

Characterizing Wind Power Forecast Uncertainty with Numerical Weather Prediction Spatial Fields

by

Nicholas J. Cutler, Jeffrey D. Kepert, Hugh R. Outhred, Iain F. MacGill

REPRINTED FROM

WIND ENGINEERING

VOLUME 32, No. 6, 2008

MULTI-SCIENCE PUBLISHING COMPANY
5 WATES WAY · BRENTWOOD · ESSEX CM15 9TB · UK
TEL: +44(0)1277 224632 · FAX: +44(0)1277 223453
E-MAIL: mscience@globalnet.co.uk · WEB SITE: www.multi-science.co.uk

Characterizing Wind Power Forecast Uncertainty with Numerical Weather Prediction Spatial Fields

Nicholas J. Cutler^{*1}, Jeffrey D. Kepert², Hugh R. Outhred¹, Iain F. MacGill¹

¹University of New South Wales, Sydney 2052 Australia, *n.cutler@unsw.edu.au

²Bureau of Meteorology, GPO Box 1289, Melbourne, Victoria, 3001, Australia

ABSTRACT

Successful integration of wind power into power systems can be facilitated through better understanding of future uncertainty in wind power generation. This paper explores a new approach to characterizing this uncertainty using measures of the variability in the wind speeds predicted at multiple grid points in a Numerical Weather Prediction (NWP) system. This approach is compared to the traditional approach of using the spread from an NWP ensemble by using two measures of uncertainty; forecast errors in single time-series forecasts and observed temporal variability. Results show that the multiple grid point approach has a comparable skill level to NWP ensembles for predicting these uncertainty measures and in particular, demonstrates very good skill in predicting large forecast errors. These results also provide a positive evaluation of a terrain standardization method described in a companion paper. A possible extension of this work is to combine the multiple grid point approach with NWP ensembles to improve uncertainty characterization.

I. INTRODUCTION

One of the ways to facilitate the high penetrations of wind power in a power system is to forecast wind power for lead times of up to 48 hours. Many wind power forecasting studies to date have focused on producing and evaluating forecasts of expected values, in the form of single time-series forecasts, eg. [1, 2]. However, the limitations of NWP models and the stochastic nature of the weather can give rise to large errors in such single time-series forecasts under some conditions. The potential for large errors limits the value of single time-series forecasts in assisting decision-making for maintaining power system security [3]. A characterization of the future uncertainty in such forecasts can improve this situation [4].

Many wind power forecasting studies argue that the main contribution to the error in single time-series forecasts comes from the Numerical Weather Prediction (NWP) system [5, 6]. The most widely used approach to characterize the uncertainty in NWP forecasts is to run an ensemble of NWP forecasts to capture one or more of the major sources of uncertainty. Some wind power forecasting studies have used NWP ensembles based on varying the initial conditions [7, 8], while others have varied the physical model parameterizations [6, 9], both of which require computationally expensive multiple NWP system runs. Others studies have created ensembles using completely different NWP systems [10] or temporal ensembles [11]. However, these ensembles typically contain less than ten members, which may be too few to sufficiently characterize the forecast uncertainty. Other methods in use to characterize wind power forecast uncertainty include synoptic classification [4] and categorizing the non-linear

power curve into different parts [12]. This paper explores a new approach using measures of the variability in the forecast wind speeds at multiple grid points in an NWP system. This is in effect an NWP ensemble based on multiple grid points and does not entail any extra computational effort since it only requires data from one NWP run.

A companion paper [13] introduced a chosen-grid-point-equivalent (cgp-equivalent) methodology where forecast wind speeds at wind turbine hub height for multiple grid points around a wind farm location are transformed to equivalent wind speeds that represent the surface roughness and terrain at a chosen grid point (*cgp*). The *cgp* is chosen as the most representative of the wind farm location. The cgp-equivalent methodology is used to develop a visually-based decision support tool to characterize future uncertainty, particularly in regards to large, rapid changes in wind power. The companion paper discusses how the uncertainty sources in NWP systems mentioned above often result in the spatial misplacement of synoptic features in the forecasts.

This study builds on the work presented in the companion paper by providing some evidence that misplacement errors in NWP forecasts make a significant contribution to forecast uncertainty. Two ex-post quantities to describe forecast uncertainty are used; the single time-series forecast error and observed temporal variability. This paper examines how a measure of the spatial variability in NWP wind speeds could be to predict these quantities, compared with the spread of an NWP ensemble. Neither of these quantities for measuring forecast uncertainty are a complete measures of the underlying forecast uncertainty, which is probably impossible to achieve. However, comparing the spatial variability with the single time-series forecast error illustrates the significance of misplacement errors, while comparing with temporal variability illustrates the extent to which uncertainty due to the misplacement of moving weather features is translated into temporal uncertainty at a site. A further outcome of this work is that comparing the spatial variability calculations made with cgp-equivalent and raw NWP spatial fields, allows a quantitative assessment of the cgp-equivalent methodology.

