promoting access to White Rose research papers

White Rose Research Online

Universities of Leeds, Sheffield and York
http://eprints.whiterose.ac.uk/

This is an author produced version of a paper published in **Renewable Energy**

White Rose Research Online URL for this paper:

http://eprints.whiterose.ac.uk/id/eprint/78143

Paper:
Weekes, SM and Tomlin, AS (2014) *Data efficient measure-correlate-predict approaches to wind resource assessment for small-scale wind energy.* Renewable Energy, 63. 162 - 171. ISSN 0960-1481.

http://dx.doi.org/10.1016/j.renene.2013.08.033

White Rose Research Online
eprints@whiterose.ac.uk

Data Efficient Measure-Correlate-Predict Approaches to Wind Resource Assessment for Small-Scale Wind Energy

S. M. Weekes and A. S. Tomlin*

Doctoral Training Centre in Low Carbon Technologies, Energy Research Institute, School of Process, Environmental and Materials Engineering, University of Leeds, Leeds, LS2 9JT, UK

*a.s.tomlin@leeds.ac.uk

Highlights:

- Long-term wind energy resource predicted using just three months onsite wind speed measurements.
- Robust error statistics obtained using 120 test periods at 22 sites over 11 years.
- Link between measurement season and prediction accuracy investigated.
- Data-driven prediction approaches compared to semi-empirical boundary layer model.

Abstract

The feasibility of predicting the long-term wind resource at 22 UK sites using a measure-correlate-predict (MCP) approach based on just three months onsite wind speed measurements has been investigated. Three regression based techniques were compared in terms of their ability to predict the wind resource at a target site based on measurements at a nearby reference site. The accuracy of the predicted parameters of mean wind speed, mean wind power density, standard deviation of wind speeds and the Weibull shape factor was assessed, and their associated error distributions were investigated, using long-term measurements recorded over a period of 10 years. For each site, 120 wind resource predictions covering the entire data period were obtained using a sliding window approach to account for inter-annual and seasonal variations. Both the magnitude and sign of the prediction errors were found to be strongly dependent on the season used for onsite measurements. Averaged across 22 sites and all seasons, the best performing MCP approach resulted in mean absolute and percentage errors in the mean wind speed of 0.21 ms^{-1} and 4.8% respectively, and in the mean wind power density of 11 wm^{-2} and 14%. The average errors were reduced to 3.6% in the mean wind speed and 10% in the mean wind power density when using the optimum season for onsite wind measurements. These values were shown to be a large improvement on the predictions obtained using an established semi-empirical model based on boundary layer scaling. The results indicate that the MCP approaches applied to very short onsite measurement periods have the potential to be a valuable addition to the wind resource assessment toolkit for small-scale wind developers.

1. Introduction

Small-scale wind energy (typically defined as < 50kW [1]) is achieving increasing interest as individuals, organisations and governments seek to decarbonize electricity supply [2]. The UK, which has put in place a legally binding commitment to reduce carbon equivalent emissions in 2050 by 80% compared to 1990 levels [3], is particularly well placed to capitalize on small-scale wind energy as a decentralized, well-established renewable energy technology due to its favorable wind resource and growing expertise in the industry. These factors, along with the introduction of a feed-in-tariff in 2010 which pays a fixed tariff for every kilowatt hour of electricity generated from small-scale installations, have resulted in significant growth in the UK small-scale wind energy industry with installed capacity projected to reach 1.3 GW by 2020 [4].

However, in order for small-scale wind energy to reach its full potential, tools capable of predicting the wind energy resource quickly, cheaply and accurately are urgently required [5]. Such tools will allow consumers to make informed decisions regarding the financial viability of a proposed installation as well as the carbon savings that may be achieved.

Thanks to decades of development, wind resource assessment is well established in the large-scale wind industry. To account for the spatial and temporal variability inherent in wind flows, wind speed and direction are monitored at a number of locations across a potential site to establish the statistical characteristics of the wind resource. These data are used in conjunction with wind turbine power curves to predict the long-term wind energy resource at the proposed site. Data collected over 1-3 years [6] along with longer-term correlations to reference sites are typically required to obtain statistics which are robust enough to justify the large financial investments involved in such projects. However, in the small-scale wind industry, where investment costs are considerably lower, onsite measurement campaigns on these timescales are often impractical and prohibitively expensive, necessitating the use of indirect methods.

Indirect approaches may be broadly categorized as modelling or data-driven. Any indirect approach must be capable of estimating the long-term average hub-height wind speed and the form of the wind speed distribution as a minimum requirement to predicting the wind energy resource. Building on previous work related to boundary layer meteorology and descriptors of surface roughness, the UK Met Office developed a promising modelling methodology which may be used to estimate the mean wind speed at any UK site without the need for onsite measurements [7]. However, despite the utility of such techniques in identifying potential wind energy locations, the uncertainties present in simple boundary layer scaling approaches may not be tolerable where significant investments are to be made [8]. The UK Met Office are currently developing more sophisticated modelling approaches based on long-term historical weather forecasts which may result in reduced uncertainties [9]. However, since uncertainties increase with site complexity [10], onsite wind measurements may still be required to achieve predictions with a sufficient level of confidence.

In contrast to modelling techniques, data-driven approaches to small-scale wind resource assessment have received relatively little attention. This work is concerned with one such data-driven approach known as measure-correlate-predict (MCP). MCP increases the value of short-term wind speed measurements recorded at a target site by correlating these with concurrent data recorded at a reference site. The correlation is then used to predict the long-term wind resource at the target site using long-term historical data from the reference site [11]. Since long-term wind speed records are routinely held by airports and national weather forecasters, this technique provides a means of reducing the onsite measurement time required at the target site. MCP is already utilized in the large-scale wind industry where the relationship between the reference and target sites is typically estimated from a concurrent measurement period covering a year or more [12]. A number of studies have considered the application of different regression techniques to the MCP approach [13-15] but very little work has been done regarding the application of MCP to measurement periods of much less than one year [16]. While long-term onsite measurements are clearly desirable in reducing uncertainty, the focus of this work is to establish the feasibility of applying MCP approaches to measurement periods of much less than one year. The aim of this is to determine whether data-driven resource assessment can be made more accessible to the small-scale wind industry. In practice, the short-term onsite measurements required for MCP could be obtained using a portable meteorological mast or LiDAR (light detection and ranging) equipment.

In this study, we constrain the onsite measurement period to just three months which is a more practically and economically viable time period in the case of small-scale wind installations. We compare the performance of three MCP techniques, simple linear regression (LR), linear regression augmented by a Gaussian scatter term (LR2) and variance ratio regression (VR). The success of the techniques in predicting the long-term wind resource at 22 UK sites located in a variety of terrains is assessed through comparison with onsite measurements recorded over a period of 10 years. The three month training period is applied throughout the whole data record using a sliding window technique to account for inter-annual and season variability. Metrics are applied to quantitatively assess the average errors in the predicted mean wind speed, mean wind power density, standard deviation of wind speeds and Weibull shape factor.

