: 9^{th} June, 2013.

[128] Bossard, M. and Feranec, J. and Otahel, J., "CORINE Land Cover technical guide - Addendum 2000," tech. rep., European Environment Agency, 2000.

[129] European Environment Agency, "CORINE Land cover - Part 2: Nomenclature." http://www.eea.europa.eu/publications/COR0-part2/page001.html. Last checked: June 2013.

[130] J. Silva, C. Ribeiro, and R. Guedes, "Roughness length classification of CORINE land cover classes," in *European Wind Energy Conference and Exhibition (EWEC)*, (Milan, Italy), May 7–10 2007.

[131] Google Earth 7.03.8542, "BADC-57 stations, elevation 100m." http://www.google.com/earth/index.html. Last checked: June 2013.

[132] T. J. Teorey, S. S. Lightstone, and T. Nadeau, *Database Modeling and Design: Logical Design*. Diane Cerra, 4^{th} ed., 2005.

[133] ORACLE, "MySQL – The world's most popular open source database." http://www.mysql.com/. Last checked: June 2013.

[134] ORACLE, "MySQL Connector: ODBC." http://dev.mysql.com/doc/refman/5.0/en/connector-odbc.html. Last checked: June 2013.

[135] ORACLE, "The MySQL Command Line Tool." http://dev.mysql.com/doc/refman/5.0/en/mysql.html. Last checked: June 2013.

[136] ORACLE, "MySQL Administrator." http://downloads.mysql.com/docs/administrator-en.a4.pdf. Last checked: June 2013.

[137] ORACLE, "MySQL Query Browser." http://downloads.mysql.com/docs/query-browser-en.a4.pdf. Last checked: June 2013.

[138] A. Becker, "HeidiSQL." http://www.heidisql.com. Last checked: June 2013.

[139] ORACLE, "MySQL Reference Manual." http://dev.mysql.com/doc/refman/5.0/en/index.html. Last checked: June 2013.

[140] International Energy Agent (IEA), "Wind Energy Projects in Cold Climates." http://arcticwind.vtt.fi, 2012. June 2014.

[141] E. Mason and A. M. P. Santos and Á. J. Peliz, "Western Iberian winter wind indices based on significant wind events," *Ocean Science Discussions*, vol. 2, no. 2, pp. 105–127, 2005.

[142] K. Harman and C. Morgan, "Use of regional wind energy indices to predict long-term wind farm production and to assess portfolio effect," in *Proceedings of the World Renewable Energy Congress (WREC)*, (Aberdeen, UK), May 22-27 2005.

[143] G. J. McCabe, M. P. Clark, and M. C. Serreze, "Trends in Northern Hemisphere Surface Cyclone Frequency and Intensity," *Journal of Climate*, vol. 14, no. 12, pp. 2763–2768, 2001.

[144] C. J. Paciorek, J. S. Risbey, V. Ventura, and R. D. Rosen, "Multiple Indices of Northern Hemisphere Cyclone Activity, Winters 1949–99," *Journal of Climate*, vol. 15, no. 13, pp. 1573–1590, 2002.

[145] X. Wang, F. Zwiers, V. R. Swail, and Y. Feng, "Trends and variability of storminess in the Northeast Atlantic region, 1874–2007," *Climate Dynamics*, vol. 33, no. 7–8, pp. 1179–1195, 2009.

[146] L. Bärring and K. Fortuniak, "Multi-indices analysis of southern Scandinavian storminess 1780–2005 and links to interdecadal variations in the NW Europe–North Sea region," *International Journal of Climatology*, vol. 29, no. 3, pp. 373–384, 2009.

[147] S. C. Pryor and R. J. Barthelmie, "Long-term trends in near-surface flow over the Baltic," *International Journal of Climatology*, vol. 23, no. 3, pp. 271–289, 2003.