2. DATA SOURCES

The observation data used for the results in this paper are obtained from two wind farm sites within Australia; Yambuk (YB) and Challicum Hills (CH). Both wind farms are owned and operated by Pacific Hydro Pty Ltd and are located in the Australian state of Victoria. Figure 1 shows the locations of the two sites and table 1 summarizes some important characteristics of the sites. The wind speed observation data used for the wind farms is the average wind speed measured on the wind turbine nacelles. Although it is known that wind speed observations made with nacelle based anemometry are affected by the moving blades and the nacelle itself [14], it has been shown that an averaged measurement from these anemometers is much more representative wind speed for characterizing a wind farm power curve than a wind speed measured on a met mast at or near the wind farm [15]. This is likely to be due to the benefits of spatial smoothing. All observations are averaged to hourly values using the observation data obtained for half an hour before and after the hour. This definition is chosen as one of the possibilities to be representative of the NWP model forecasts, which are available on the hour, every 3 hours.

The single time-series and spatial field NWP forecasts used for this work come from a new NWP model developed by the Australian Bureau of Meteorology; Mesoscale Assimilation Limited Area Prediction System (MALAPS), which went operational in December 2007. MALAPS has a 0.1° (≈ 10 km) resolution in the horizontal, 61 levels in the vertical and has its own analysis at the same resolution [16]. The system was developed partly for wind power

forecasting applications, and at least one verification study on MALAPS includes the use of wind power monitoring data [16]. The MALAPS data used is for 7.5 weeks from 2nd January to 24th February 2007, and 36 weeks from 1st June 2007 to 10th February 2008. For each day within these periods, prediction horizons from 6 to 48 hours are used (available in 3-hour increments) from each of two available MALAPS forecast runs made at 0:00 and 12:00 UTC (Coordinated Universal Time). The total period of 10.5 months of available MALAPS data gives in excess of 8000 hourly data points for each site when combining the forecasts at multiple prediction horizons with the available observation data. The chosen grid points (including sigma levels¹) used for each site were chosen to minimize the bias between the forecasts and the observations.

Table 1 presents a summary of the sites and [13] describes them in more detail. The mean wind speeds for each site were around 8 ms^{-1} with a standard deviation of 3 ms^{-1} over the 10.5 months of data. More precise figures for these statistics could not be shown due to confidentiality reasons and they would provide very little additional information for the analysis in this paper.

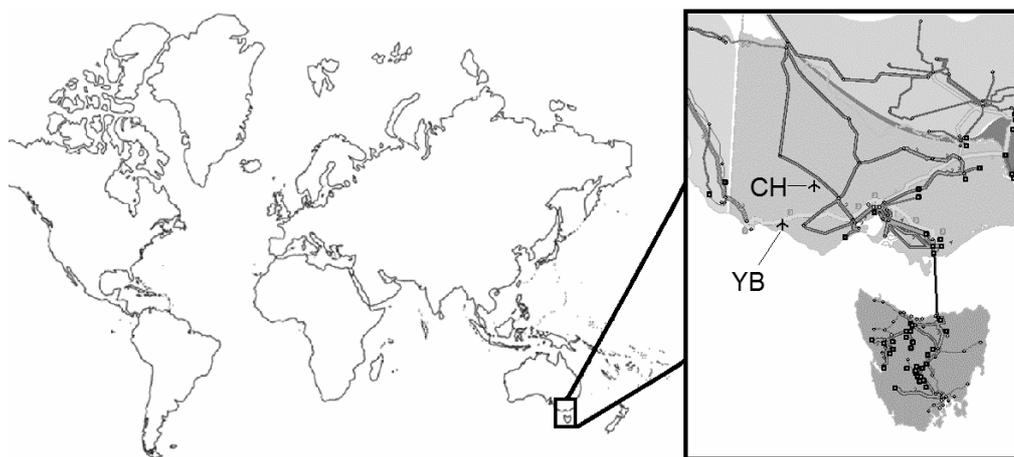


Figure 1: Location of south eastern Australia with an inset of the particular region being studied showing the major transmission lines and the relative locations of the two wind farm sites (source of transmission map: [17])

The ensemble data used as a benchmark for the results is from the Multi-Scheme Ensemble Prediction System (MSEPS), which is run by Danish company, Weprog ApS [18]. MSEPS has 75 ensemble members, which differ by the physical assumptions (parameterizations) made. The model is currently forecasting over the Australian region with a resolution of 0.45° ($\approx 45 \text{ km}$) and 32 vertical levels. A concurrent data set of MSEPS wind speed forecasts for each ensemble member at a representative grid point for each site is constructed such that the base times and prediction horizons match with the corresponding MALAPS forecasts. Only two sigma levels¹ from MSEPS were available for this study and the chosen levels (to minimize bias with the observations) are shown for each site in table 1. The latitudes and longitudes of grid points used for the sites were chosen by Weprog after their own analysis. Unfortunately no NWP data set containing ensemble data at multiple grid points was available for this work.

¹In MALAPS and MSEPS, the 0 sigma level is defined as 0 pressure.