2. Methodology

2.1. Measure-Correlate-Predict

The MCP strategy involves three stages: (I) measurement of wind speeds at a proposed installation site (the target site), (II) identifying a correlation between the target site and concurrent measurements at a local long-term reference site such as an airport or meteorological station and (III) predicting the long-term wind resource at the target site using long-term historical data from the reference site.

Typically, a simple correlation is sought whereby the long-term historical reference data may be used to construct a time-series of wind speeds (and possibly wind directions) at the target site. From this time-series, statistical descriptors may be extracted which represent the long-term target site wind resource. In the case of the large-scale wind industry, onsite measurements at the target site covering a year or more may be extrapolated to several decades to predict the likely energy output over the lifetime of the installation. The main purpose of such a process is to take account of inter-annual variation in the wind resource. In the scheme proposed in this study, the analysis is based on just three months onsite measurements and thus a larger burden is placed on the MCP process since it must account for both inter-annual and inter-seasonal variations and it must achieve this using a much reduced data set.

The majority of MCP techniques involve the use of a parametric linear relationship between target and reference site wind speeds [11], although a number of alternatives have also been proposed [15, 17, 18] with variable results. In this work we have used two established regression based MCP techniques, LR and VR, and a third technique LR2, which employs a Gaussian scatter term in an attempt to improve the predictions of the wind speed distribution compared to simple linear regression. The 10 year wind resource at the target site was estimated using just three months concurrent measurements at the reference and target sites (the training data), and a further 10 years historical measurements at the reference site. In order to account for both seasonal and inter-annual variability, as well as to produce robust error statistics, multiple three month training periods were selected from the 11 year data record using a sliding window technique. The approach is shown schematically in Figure 1 and can be summarized as follows:

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

(II) At each step, a three month training period at the start of the training window is used to extract the regression parameters for the MCP approaches of LR, LR2 and VR. This represents the short-term onsite measurement period proposed in this study.

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

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

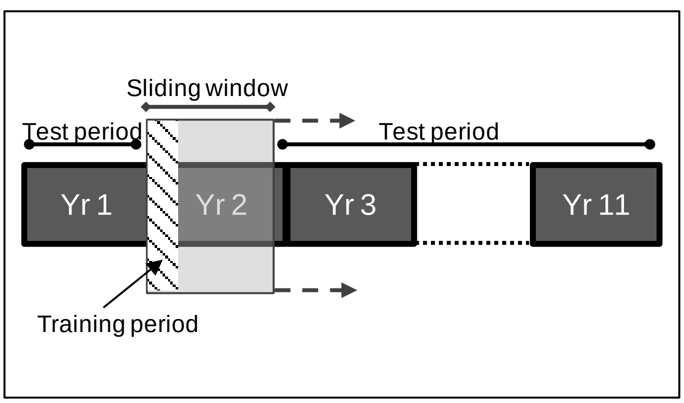

Figure 1: Schematic diagram of the sliding window technique used to test the MCP predictions across the entire data record. The test periods move with the training window such that the two never overlap.

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

Each of the MCP techniques was applied sector-wise to data which was first binned according to the reference site wind direction. The sector approach was used to account for the directional dependence of the upwind roughness which previous work [8] has shown affects the scaling between the reference and target site wind speeds. In line with previous studies [14, 19], angular bins of width 30º were used resulting in 12 separate regressions for each target-reference site pair. For angular bins with less than 20 data entries, the regression parameters were obtained by applying a global fit to data from all bins. The individual regression techniques will now be described in more detail.

Linear regression (LR)
The individual wind speeds at the target site as described by LR are given by the equation:

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

Equation 1

where u_{tar} is the observed wind speed at the target site, u_{ref} is the corresponding observed wind speed at the reference site, α and β are the regression coefficients obtained using a least squares fit to the training data and ε is an error term which represents the residual scatter. Assuming the data is well represented by a linear fit, $\hat{u}_{tar} = \alpha + \beta u_{ref}$ represents the mean prediction at the target site given a reference wind speed u_{ref}. Despite its simplicity, standard linear regression has been shown to be surprisingly effective in predicting the target site mean wind speed [11, 18] and is included in many commercial MCP software packages.

Linear regression with Gaussian scatter (LR2)
While the LR approach is frequently used as a baseline for assessing the performance of more sophisticated MCP techniques, it is noteworthy that that the residual scatter term ε in Equation 1 is often not considered explicitly. This term represents a variety of processes which are not accounted for in the simple linear model and which result in scatter in the individual data points about the mean prediction $\hat{u}_{tar}$. Assuming the residuals are normally distributed about $\hat{u}_{tar}$, this scatter will not have a significant effect on the predicted mean wind speed. However, this term will have an impact on the shape of the predicted wind speed distribution, and due to the cubic relationship between wind speed and wind power, this will in turn impact on the predicted wind power.

In order to account for this scatter, the error term was explicitly modelled as a zero-mean Gaussian distribution centred on each of the mean target site wind speed predictions $\hat{u}_{tar}$ [20]:

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

Equation 2

where σ_{res} is the sample standard deviation of the residuals about the predicted target site wind speeds $\hat{u}_{tar}$, as calculated from the training data using [20]:

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

Equation 3

where i represents the i^{th} data point, N is the number of observations and σ_{res} represents the average, wind speed independent value for a particular 30° angular sector. The residual scatter ε is then reconstructed by adding random draws from the Gaussian distribution in Equation 2 to the individual mean predictions at the target site, $\hat{u}_{tar}$. While such a model is in line with the standard assumptions of linear regression, it represents a simplification since the underlying joint distribution of u_{tar} and u_{ref} is expected to be joint Weibull rather than joint Gaussian [15].

Variance Ratio Regression (VR)
In the simple linear regression case (LR) where no account is taken of the residual scatter ε, it is known that the standard deviation of the predicted wind speeds $\sigma(\hat{u})$ about the whole sample mean will be smaller than the standard deviation of the observed values $\sigma(u)$ about the whole sample mean by a factor r, where r is the linear correlation coefficient [11]. Here $\hat{u}$ and u are used to indicate that the standard deviations are a function of the predicted or observed wind speeds respectively. Note that the standard deviation referred to here, describes the spread of wind speeds about the whole sample mean, in contrast with σ_{res} defined by Equation 3 which describes residual scatter about individual wind speed predictions. The underestimation of the standard deviation will impact on the shape of the predicted wind speed distribution and hence will result in an under prediction of the available wind power at the target site. To account for this, Rogers *et al*. [11] proposed a variation of simple LR based on setting $\sigma(\hat{u}) = \sigma(u)$. Using least squares theory it can be shown that this condition results in a linear equation of the form:

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

Equation 4

where σ_{tar} and σ_{ref} represent the standard deviation about the mean wind speeds at the target ($\bar{u}_{tar}$) and reference ($\bar{u}_{ref}$) sites respectively, as calculated from the short-term training data. This approach differs from that described in Equation 1 and Equation 2 in that no direct attempt is made to model the error term in order to reconstruct the residual scatter. Instead, Equation 4 postulates zero scatter (r = 1) resulting in a gradient which is a factor of $1/r$ larger than that obtained using simple LR thus forcing the required increase in $\sigma(\hat{u})$.