[148] J. Pinto, U. Ulbrich, G. Leckebusch, T. Spangehl, M. Reyers, and S. Zacharias, "Changes in storm track and cyclone activity in three SRES ensemble experiments with the ECHAM5/MPI-OM1 GCM," *Climate Dynamics*, vol. 29, no. 2–3, pp. 195–201, 2007.

[149] S. C. Pryor, R. J. Barthelmie, and J. T. Schoof, "Inter-annual variability of wind indices across Europe," *Wind Energy*, vol. 9, no. 1-2, pp. 27–38, 2006.

[150] L. C. Cradden, G. P. Harrison, and J. P. Chick, "Will climate change impact on wind power development in the UK?," *Climatic Change*, vol. 115, no. 3–4, pp. 837–852, 2012.

[151] G. C. Leckebusch and U. Ulbrich, "On the relationship between cyclones and extreme windstorm events over Europe under climate change," *Global and Planetary Change*, vol. 44, no. 1–4, pp. 181 – 193, 2004.

[152] P. Nolan, P. Lynch, R. McGrath, T. Semmler, and S. Wang, "Simulating climate change and its effects on the wind energy resource of ireland," *Wind Energy*, vol. 15, no. 4, pp. 593–608, 2012.

[153] M. G. Donat, G. C. Leckebusch, J. Pinto, and U. Ulbrich, "European storminess and associated circulation weather types: future changes deduced from a multi-model ensemble of GCM simulations," *Climate Research*, vol. 42, no. 1, pp. 27–43, 2010.

[154] M. Beniston, D. B. Stephenson, O. B. Christensen, C. A. T. Ferro, C. Frei, S. Goyette, K. Halsnaes, T. Holt, K. Jylhä, B. Koffi, J. P. Paluticof, R. Schöll, T. Semmler, and K. Woth, "Future extreme events in European climate: an exploration of regional climate model projections," *Climatic Change*, vol. 81, no. 1, pp. 267–280, 2007.

[155] B. Rockel and K. Woth, "Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations," *Climatic Change*, vol. 81, no. 1, pp. 267–280, 2007.

[156] S. Brown, P. Boorman, R. McDonald, and M. J., "Interpretation for use of surface wind speed projections from the 11- member Met Office regional climate model ensemble," tech. rep., Met. Office, Hadley Centre, Exeter, 2009.

[157] D. M. H. Sexton and M. J., "UKCP09: probabilistic projections of wind speed," tech. rep., Met. Office, Hadley Centre, Exeter, 2010.

[158] G. P. Harrison, L. C. Cradden, and J. P. Chick, "Preliminary assessment of climate change impacts on the UK onshore wind energy resource," *New Energy Sources*, vol. 30, pp. 1286–1299, 2009.

[159] W. G. Früh, "Long-term wind resource and uncertainty estimation using wind records from Scotland as example," *Renewable Energy*, vol. 50, pp. 1014 – 1026, February 2013.

[160] Nielsen, P., "New Danish Wind energy index calculation." http://www.vindstat.dk/, 2010. Last checked: December 2012.

[161] A. M. R. Bakker and B. J. J. M. van den Hurk, "Estimation of persistence and trends in geostrophic wind speed for the assessment of wind energy yields in Northwest Europe," *Climate Dynamics*, vol. 39, no. 3–4, pp. 767–782, 2012.

[162] A. M. R. Bakker, B. J. J. M. Van den Hurk, and J. P. Coelingh, "Decomposition of the windiness index in the Netherlands for the assessment of future long-term wind supply," *Wind Energy*, 2012.

[163] D. J. Rasmussen, T. Holloway, and G. F. Nemet, "Opportunities and challenges in assessing climate change impacts on wind energy—a critical comparison of wind speed projections in California," *Environmental Research Letters*, vol. 6, no. 2, p. 024008, 2011.