Table 1: Summary of main attributes for the wind farm data

Attribute	YB	CH
Wind farm location type	Coastal	Inland
Wind farm rating	30 MW	52.5 MW
Number of turbines	20	35
Site elevation (average)	35 m	470 m
Measurement height	69 m	70 m
Nearest MALAPS grid point		
MALAPS latitude	38.3° S	37.4° S
MALAPS longitude	142° E	143.1° E
MALAPS sigma level	0.995	0.972
Nearest MSEPS grid point		
MSEPS latitude	38.65° S	37.33° S
MSEPS longitude	142.17° E	143.09° E
MSEPS sigma level	0.996	0.988

3. DEFINITIONS OF QUANTITIES

The main conjectures under investigation in this work are concerned with evaluating the degree of correspondence between a predictor and an ex-post measure of forecast uncertainty. The three predictors can be summarized as follows.

1. $\hat{V}^{fld,raw}$: A variability measure of the raw (direct model output) wind speeds predicted at multiple grid points from MALAPS in the vicinity of a site - see equation (1) below.
2. $\hat{V}^{fld,cgpe}$: As above, except on chosen-grid-point-equivalent (cgpe) wind speeds from MALAPS (adjusted wind speeds to standardize the effects of terrain) - see equation (1) below.
3. \hat{V}^{ens} : A variability measure of the wind speeds predicted by the ensemble members of MSEPS for a site. This is the variance (non-weighted) of the 75 wind speeds predicted at a representative grid point for the site.

Two different measures of forecast uncertainty are used to compare the three predictors and these are summarized as follows.

1. $STDE$: The standard deviation of forecast errors comparing the wind speed observations at a site and the single time-series forecast from the MALAPS single grid point chosen for that site - see equation (4) below.
2. V^{obs} : A measure of the temporal variability in the observations over a 5 hour period - see equation (5) below.

The following describes the variability measures and forecast uncertainty measures in more detail.

A predicted wind speed field is defined as the set of wind speeds predicted at multiple grid points at a constant sigma level in an NWP system. The spatial variability is defined as a weighted variance of the wind speeds in the field, as in equation (1). The weighting is chosen to decrease with distance from the chosen grid point with a cosine function, since a cosine function depicts an assumed nature of the change in the probability of misplacement with distance. The first zero points of the cosine function are defined at a prescribed distance above which the weighting is zero. This prescribed distance is chosen to be 140 km, which allows for the representation of fast moving features (of up to 70 kmh⁻¹) in the field if they are forecast with a timing error of up to two hours.

$$\hat{V}_{th}^{fld} = \frac{\sum_{\forall ij} w_{ij} \left(\hat{S}_{ijth} - \bar{S}_{w(ij),th} \right)^2}{\sum_{\forall ij} w_{ij}} \tag{1}$$

where the subscripts i and j refer to a specific grid point by its latitudinal and longitudinal position respectively, the subscript t refers to the valid time and h refers to the prediction horizon, \hat{S} is the predicted wind speed, $\bar{S}_{w(ij),th}$ is the weighted mean (weightings defined in equation (2)) of the wind speeds predicted at each grid point ij , valid at time t with prediction horizon h , and w_{ij} is the weighting of the grid point ij , and is defined by the following equation.

$$w_{ij} = \begin{cases} \cos\left(\frac{\pi}{2} \times \frac{D_{ij,cgp}}{140}\right); & D_{ij,cgp} \leq 140 \text{ km} \\ 0; & \text{otherwise} \end{cases} \tag{2}$$

where

$D_{ij,cgp}$ is the great-circle distance in kilometers along the earth's surface between the grid point ij and the grid point chosen as most representative of the wind farm site (cgp).

Many standard average time-series forecast error measures have been trialed for this work including the root mean square error (RMSE). However, RMSE is influenced by bias and although the chosen grid point from MALAPS minimizes forecast bias, no additional downscaling techniques were applied to further correct for it. Hence for this study the standard deviation of the forecast error, STDE, is used to measure single time-series forecast error since it is independent of bias. This allows a more direct comparison between MALAPS and MSEPS. Note, however that the results using RMSE and STDE proved to be very similar. The absolute value of the single time-series forecast error, (ATFE), is defined in equation (3) and STDE is defined in terms of ATFE in equation (4).

$$ATFE_{th} = \left| \hat{S}_{cgp,th} - S_t \right| \tag{3}$$

where

$\hat{S}_{cgp,th}$ is the wind speed forecast at the chosen grid point, cgp , to represent the site that is valid at time t , with prediction horizon h .

$$STDE_B = \sqrt{\frac{\sum_{t,h \in B} \left(ATFE_{th} - \frac{\sum_{t,h \in B} ATFE_{th}}{N_B} \right)^2}{N_B}} \tag{4}$$

where

B is a subset of the forecast data and N_B is the total number of data in B , and $t,h \in B$ refers to all forecasts valid a time t , with a prediction horizon h , that belong to the set B .