2.2. Meteorological measurements

In order to assess the success of the MCP approaches, long-term meteorological data (hourly averaged wind speed and direction) were obtained from the Met Office anemometer network [21] for a number of UK sites. The wind speed and wind angle were recorded with a resolution of one knot (0.51 ms^{-1}) and ten degrees respectively. Each site was designated as either a *target site* (a location where we wish to predict the wind resource), or a *reference site* (used as a predictor for the target site wind resource). The target sites were specifically chosen to be representative of different terrain types, *urban*, *sub-urban*, *rural* and *coastal* to reflect the range of scenarios which may be encountered by small-scale wind energy developers. The site classification was achieved through examination of satellite images. Wherever possible, reference sites were located either in exposed rural areas, or in coastal locations in the case of pairings with coastal sites. Anemometers were assumed to be located at a height of 10 m above ground level in line with the standard Met Office observational practice. Where anemometers were known to differ from this height, they are noted below.

The data included 22 target sites and 15 reference sites, (depending on geographical location, a single reference site may serve multiple target sites), and covered the period August 2001 to July 2012. The training and test periods were varied throughout the data set using the sliding window approach described above, such that the two periods did not overlap. In a real-world prediction scenario, we would expect to obtain short-term measurements at the target site and to have access to long-term historical data only from the reference site. In this study however, we require long-term data at both the reference and target sites to allow the predictions made using MCP to be tested against long-term measurements. Details of all sites used in this work can be found in Table 1 and approximate geographical locations are shown in Figure 2.

Target Sites				**Reference Sites**					
Site	OS grid	Elev (m)	$\bar{u}_{obs}$ (ms^{-1})	Site	OS grid	Elev (m)	$\bar{u}_{obs}$ (ms^{-1})	d (km)	r
U1*	SJ8396	33	3.2	**Rf1**	SD6614	440	6.6	25	0.78
U2**	SU4210	26	4.4	**Rf2**	SU5501	9	6.0	16	0.85
SU1	NJ8712	65	4.5	**Rf3**	NO4620	10	4.8	101	0.57
SU2	SK5045	117	3.5	**Rf4**	TF0049	63	5.1	49	0.88
SU3	SU8554	65	3.6	**Rf5**	SU3039	90	4.1	58	0.87
SU4	SU1344	132	3.7	**Rf5**	SU3039	90	4.1	17	0.90
SU5	SU1740	126	4.5	**Rf5**	SU3039	90	4.1	13	0.91
SU6	SD8812	110	2.2	**Rf1**	SD6614	440	6.6	22	0.75
SU7	SP3180	119	3.1	**Rf6**	SP2186	96	3.6	12	0.79
C1	NK1345	15	5.5	**Rf7**	NJ2169	7	5.2	96	0.44
C2	NU2514	23	5.1	**Rf3**	NO4620	10	4.8	133	0.72
C3	TA1967	15	5.0	**Rf8**	TA0243	7	4.1	30	0.63
C4	NM8834	3	3.9	**Rf9**	NR6622	10	6.1	113	0.68
C5	SN2452	133	6.6	**Rf10**	SM8905	44	5.2	59	0.79
C6	SX9456	58	6.1	**Rf11**	SX4952	50	5.2	46	0.66
C7	SD3000	9	5.9	**Rf12**	SD3131	10	5.3	31	0.78
R1	NH8914	228	2.8	**Rf13**	NJ0662	5	4.6	51	0.85
R2	SE5238	8	4.3	**Rf14**	SE4961	14	3.6	24	0.57
R3	SK5026	43	3.5	**Rf4**	TF0049	63	5.1	55	0.88
R4	SO9749	35	3.5	**Rf6**	SP2186	96	3.6	44	0.87
R5	SU7349	118	4.6	**Rf5**	SU3039	90	4.1	45	0.90
R6	NS8264	277	5.9	**Rf15**	NT2302	236	3.5	74	0.91

Table 1: Summary of the meteorological monitoring sites used in this study. Target sites are defined as Urban, Sub-Urban, Coastal or Rural, reference sites are denoted as Rf. The elevation above sea level (Elev), ratio of target and reference site mean wind speeds ($\bar{u}_{tar}/\bar{u}_{ref}$), distance between target and reference sites (d) and linear correlation coefficient between the reference and target site wind speeds (r) are also shown. Anemometer heights differing from 10 m: *h = 20.6 m, **h = 22.5 m.

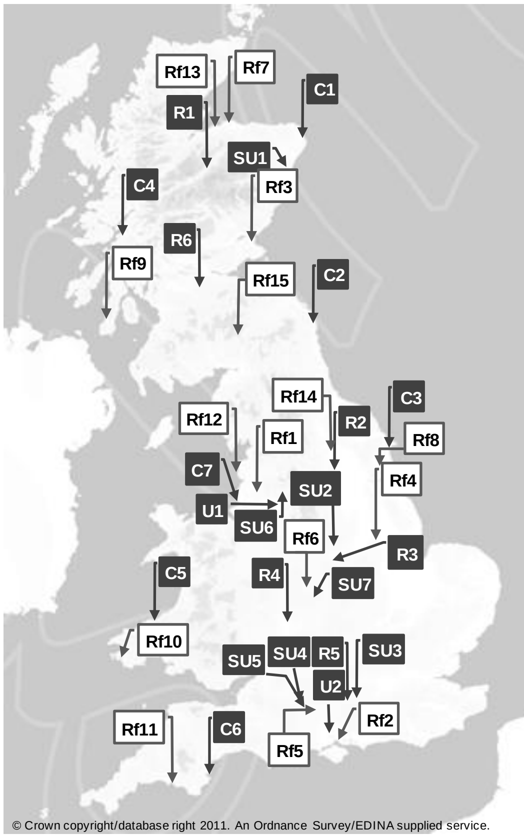

Figure 2: Approximate geographical locations of the meteorological monitoring sites used in this study. Target sites are defined as Urban, Sub-Urban, Coastal or Rural, reference sites are denoted as Rf.

2.3. Assessing the accuracy of predictions

The MCP approaches described above were used to predict the entire time-series of hourly averaged wind speeds over a 10 year prediction period at each test site. From this time series, a number of parameters related to the wind resource can be extracted. Of particular importance are the predicted mean wind speed and the mean wind power density. Also of interest are the predicted standard deviation and Weibull shape factor since they give insight into the form of the predicted wind speed distribution. These parameters are defined below.

The mean wind speed is simply the average of the long-term time series of predicted wind speeds. The mean Betz wind power density $\bar{p}_d$, defined as the mean Betz power in the wind per unit swept area is given by [22]:

$$\bar{p}_d = (16/27)0.5\rho\overline{u^3}$$

Equation 5

where ρ = 1.225 kgm^{-3} is the air density, 16/27 is the Betz limit and $\overline{u^3}$ represents the mean of the cubed wind speeds.

The standard deviation of wind speeds σ is defined as:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(u_i - \bar{u})^2}$$

Equation 6

where u_i is the i^{th} wind speed observation, $\bar{u}$ is the long-term mean wind speed and N is the total number of observations.

Along with the wind speed u and scale factor c, the Weibull shape factor k is used to define the Weibull probability distribution $f(u)$ as:

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

Equation 7

The Weibull shape factor is obtained using the method of maximum likelihood to fit a Weibull probability distribution to the predicted target site wind speeds.