[164] S. C. Pryor, R. J. Barthelmie, D. T. Young, E. S. Takle, R. W. Arritt, D. Flory, W. J. Gutowski, A. Nunes, and J. Roads, "Wind speed trends over the contiguous United States," *Journal of Geophysical Research: Atmospheres*, vol. 114, no. D14, pp. n/a–n/a, 2009.

[165] S. C. Pryor and R. J. Barthelmie, "Climate change impacts on wind energy: A review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1, pp. 430–437, 2010.

[166] Höglund, A. and Meier, H. E. M. and Broman, H. and Kriez, E., "Validation and Correction of Regionalised ERA-40 Wind Fields Over the Baltic Sea Using the Rossby Centre Atmosphere Model RCA3.0," Tech. Rep. 97, Oceanografi Swedish Meteorological and Hydrological Institute, 2009.

[167] N. Wever, "Quantifying trends in surface roughness and the effect on surface wind speed observations," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D11, 2012.

[168] M. Perry and D. Hollis, "The development of a new set of long-term climate averages for the uk," *International Journal of Climatology*, vol. 25, no. 8, pp. 1023 – 1039, 2005.

[169] M. Perry and D. Hollis, "The generation of monthly gridded datasets for a range of climatic variables over the uk," *International Journal of Climatology*, vol. 25, no. 8, pp. 1041 – 1054, 2005.

[170] P. Thomas, S. Cox, and A. Tindal, "Long-Term Wind Speed Trends in NorthWestern Europe," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (Marseille, France), March 16-19 2009.

[171] G. Ebsworth, C. Ribeiro, and B. Hodgetts, "A Study of Recent Wind Speed Trends in the United Kingdom," Tech. Rep. 1026/BT/08, GL–Garrad Hassan, 2011.

[172] P. Kritharas and S. J. Watson, "Long term wind speed forecasting based on seasonal trends," in *Proceedings of the ASME 3^{rd} Int. Conf. on Energy Sustainability*, vol. 2, (San Francisco, CA, USA), pp. 897–904, July 19-23 2009.

[173] I. Vanderhoven, "Power Spectrum of Horizontal Wind Speed in the Frequency Range from 0.0007 to 900 Cycles Per Hour," *Journal of Meteorology*, vol. 14, no. 2, pp. 160–164, 1957.

[174] D. A. Dickey and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, vol. 74, pp. 427–431, June 1979.

[175] MathWorks®, "Econometrics Toolbox™ User's Guide." http://www.mathworks.com/help/pdf_doc/econ/econ.pdf, 2011. Last checked: December 2012.

[176] G. Plaut and E. Simonnet, "Large-scale circulation classification, weather regimes, and local climate over France, the Alps and western Europe," *Climate Research*, vol. 17, no. 3 special 9, pp. 303–324, 2001.

[177] C. Cassou, L. Terray, J. W. Hurrell, and C. Deser, "North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing," *Journal of Climate*, vol. 17, no. 5, pp. 1055–1068, 2004.

[178] C. Fil and L. Dubus, "Winter climate regimes over the North Atlantic and European region in ERA40 reanalysis and DEMETER seasonal hindcasts," *Tellus A*, vol. 57, no. 3, pp. 290–307, 2005.

[179] J. P. Paluticof, X. Guo, and J. A. Halliday, "Climate variability and the UK wind resource," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 39, no. 1–3, pp. 243–249, 1992.

[180] R. T. Sutton and M. R. Allen, "Decadal predictability of North Atlantic sea surface temperature and climate," *Nature*, vol. 388, pp. 563–567, 1997.

[181] M. N. Lorenzo, I. Iglesias, J. J. Taboada, and M. Gómez-Gesteira, “Relationship between monthly rainfall in northwest Iberian Peninsula and North Atlantic sea surface temperature,” *International Journal of Climatology*, vol. 30, no. 7, pp. 980–990, 2010.