The observed temporal variability in wind speed, V^{obs} , is calculated every 3 hours to coincide with the available MALAPS forecast data. V^{obs} is calculated as the standard deviation of 5 consecutive hourly wind speed observations from 2 hours before to 2 hours after the 3-hourly time stamp of the MALAPS data, as shown in equation (5). Although this definition is independent of the chronology of the 5 observations, this definition matches more closely the above definitions of spatial variability and ensemble spread.

$$V_t^{obs} = \sqrt{\frac{\sum_{k=t-2}^{t+2} (S_k - \bar{S}_{t5})^2}{5}} \quad (5)$$

where

$$\bar{S}_{t5} = \frac{\sum_{k=t-2}^{t+2} S_k}{5} \quad (6)$$

The main method used to compare the three predictors involves converting them to categorical forecasts and then using standard categorical forecast skill scores. In this work the forecast categories are “yes” or “no”, referring to whether the forecast uncertainty (ATFE or V^{obs}) exceeds a particular threshold or not. To assign categories for the predictor data, we need to select the number of predictor data that will be assigned to the “yes” category, starting from the largest predictor value. We have selected this number to be three times the number of data for which the forecast uncertainty exceeds the threshold. The multiplier of three was chosen after trialing several multipliers between one and four, as three gave the best skill scores. Note that many occurrences of forecast uncertainty above the threshold may still be classified in the “no” category. Some of the experiments are for predicting the forecast uncertainty to be *below* a particular threshold and in this involves a similar methodology. The skill scores to evaluate categorical measures chosen for this paper are the Probability of Detection (POD), False Alarm Ratio (FAR) and Odds Ratio Skill Score (ORSS). POD and FAR are the most common categorical error measures used in weather forecasting and are described in many publications, including [19]. POD is the fraction of “yes” forecasts that are correct and FAR is the fraction of “yes” forecasts where the event did not occur. ORSS is described in detail in [20], and measures the odds of a “yes” forecast being right to the odds of a “yes” forecast being wrong. ORSS has the advantage that it can be used to directly compare forecasts of events with different probabilities. Categorical forecasts are starting to be used for the evaluation of forecasts of large changes in wind power, eg. [21]. POD was used in [3] and POD, FAR and a version of ORSS were used in [6].

To test for significance, a Monte Carlo test at the 1% level [22] provides a useful comparison. A Monte Carlo test makes the assumption that the data is from a certain distribution and seeks to identify where the data item lies in that distribution. It involves producing a set of random results through independent random sampling from that distribution. A Monte Carlo test at the 1% level involves selecting the random result that is better than 99% of all random results in the set. If the result using observed data is better than the 1% level random result then it is said that the Monte Carlo test result is significant at the 1% level.

The results in this paper are essentially concerned with comparing the rank of the predictors with forecast uncertainty. Hence, for the Monte Carlo test, each random result uses a random permutation of the rank of the data. [22] describes how 500 random results are enough to perform an acceptable test at the 1% level. To be conservative, 1000 random results were generated for all Monte Carlo tests in this paper. Hence, all Monte Carlo 1% level results shown in this paper are the 10th best value from 1000 random results produced from 1000 predictors using random ranking orders.

4. HYPOTHESES

The hypotheses assessed in this paper are numbered as follows, along with the tests used in this paper to provide evidence for each. In the context of this paper, degree of correspondence

means that the larger the values of the predictors, the higher the probability that the forecast uncertainty is large (and visa versa).

1. Misplacement of weather features in NWP forecasts is a significant source of forecast uncertainty.
 - a. $\hat{V}^{fld, cgpe}$ has a degree of correspondence with temporal variability, V^{obs} , that is significant compared with a Monte Carlo 1% test. If true, this indicates that a significant amount of spatial variability in the wind is translated to temporal variability observed at a site.
 - b. The NWP ensemble spread, \hat{V}^{ens} , has a degree of correspondence with V^{obs} that is significant compared with a Monte Carlo 1% test. If true, this indicates that a significant amount of variability between the NWP ensemble members stems from misplacing weather features with a high spatial variability in the wind.
 - c. $\hat{V}^{fld, cgpe}$ has a degree of correspondence with ATFE that is significant compared with a Monte Carlo 1% test.
 - d. $\hat{V}^{fld, cgpe}$ has a degree of correspondence with ATFE that is comparable to \hat{V}^{ens} .
 - e. Large values of $\hat{V}^{fld, cgpe}$ have a strong correspondence with large values of ATFE. \hat{V}^{ens} and the Monte Carlo 1% level are used for comparison. The hypothesis here is that a significant number of large single time-series forecast errors are due to the misplacement of large spatial variability in the wind.
2. The terrain standardization method from the companion paper [13] performs well in transforming the wind speeds at multiple grid points to equivalent values that correctly represent the terrain at the chosen grid point.
 - a. $\hat{V}^{fld, cgpe}$ has a stronger degree of correspondence with forecast uncertainty (either measure) than $\hat{V}^{fld, raw}$ since $\hat{V}^{fld, cgpe}$ only depends on spatial variability due to the weather across a region while $\hat{V}^{fld, raw}$ also includes terrain-related effects on spatial variability.
 - b. Given that there is a significant relationship between forecast uncertainty and the spatial variability due to the synoptic weather phenomena, $\hat{V}^{fld, cgpe}$ should provide a more useful improvement over $\hat{V}^{fld, raw}$ in distinguishing large values of forecast uncertainty, than for distinguishing small values of forecast uncertainty. A large spatial variability implicitly implies that there are strong wind speeds in the field, and values of $\hat{V}^{fld, cgpe}$ are likely to differ more greatly from $\hat{V}^{fld, raw}$ in these cases since the terrain will have a larger contribution to $\hat{V}^{fld, raw}$. Conversely, small values of $\hat{V}^{fld, raw}$ are likely to occur when most wind speeds in the field are quite weak and near zero. In this case the terrain standardization is not likely to affect the wind speeds much and $\hat{V}^{fld, cgpe}$ will also have a small value.