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

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

Equation 8

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

Equation 9

$$\text{MBE} = \sum_j \bar{u}_{obs,j} - \bar{u}_{pred,j}\,/n$$

Equation 10

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

Note that a negative MBE represents an overestimate compared to the observed values of the parameter under consideration. Using the sliding window approach, 120 predicted time-series, and hence 120 values of the error metrics (Equation 8 - Equation 10) corresponding to each window position, were obtained for the target sites. The final error statistics were calculated as the average of these error metrics across all training/test periods.

3. Results and Discussion

Figure 3 shows hourly averaged target and reference site wind speeds for a single 30º angular sector from the target/reference site pair C6-Rf11. Since there is a large amount of variability at each site and within each angular sector, these results are intended as an example rather than to be fully representative. Wind speed observations are recorded with a resolution 0.51 ms^{-1}, hence in order to fully represent the discretized data, the wind speeds are presented as two dimensional density plots where the shading indicates the frequency of observations at each wind speed. In the case of the predicted wind speeds, (Figure 3C and D), the data are also discretized using wind speed bins of width 0.51 ms^{-1} for consistency with the recorded data.

Figure 3A shows the observed wind speeds over the three month training period along with the three regression fits. The fits are identical for LR and LR2 since the approaches only differ in the prediction phase. It can be seen that VR results in a steeper gradient than LR due to the forced increase in $\sigma(\hat{u})$ discussed previously. Figure 3B shows the recorded wind speeds over the entire 10 year prediction period and Figure 3C and D show the attempt to predict these observations using the MCP approaches. The predictions using LR and VR (Figure 3C) all lie along the straight lines defined by the regression equations. It can be seen that these predictions differ markedly from the instantaneous observations (Figure 3B) which exhibit considerable scatter. In contrast, the LR2 approach (Figure 3D) is capable of reproducing at least the general form of the scatter about the mean prediction. However, the predicted scatter has a narrower range than that observed in Figure 3B, implying that σ_{res}, as modeled from the short-term training data, is lower than the long-term observed value.

It can be seen from Figure 3D that at low reference site wind speeds, the predicted target site wind speed may be less than zero due to the effect of the residual scatter term. Previous studies have dealt with negative predictions arising from simple linear models by removing the values [17] or setting them to zero [11]. However, where an attempt is made to model the residual scatter, negative predictions are more frequent and simply removing them will reduce the number of entries at low wind speeds, resulting in a positive bias in the predicted mean wind speed and mean wind power. The opposite will be true if the values are simply set to zero. In this work we use a compromise whereby negative predictions are set to the mean value of the function before the residual scatter term is applied. In the small number of cases where the mean value of the function is also less than zero, the value is removed from the prediction.

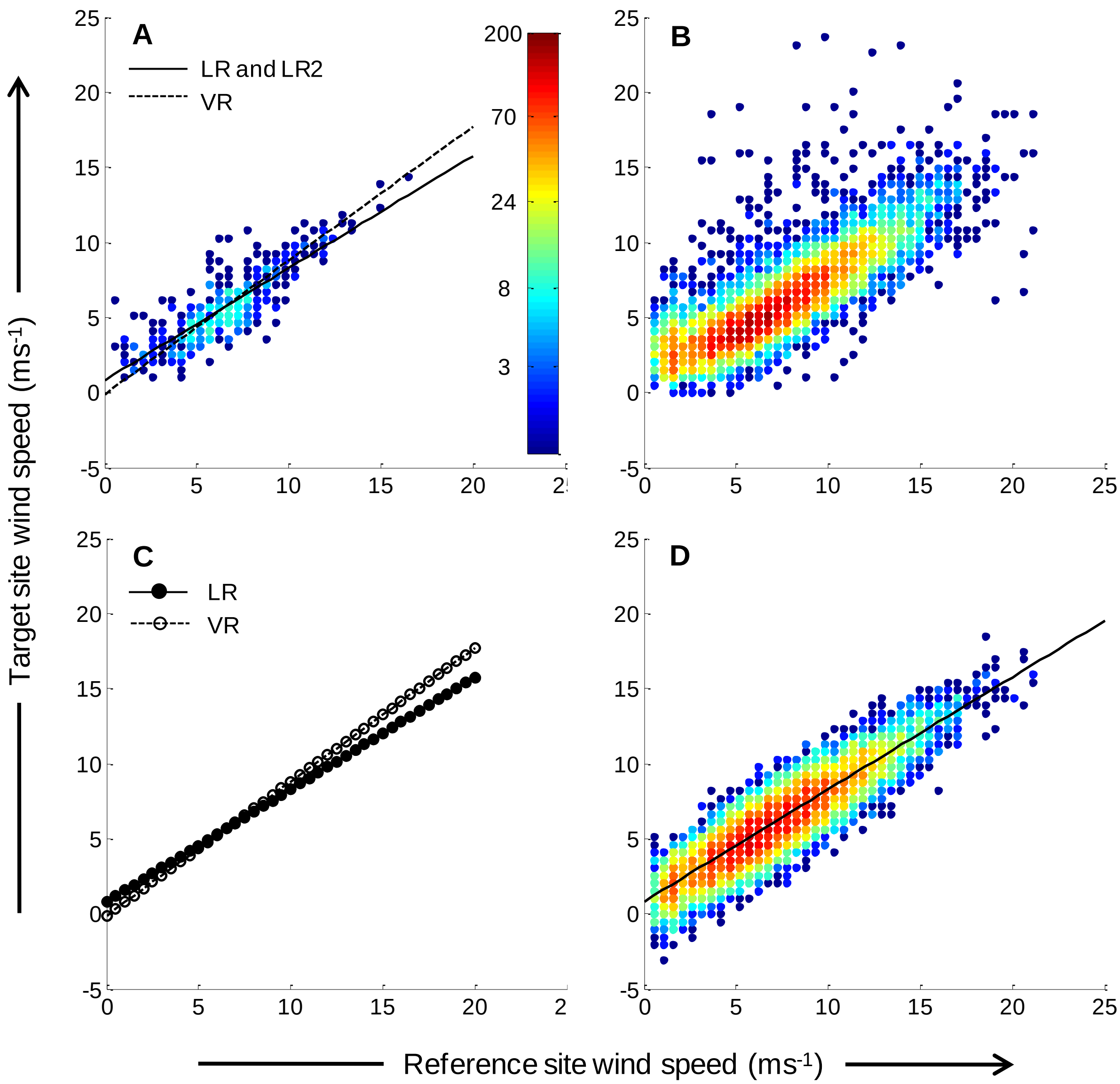

Figure 3: Target and reference site wind speeds for a single 30º angular sector from the reference/target site pair C6-Rf11. A) Observed target and reference site wind speeds over the 3 month training period along with the LR, LR2 and VR fits, B) Long-term observed target and reference site wind speeds over 10 years, C) Long-term predicted target site wind speeds over 10 years using LR and VR, D) Long-term predicted target site wind speeds over 10 years using LR2, the solid line represents the mean prediction, the dots show the individual predictions drawn from a zero mean Gaussian distribution. The shading represents the frequency of data points at each discretized wind speed.