[182] R. S. Lindzen and S. Nigam, “On the Role of Sea Surface Temperature Gradients in Forcing Low-Level Winds and Convergence in the Tropics,” *Journal of the Atmospheric Sciences*, vol. 44, no. 17, pp. 2418–2436, 1987.

[183] W. Drosdowsky and L. E. Chambers, “Near-Global Sea Surface Temperature Anomalies as Predictors of Australian Seasonal Rainfall,” *Journal of Climate*, vol. 14, no. 7, pp. 1677–1687, 2001.

[184] S. E. Zebiak, “A Simple Atmospheric Model of Relevance to El Niño,” *Journal of the Atmospheric Sciences*, vol. 39, no. 9, pp. 2017–2027, 1982.

[185] J. C. H. Chiang, S. E. Zebiak, and M. A. Cane, “Relative Roles of Elevated Heating and Surface Temperature Gradients in Driving Anomalous Surface Winds over Tropical Oceans,” *Journal of the Atmospheric Sciences*, vol. 58, no. 11, pp. 1371–1394, 2001.

[186] B. Qu, A. J. Gabric, J. Zhu, D. Lin, F. Quian, and M. Zhao, “Correlation between sea surface temperature and wind speed in Greenland Sea and their relationships with NAO variability,” *Water Science and Engineering*, vol. 5, no. 3, pp. 304–315, 2012.

[187] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: methods and applications*. John Wiley & Sons, 1998.

[188] L. E. Back and C. S. Bretherton, “On the relationship between sst gradients, boundary layer winds, and convergence over the tropical oceans,” *Journal of Climate*, vol. 22, no. 15, pp. 4182–4196, 2009.

[189] S. E. Zebiak, “Atmospheric Convergence Feedback in a Simple Model for El Niño,” *Monthly Weather Review*, vol. 114, no. 7, pp. 1263–1271, 1986.

[190] J. D. Neelin, “On the Intepretation of the Gill Model,” *Journal of the Atmospheric Sciences*, vol. 46, no. 15, pp. 2466–2468, 1989.

[191] B. Wang and T. Li, “A Simple Tropical Atmosphere Model of Relevance to Short-Term Climate Variations,” *Journal of the Atmospheric Sciences*, vol. 50, no. 2, pp. 260–284, 1993.

[192] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Second International Symposium on Information Theory* (B. N. Petrov and F. Csaki, eds.), (Budapest, Hungary), pp. 267–281, Akademiai Kiado, 1973.

[193] MathWorks®, “Systems Identification Toolbox™ User’s Guide.” http://www.mathworks.com/help/pdf_doc/system-identification-toolbox.pdf, 2011. Last checked: December 2012.

[194] I. Youm, J. Sarr, M. Sall, A. Ndiaye, and M. Kane, “Analysis of wind data and wind energy potential along the northern coast of Senegal,” *Rev. Ene. Ren.*, vol. 8, pp. 95–108, 2005.

[195] 4C Offshore, "Project Details for Arklow Bank Phase 1." http://www.4coffshore.com/windfarms/arklow-bank-phase-1-ireland-ie01.html. Last checked: March 2014.

[196] 4C Offshore, "Project Details for Dogger Bank Tranche D." http://www.4coffshore.com/windfarms/dogger-bank-united-kingdom-uk43.html. Last checked: March 2014.

[197] G. Kariniotakis, I. Martí, D. Casas, P. Pinson, T. S. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A. M. Palomares, J. Tambke, U. Focken, M. Lange, P. Louka, G. Kallos, C. Lac, G. Sideratos, and N. Hatziargyriou, "What performance an can be expected by short-term wind power prediction models depending on site characteristics?," in *Proceedings of the European Wind Energy Conference and Exhibition (EWEC)*, (London, UK), November 22–25 2004.

[198] W. A. Fuller, *Introduction to Statistical Time Series (*2^{nd} *ed.)*. John Wiley Sons, Inc., 1976 (1^{st} ed.), 1996 (2^{nd} ed.).