5. RESULTS AND DISCUSSION

5.1. Absolute single time-series forecast error (ATFE)

This section compares the performance of the various predictors in forecasting the absolute single time-series forecast error (ATFE). We first assess a general rank correlation statistic between the predictors and ATFE and then assess the ability of the predictors to distinguish between very large and very small values of ATFE.

The degree of correspondence between two data sets can be calculated using Kendall's ranked correlation statistic, τ [23]. τ has a value between -1 and 1, where 1 means that the rank orders of the two data sets are exactly the same, and -1 where the ranks are exactly opposite. In between -1 and 1, the value for τ is linearly related to the number of inversions of pairs of objects required to transform the rank of one data set into the rank of the other. The statistic is calculated between each of the three predictors and ATFE using the method described in [24], and the results are shown in table 2. The Monte Carlo 1% test result is also shown and has a value less than half of all the values for the predictors indicating that they all have a significant degree of correspondence (hypothesis 1.c). \hat{V}^{ens} gives the largest correlation at each site by around 0.02. However compared to the difference between the correlations for all predictors and the Monte Carlo results, the results for the three predictors could be considered comparable (hypothesis 1.d). $\hat{V}^{fld,cgpe}$ gives a larger value compared with $\hat{V}^{fld,raw}$ for YB but not for CH giving an inconclusive evaluation of the terrain standardization procedure (hypothesis 2.a). However, these correlations are very similar compared with the other values in the table. Furthermore, overall ranked correlations with forecast uncertainty may not be the best way to test hypothesis 2 since the terrain standardization procedure is only expected to improve the correlation for large values of forecast uncertainty. As explained in hypothesis 2.b, a large spatial variability implicitly implies that there are strong wind speeds in the field, and values of $\hat{V}^{fld,cgpe}$ are likely to differ more greatly from $\hat{V}^{fld,raw}$ when there are strong wind speeds in the field since the terrain will have a larger contribution to $\hat{V}^{fld,raw}$.

Table 2: Kendall's correlation coefficients between predictors and ATFE at each site

Predictor	YB	CH
$\hat{V}^{fld,raw}$	0.088	0.080
$\hat{V}^{fld,cgpe}$	0.089	0.078
\hat{V}^{ens}	0.107	0.098
Monte Carlo 1% test	0.017	0.017

Figure 2 presents a cumulative STDE plot, which shows how the three predictors compare in distinguishing the largest single time-series forecast errors from MALAPS at Yambuk. The $\hat{V}^{fld,raw}$ plot, for example, is constructed by sorting the data on $\hat{V}^{fld,raw}$ from the largest to smallest values. Then the data with the 300 largest values are taken to define a subset B and $STDE_B$ is calculated as described by equation (4) to obtain the left-most value in the $\hat{V}^{fld,raw}$ plot. Successive cumulative subsets are taken in hundreds (eg. the 400 largest values, then 500 largest values, and so on) and $STDE_B$ is calculated for each to produce the entire $\hat{V}^{fld,raw}$ plot. The bar plot in figure 2 indicates the number of data used in each cumulative subset of data. The STDE values for each predictor converge to the overall STDE of around 2.4 ms^{-1} when all data is used. The upper and lower x-axes show the minimum values for some subsets for the predictors $\hat{V}^{fld,cgpe}$ and $\hat{V}^{fld,raw}$, respectively.

The cumulative STDE plot is a good test for many of the hypotheses as described in the previous section. Firstly, all three variability predictors give significantly larger values for STDE compared with the Monte Carlo 1% level indicating a significant degree of correspondence (hypothesis 1.c). Next, the x-axes in figure 2 show that the terrain standardization procedure has reduced the overall wind speed variability in the fields since $\hat{V}^{fld,raw}$ has consistently larger values than $\hat{V}^{fld,cgpe}$. Furthermore, the largest values of $\hat{V}^{fld,cgpe}$ give much larger values for STDE than $\hat{V}^{fld,raw}$, showing that the cgpe-equivalent field is better able to discriminate large errors (hypothesis 2.b). The two plots do not converge until around 2600 data (32%) are used. Also, the $\hat{V}^{fld,cgpe}$ plot shows comparable values of STDE compared with \hat{V}^{ens} with slightly larger values for the first 1800 data (22%). This excellent result is a good

indication that hypotheses 1.d and 1.e are valid. It must be reiterated however, that the difference in the results between $\hat{V}^{fld, cgpe}$ and \hat{V}^{ens} is due to a combination of factors including the fact that variations in the ATFE are partly due to specific attributes of MALAPS and the ensemble spread of MSEPS cannot characterize that component of the uncertainty by design. \hat{V}^{ens} does however provide a good benchmark comparison.