Figure 4 shows the average $\bar{u}_{pred}$ and $\bar{u}_{obs}$ 10 year mean wind speeds at the 22 target sites using the three MCP approaches. Note these are averages over all training periods and hence only systematic biases will be visible since seasonal and inter-annual biases in the predictions will be smoothed out. The long-term mean wind speeds appear to be well predicted at all sites using each of the MCP approaches. However, there appears to be a small tendency for VR to overestimate the mean wind speed compared to the other two approaches. This is not surprising since the VR method enforces a steeper gradient compared to LR. Note that the LR and LR2 approaches result in very similar values for $\bar{u}_{pred}$ indicating that the residual scatter term does not significantly affect the mean wind speed prediction.

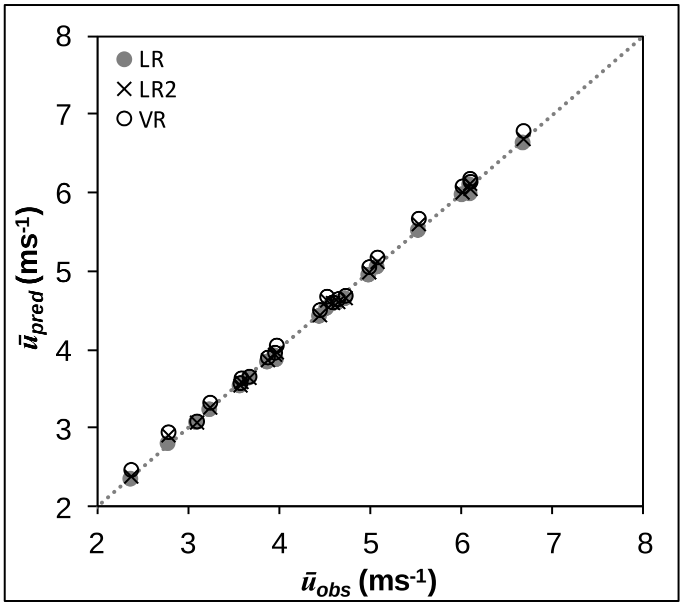

Figure 4: Predicted and observed 10 year mean wind speeds at 22 target sites using three MCP approaches averaged across all training/test periods. The dotted line shows a one-to-one relationship.

Table 2 shows the error metrics of %Error, MAE and MBE averaged across all target sites, and all training/test periods for the three MCP approaches. Predictions obtained using a semi-empirical boundary layer scaling model (SE) are also included and these are discussed further in Section 3.2. The results indicate that on average, $\bar{u}$ and $\bar{p}_d$ can be predicted to within 4.8% and 14% respectively using just three months onsite measurements with the best performing MCP approach. For the predicted $\bar{u}$ there is no clear difference between MCP approaches while for $\bar{p}_d$, the LR2 approach performs best, closely followed by VR with the largest errors observed for LR. In the case of σ and k, VR performs best indicating that this approach is more successful at predicting the wind speed distribution, again the largest errors are observed for LR. The bias errors indicate the degree to which the approaches systematically overestimate (negative bias) or underestimate (positive bias) a particular parameter. It is noteworthy that while LR results in very low bias in $\bar{u}$, the approach underestimates σ and overestimates k resulting in a narrower predicted wind speed distribution which in turn leads to a large underestimate of $\bar{p}_d$. LR2, which accounts for the residual scatter, results in the lowest bias for $\bar{p}_d$, followed closely by VR. Overall, LR2 tend to underestimates $\bar{p}_d$ while VR overestimates it.

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

Table 2: Error metrics averaged across 22 target sites and all training/test periods for three MCP approaches, (LR, LR2 and VR), as well as a semi-empirical boundary layer scaling model (SE).

In addition to the average error metrics, it is useful to consider the distribution of errors across individual sites. The distributions of percentage errors across all target sites are shown in Figure 5, averaged across all training/test periods. The error distributions for $\bar{u}$ are very similar for all approaches although VR shows a slight tendency to overestimate. For $\bar{p}_d$, which is perhaps the most significant parameter given that the aim is to predict the wind energy resource, LR can be seen to exhibit a strong positive bias (tendency to underestimate). This is expected due to the failure of LR to represent the residual scatter. On average, VR has a significantly lower bias than LR but the error distribution is negatively skewed indicating a tendency to overestimate. The best predictions (small bias and low error range) are obtained using LR2 highlighting the value of explicitly accounting for the residual scatter term.

These observations are also reflected in the error distributions for of σ and k. LR underestimates σ and overestimates k leading to a narrower predicted wind speed distribution. LR2 does significantly better at estimating these parameters although it still leads to an underestimate of the width of the wind speed distribution, likely because the short training period makes it challenging to accurately estimate σ_{res}(Equation 3). VR does particularly well at estimating σ and k resulting in a very small bias and a low error range. It is likely that the tendency of VR to slightly overestimate $\bar{u}$, which has a more significant effect on the power density compared to k [8], prevents this approach from outperforming LR2 in terms of predicted $\bar{p}_d$.