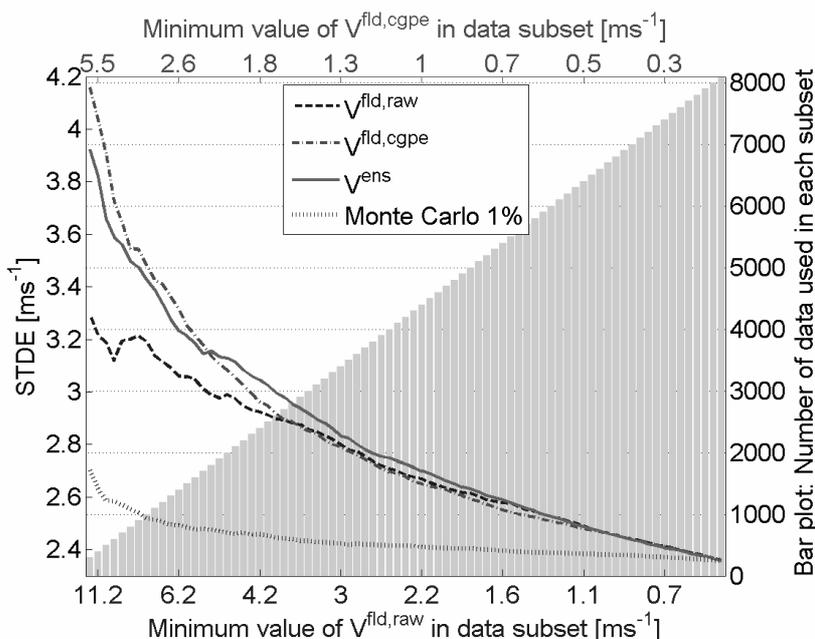


Figure 2: Cumulative STDE at Yambuk. The data is sorted from largest to smallest values of each predictor and a cumulative selection of the data is taken to calculate STDE. The grey bars show the amount of data used for each value on the curves. The x-axes show the minimum values of $V^{fld,raw}$ or $V^{fld,cgpe}$ of the relevant subset of the data for a few cases

Figure 3 shows the POD, FAR and ORSS results for Yambuk using the predictors to categorically forecast ATFE to be above a threshold. For example, an ATFE of magnitude greater than 5 ms^{-1} occurs 367 times, corresponding to a frequency of 4.5%. The plot shows that $\hat{V}^{fld, cgpe}$ has marginally the best results for predicting the occurrence of this event and predicts 41% of the events correctly (POD), 86% of the predictions made are false alarms (FAR) and the ORSS is 0.67. These are considered reasonable results for what is a difficult prediction to make. One of the reasons the FAR is quite large is that the values of the predictors used are related to the probability of large values of ATFE rather than a direct prediction of a large ATFE. There will be some times where the predictor value is large but ATFE is small because the forecast was “lucky”. The plots show that all of the variability predictors outperform the Monte Carlo 1% level indicating significant forecast skill (hypothesis 1.c). Although the POD decreases and the FAR increases with rising threshold, the ORSS increases as well indicating that the skill level is improving with the more extreme, rarer event (hypothesis 1.e). As with figure 2, these plots also show an improvement in performance using $\hat{V}^{fld, cgpe}$ compared with $\hat{V}^{fld, raw}$ (hypothesis 2.a), and the level of improvement increases with rising threshold (hypothesis 2.b). The performance of $\hat{V}^{fld, cgpe}$ is also comparable with $\hat{V}^{fld, ens}$ (hypothesis 1.d). All these points are consistent with the hypothesis that misplacement provides a significant contribution to large single time-series forecast errors (hypothesis 1).

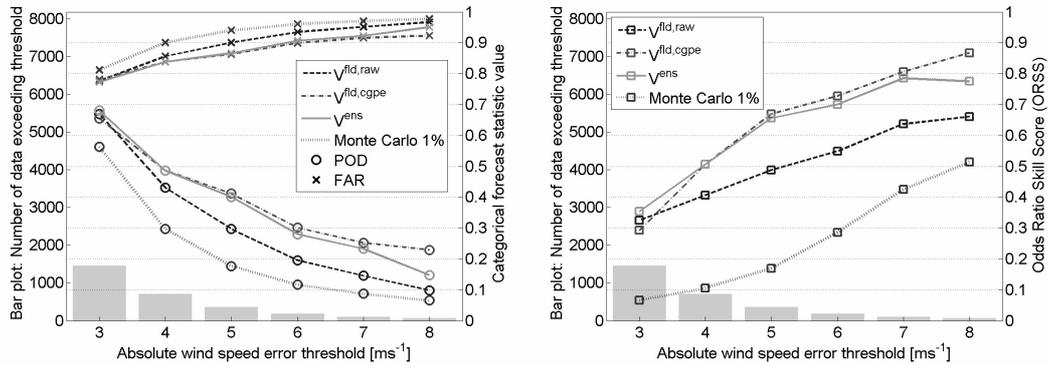


Figure 3: POD and FAR (left) and ORSS (right) for the three predictors and the Monte Carlo 1% level for predicting ATFE to exceed various thresholds at Yambuk wind farm