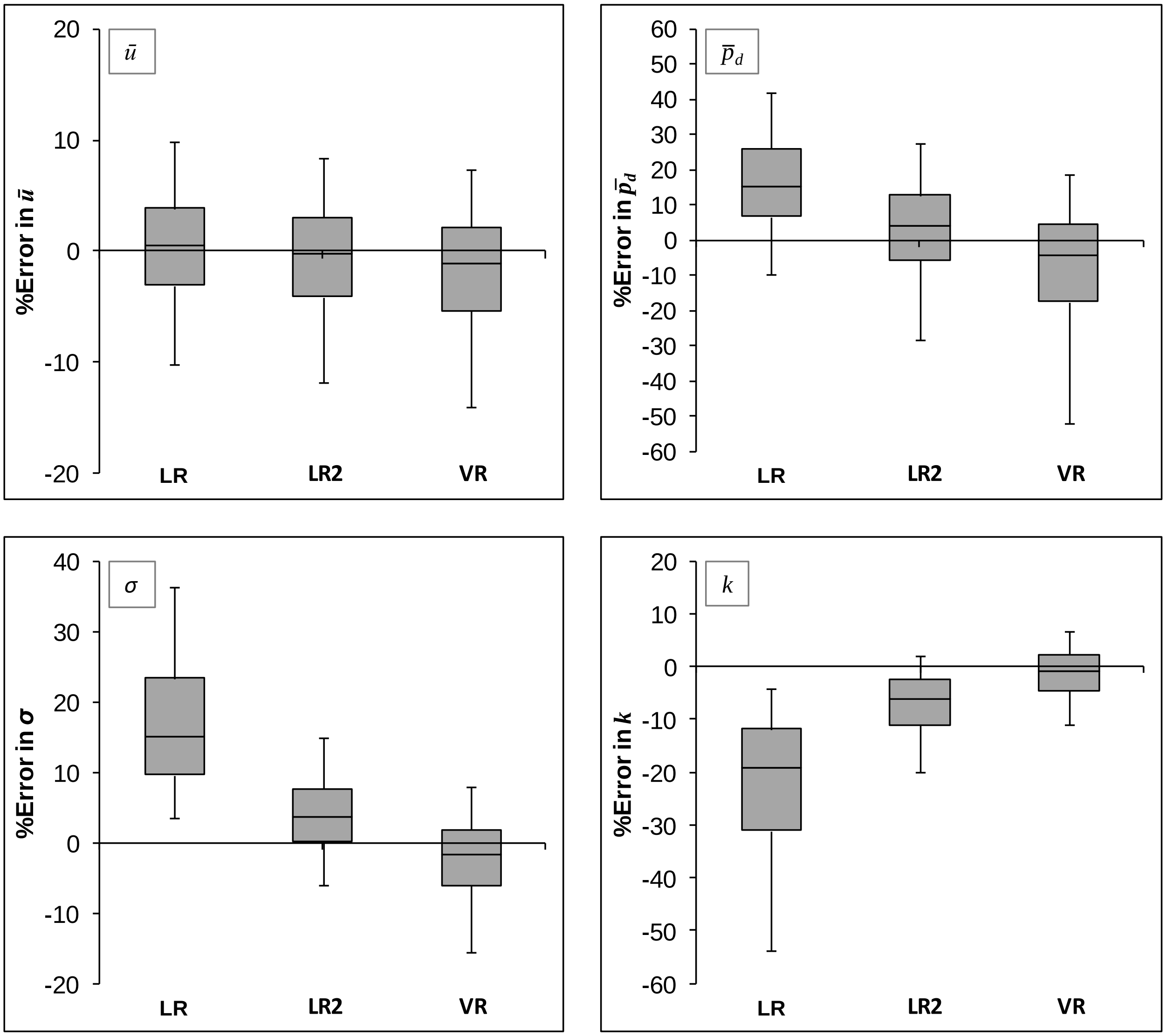

Figure 5: Distribution of percentage errors in mean wind speed $\bar{\boldsymbol{u}}$, mean wind power density $\bar{\boldsymbol{p}}_{\boldsymbol{d}}$, sample standard deviation $\boldsymbol{\sigma}$, and Weibull shape factor $\boldsymbol{k}$ averaged across 22 target sites and all training/test periods using three MCP approaches. The error bars represent the 5th and 95th percentiles, the shaded regions encloses the interquartile range.

3.1. Seasonal effects

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

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

Figure 6 shows the variation in the average percentage errors for $\bar{u}$, $\bar{p}_d$, σ, and k using different three month training periods throughout the calendar year. The vertical lines mark training periods corresponding to the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). Training data periods between these points include months which overlap more than one season. Clear seasonal variations can be seen in the prediction errors for all parameters for both LR2 and VR. For the key parameters of $\bar{u}$ and $\bar{p}_d$, the largest errors occur close to winter and summer while the smallest errors occur close to autumn and late winter/early spring (Feb – April). The results indicate that on average, significant reductions in the error of the predicted $\bar{p}_d$ can be achieved through choosing optimum seasons in which to obtain the short-term onsite measurements. For LR2, the best season results in a percentage error in $\bar{p}_d$ of ~10%, compared to ~18% for the worst season. Similarly for VR, the best season results in an error in $\bar{p}_d$ of ~12%, compared to ~20% for the worst. Interestingly, for the parameters of σ and k, which are related to the form of the wind speed distribution, the LR2 approach exhibits a stronger seasonal variation than VR with the largest errors occurring in summer.

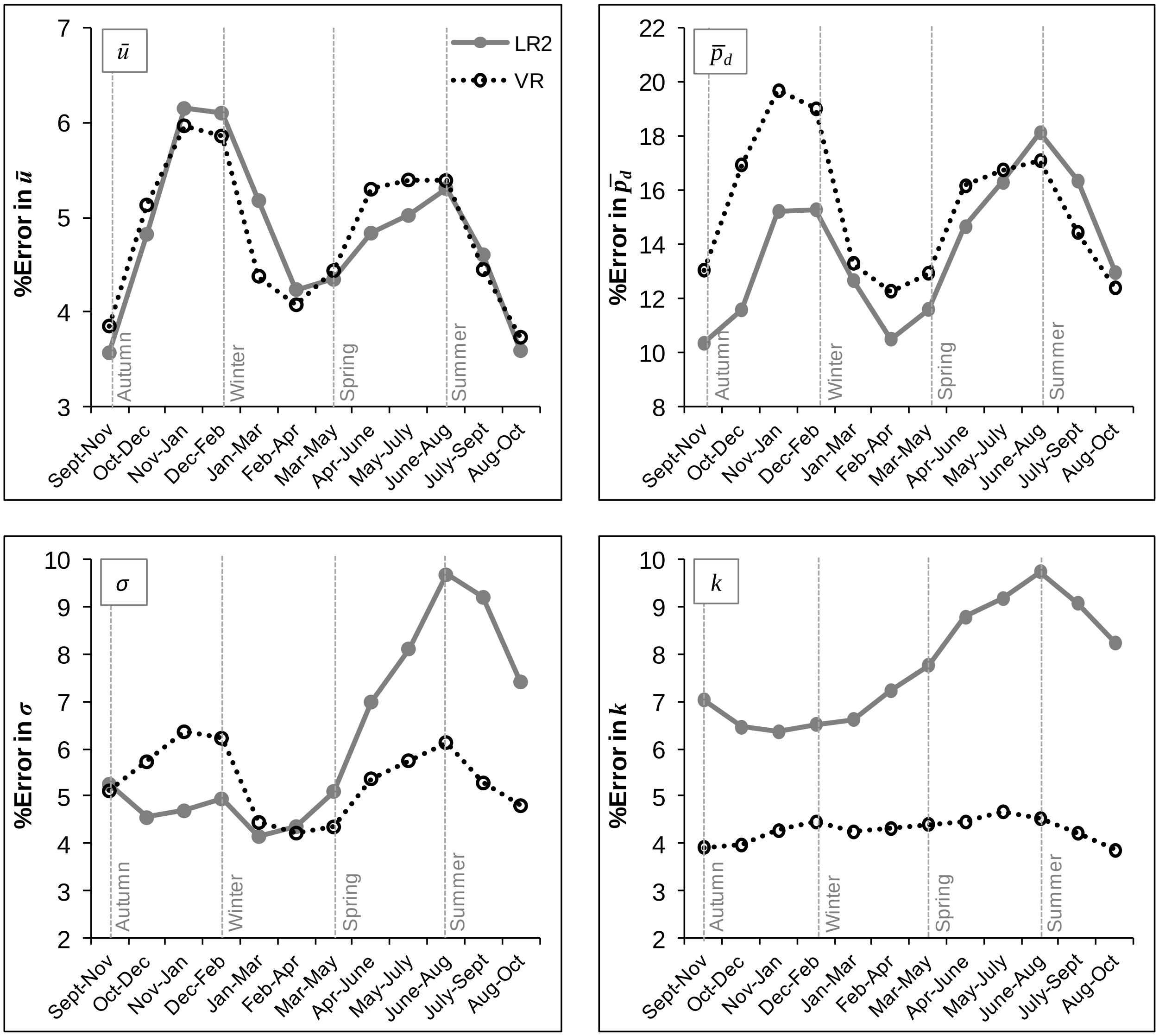

Figure 6: Seasonal variation of the percentage errors in mean wind speed $\bar{\boldsymbol{u}}$, mean wind power density $\bar{\boldsymbol{p}}_{\boldsymbol{d}}$, sample standard deviation $\boldsymbol{\sigma}$ and Weibull shape factor $\boldsymbol{k}$ averaged across 22 target sites and 10 years using two MCP approaches. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training.