The cumulative STDE plot for Chalicum Hills is shown in figure 4. The comparison of the predictors for Chalicum Hills with the Monte Carlo 1% level (hypothesis 1.c) and the comparability between $\hat{v}^{fld, cgpe}$ and $\hat{v}^{fld, ens}$ (hypothesis 1.d) are as convincing as for Yambuk. While $\hat{v}^{fld, cgpe}$ still provides an improvement over $\hat{v}^{fld, raw}$ (hypothesis 2.b) the performance of $\hat{v}^{fld, raw}$ is much better than at Yambuk. This might be due to the fact that the domain used for CH is almost all land thereby giving weaker wind speeds overall and hence less spatial variability due to terrain than for Yambuk where half of the domain is ocean. Figure 4 shows some evidence for the spatial variability being less for the CH field, since the largest $\hat{v}^{fld, raw}$ values shown on the x-axis for CH is a relatively small value of 6.7 ms^{-1} compared with 11.2 ms^{-1} for YB.

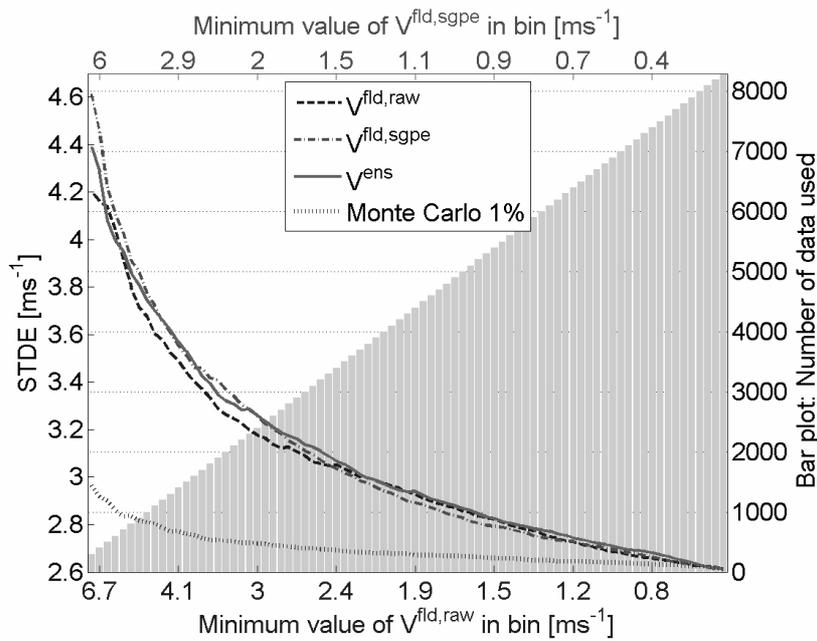


Figure 4: Cumulative STDE for Chalicum Hills. Data sorted as in figure 2

The POD, FAR and ORSS results are shown for CH in figure 5. The results show that similar performances to YB are achieved at CH with some consistency with the difference in results observed in the cumulative STDE plot. For CH, the ORSS with $\hat{V}^{fld, cgpe}$ for the 5 ms⁻¹ threshold is a bit less than for YB at 0.58.