Seasonal variations in the error metrics indicate that the regression parameters extracted from the short-term training data vary according to seasonal changes in synoptic weather patterns. These weather patterns could introduce a number of factors which contribute to this variability. In general, it might be expected that training periods which include the most variable weather patterns would lead to regression fits which are most representative of the long-term reference-target site relationship. Of interest in this regard is a recent study by Earl *et al.* [23] which investigated the variability in UK surface winds over a 30 year period using a network of 40 anemometer stations. While the UK surface winds are dominated by winds from the southwest, relatively large seasonal variations in wind direction were observed. In particular, spring was found to have a more significant northeasterly component leading to a more even spread of wind directions compared to other seasons.

For the 15 reference sites used in the current study, Figure 7 compares the percentage frequency of wind directions during winter and spring as averaged across the 11 year data record. The data shows similar seasonal variations in wind direction to those observed by Earl *et al*. with more significant northeasterly and easterly components during spring compared to winter. As the MCP approaches are applied sector-wise, it is likely that seasons which have a greater spread of wind angles will result in more robust estimates of the regression parameters since the sectors will be more uniformly populated. Conversely, seasons with a strongly dominant wind direction could result in poor estimates of the regression parameters for certain poorly represented sectors. This effect could be a contributing factor to the improved error statistics observed in specific seasons. While any such effect could in principle be reduced by using a single regression fit to data from all wind directions, it was found that on average the prediction errors were lower when using the sector approach.

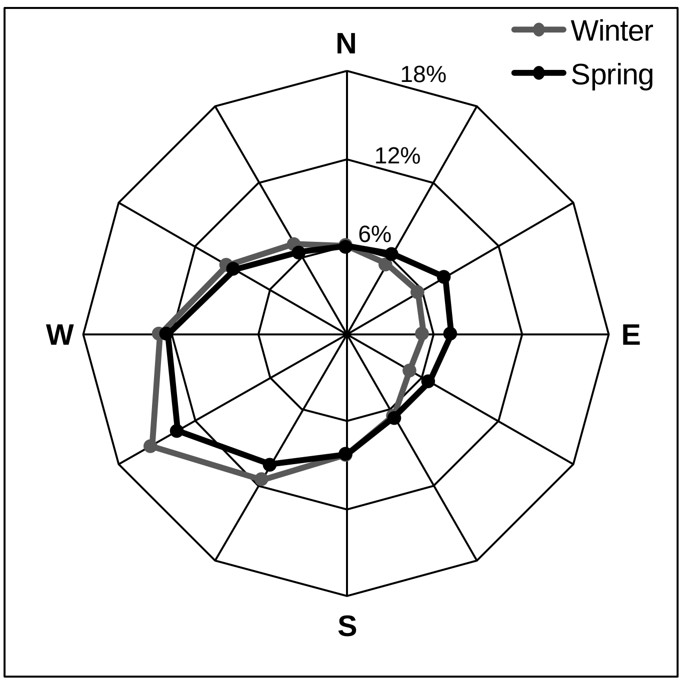

Figure 7: Percentage frequency of wind directions in 30º sectors during winter (Dec-Feb) and spring (Mar-May) averaged across the 15 reference sites and 11 year data record used in this study. The lines are included as a guide to the eye.

Another point of interest is the increased errors in σ and k observed for LR2 during summer. For the sites considered in this study, both the mean wind speed and the average variance of the wind speeds was lowest during this season. This could result in insufficient range over which to achieve an accurate least squares regression as well as impacting on the estimates of the long-term residual standard deviation σ_{res} required to reconstruct the residual scatter.

To establish whether certain seasons are more likely to result on average in over or underestimates of the wind resource, Figure 8 shows the seasonal variation in the mean bias error for $\bar{u}$ and $\bar{p}_d$. For LR2, winter training periods are more likely to result in overestimates of the long-term wind resource (negative bias) while the opposite is true for summer training periods. For VR, the same trend is visible for winter training periods while for the remaining seasons the bias is close to zero. Since Figure 6 shows that the percentage errors in $\bar{u}$ and $\bar{p}_d$ also peak for VR in the summer, this implies that for VR the sign of the error varies depending on the specific site during this season.

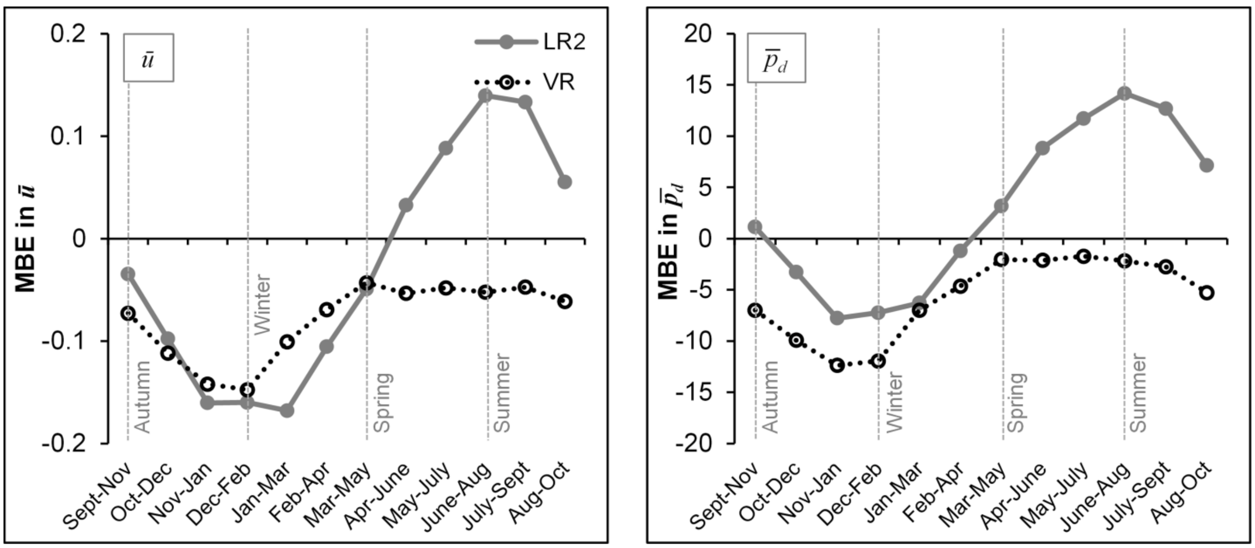

Figure 8: Seasonal variation of the mean bias error in mean wind speed $\bar{u}$ and mean wind power density $\bar{p}_d$, averaged across 22 target sites and 10 years using two MCP approaches. The vertical lines mark the nominal seasons of autumn (Sept-Nov), winter (Dec-Feb), spring (Mar-May) and summer (June-Aug). The horizontal axes show the three month period used for training.

These results indicate that on average, significant improvements in the accuracy of the predicted long-term wind resource estimated from short-term measurements can be made through choosing optimum seasons in which to obtain onsite measurements. In addition, the results also give an indication as to the average sign of the bias error as a function of the season used for training and the MCP approach used.

3.2. Comparison between modelling and data-driven approaches

Given that obtaining onsite wind speed measurements, even for a short time period, necessitates additional time and expense, it is useful to investigate to what extent a data-driven approach, based on a very short measurement period, may improve predictions of the wind resource compared to a modelling approach. The target sites used in this work formed a subset of sites considered in a previous study [8] which evaluated the accuracy of a semi-empirical, boundary layer scaling model (SE) that was developed by the UK Met Office [7] for predicting the UK wind resource. Hence, it is possible to make a direct comparison between the accuracy of the two approaches for these sites. The average error metrics are compared in Table 1 for SE and the three MCP approaches. The SE predictions were obtained using the improvements suggested in reference [8] which included taking account of the angular dependent upwind roughness.

Table 1 shows that despite the very short training period, all the MCP approaches result in a clear improvement in all of the average error metrics. For example, using LR2 compared to SE reduces the average percentage error in the predicted $\bar{p}_d$ from 26% to 14% and halves the MAE from 22 wm^{-2} to 11 wm^{-2}. It should be noted that modelling approaches are still of significant value in that they can be easily implemented in a scoping context with little prior investment. However, the results presented here indicate that the additional time and investment required for short-term onsite measurements accompanied by MCP analysis is well justified in cases where investors require greater confidence in the predicted wind resource.

4. Conclusions

We have investigated the feasibility of predicting the long-term wind resource over 10 years at 22 UK sites using MCP approaches based on short-term onsite measurements covering just three months. Using a sliding window approach over an 11 year data period, robust error statistics have been obtained which account for both inter-annual and seasonal variations. The results indicate that while such a short measurement period introduces additional challenges including seasonal variations and reduced data coverage, the approach can be successfully applied in wind resource prediction.

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

Analysis of the sensitivity of the wind resource predictions to the season in which the onsite wind speed measurements were obtained revealed clear seasonal variations in both the sign and the magnitude of the prediction errors. The results indicate that on average in the UK, the lowest prediction errors are obtained when using either autumn or early spring as the training period, while the highest errors are obtained when using winter or summer. For a three month training period, choosing the optimum measurement season can result in an average improvement of 8 percentage points in the predicted wind power density compared to the worst season.

Comparison between the MCP approaches presented in this work and a previously developed semi-empirical model demonstrate that large improvements can be made in predicting the long-term wind resource using the MCP approaches even with just three months onsite wind speed measurements. Across 22 UK sites, the best performing MCP approach resulted in mean absolute and percentage errors of 4.8% and 0.21 ms^{-1} respectively for $\bar{u}$ and 14% and 11 wm^{-2}

for $\bar{p}_d$. By way of comparison, the modelling approach resulted in errors of 9.5% and 0.42 ms^{-1} for $\bar{u}$ and 26% and 22 wm^{-2} for $\bar{p}_d$.

These results indicate that in addition to modelling techniques, MCP approaches used with very short onsite measurement periods have the potential to be a valuable addition to the wind resource assessment toolkit for small-scale wind developers. This is particularly true in cases where investors require greater confidence in the predicted wind resource and can thus justify the additional expense of a short measurement campaign.

Acknowledgements

This work was financially supported by the Engineering and Physical Sciences Research Council through the Doctoral Training Centre in Low Carbon Technologies. The authors are grateful to the UK Met Office and the British Atmospheric Data Centre (BADC) for providing access to the wind data used in this study.

References

[1] HM Government. Energy Act 2004, 2004.
[2] Department of Energy and Climate Change. Microgeneration Strategy, June 2011.
[3] OPSI 2008. Climate Change Act 2008. Available from: http://www.opsi.gov.uk/acts/acts2008/ukpga_20080027_en_1.
[4] RenewableUK. Small Wind Systems: UK market report, 2011.
[5] World Wind Energy Association. Small Wind World Report, 2012.
[6] AWS Scientific Inc. & National Renewable Energy Laboratory (U.S.). Wind Resource Assessment Handbook - fundamentals for conducting a successful monitoring program, 1997.
[7] UK Meteorological Office. Small-scale wind energy - Technical Report, 2008.
[8] S.M. Weekes, A.S. Tomlin, Evaluation of a semi-empirical model for predicting the wind energy resource relevant to small-scale wind turbines. Renew. Energy, 2013;50:280-288.
[9] C. Wilson, J. Standen, The added value and validation of mesoscale models compared to atmospheric reanalyses for resource assessment. European Wind Energy Conference, Vienna, 2013.
[10] UK Meteorological Office. Virtual Met Mast verification report, September 2012.
[11] A.L. Rogers, J.W. Rogers, J.F. Manwell, Comparison of the performance of four measure-correlate-predict algorithms. Journal of Wind Engineering and Industrial Aerodynamics, 2005;93:243-264.
[12] M.A. Lackner, A.L. Rogers, J.F. Manwell, Uncertainty analysis in MCP-based wind resource assessment and energy production estimation. J. Sol. Energy Eng. Trans.-ASME, 2008;130:10.
[13] A. Joenson, L. Landberg, H. Madsen, A new measure-correlate-predict approach for resource assessment. Proceedings of the EWEC, 1999.
[14] A. Derrick, Development of the measure-correlate-predict strategy for site assessment. Proceedings of the BWEA, 1992.
[15] A. Romo Perea, J. Amezcua, O. Probst, Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource. Journal of Renewable and Sustainable Energy, 2011;3.
[16] M.A. Lackner, A.L. Rogers, J.F. Manwell, The round robin site assessment method: A new approach to wind energy site assessment. Renew. Energy, 2008;33:2019-2026.
[17] S. Velazquez, J.A. Carta, J.M. Matias, Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kW h produced by a wind turbine at a candidate site: A case study in the Canary Islands. Applied Energy, 2011;88:3869-3881.

[18] J.H. Bass, M. Rebbeck, L. Landberg, M. Cabre, A. Hunter. An improved measure-correlate-predict algorithm for the prediction of the long term wind climate in regions of complex environment, Joule Project JOR3-CT98-0295, 2000.
[19] J.C. Woods, S.J. Watson, A new matrix method of predicting long-term wind roses with MCP. Journal of Wind Engineering and Industrial Aerodynamics, 1997;66:85-94.
[20] S.L.R. Ellison, V.J. Barwick, T.J.D. Farrant, Practical statistics for the analytical scientist, a bench guide, RSC Publishing, 2009.
[21] UK Meteorological Office. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853 - current). NCAS British Atnospheric Data Centre 2012.
[22] P. Ramirez, J.A. Carta, Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study. Energy Conv. Manag., 2005;46:2419-2438.
[23] N. Earl, S. Dorling, R. Hewston, R. von Glasow, 1980-2010 Variability in UK Surface Wind Climate. J. Clim., 2013;26:1172-1191.