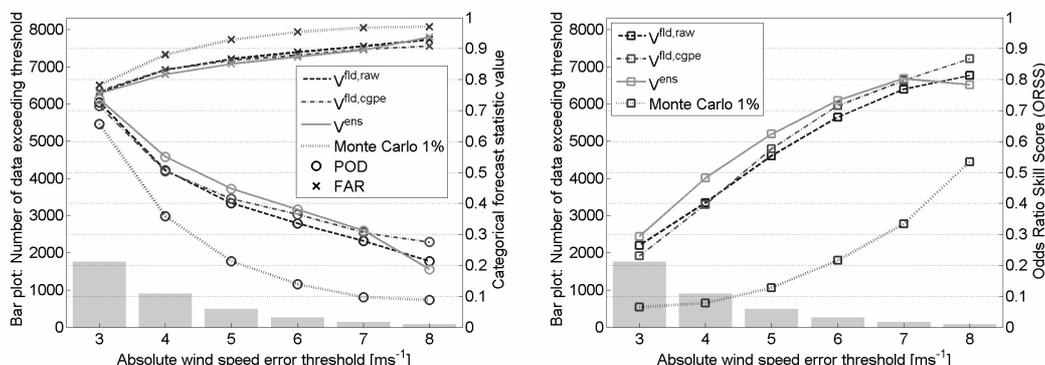


Figure 5: As in figure 3 except for Challicum Hills

Along with forecasting periods of large uncertainty, we can also study the ability of the predictors to distinguish periods when the single time-series forecasts will be more accurate than usual. Figure 7 shows the cumulative STDE plots for Yambuk and Challicum Hills equivalent to figure 2 except that the data is sorted on the predictors from smallest to largest. The figure shows how the three predictors compare in distinguishing the smallest single time-series forecast errors from MALAPS. As with large errors, all results are significantly better than the Monte Carlo 1% level (hypothesis 1.c). The values for $\hat{V}^{fld, cgpe}$ are comparable with \hat{V}^{ens} at YB (hypothesis 1.d) but the results for $\hat{V}^{fld, cgpe}$ are not so good at CH. The performance of $\hat{V}^{fld, cgpe}$ is similar or a little worse than $\hat{V}^{fld, raw}$ at YB and in the case of CH $\hat{V}^{fld, cgpe}$ performs considerably worse in distinguishing small errors than $\hat{V}^{fld, raw}$ (however in this case using RMSE instead of STDE showed the opposite result). The ORSS values for all sites (plots not shown) are around 0.1 to 0.2 and do not increase with rarer events (smaller error thresholds). The POD and FAR results are close to identical for all predictors at all sites. These ordinary results give some further evidence for hypothesis 2.b.

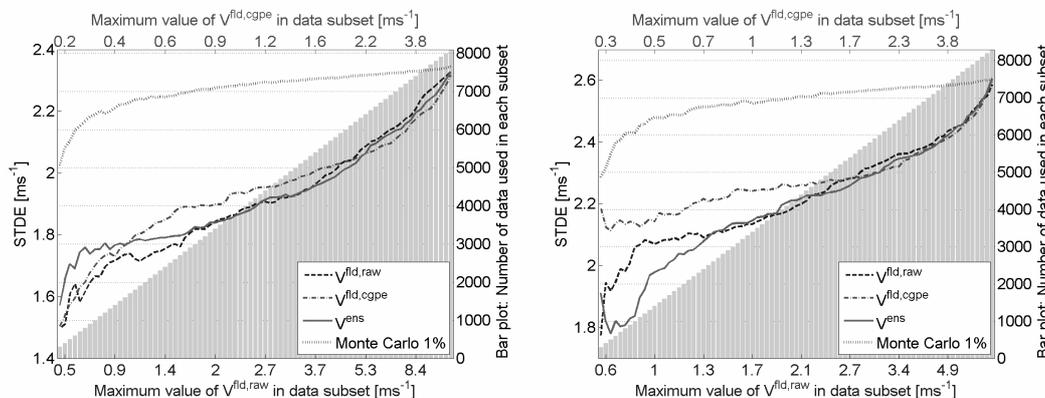


Figure 6: Cumulative STDE plot for Yambuk (left) and Challicum Hills (right) as in figure 2 except sorting on the predictors from smallest to largest and the x-axes show maximum values

5.2. Temporal variability (V^{obs})

In this section, the performances of the various predictors are compared with respect to their ability to forecast occurrences of extreme values in temporal variability, V^{obs} , as described by equation (5) in section 3. Unlike ATFE, which is based on the MALAPS single time-series forecast, V^{obs} is based solely on observations, giving MSEPS and MALAPS equal conditions on which to compare results. However, the field variability and ensemble spread techniques still cannot be directly compared since the differences in the results are also due to differences between the NWP systems. As in the previous section, Kendall's rank correlation is used to compare the predictors along with the cumulative plot and categorical forecast scores to study the ability of the predictors to distinguish very large and very small values for V^{obs} .

Kendall's ranked correlations between the predictors and V^{obs} are shown in table 3. As with ATFE, all predictors give correlations significantly larger than the Monte Carlo 1% values (hypotheses 1.a and 1.b). The correlations are larger for most of the predictors than they were for ATFE indicating a stronger degree of correspondence. In this case, $\hat{V}^{fld, cgpe}$ gives a smaller value than $\hat{V}^{fld, raw}$ for YB and a larger value for CH, which is opposite to ATFE. However, as for ATFE, the values are close and hence the change may not be significant. For V^{obs} , \hat{V}^{ens} gives the strongest correlation for both sites and is more than double the $\hat{V}^{fld, cgpe}$ value at YB.

Table 3: Kendall's correlation coefficients between predictors and V^{obs} at each site

Predictor	YB	CH
$\hat{V}^{fld, raw}$	0.092	0.119
$\hat{V}^{fld, cgpe}$	0.077	0.130
\hat{V}^{ens}	0.198	0.181
Monte Carlo 1% test	0.017	0.017

The cumulative V^{obs} plots for YB and CH are shown in figure 7. As with the correlations, all results are considerably better than the Monte Carlo 1% level. The plots show that $\hat{V}^{fld, cgpe}$ provides an improvement over $\hat{V}^{fld, raw}$ in distinguishing large values of V^{obs} at both sites (hypothesis 2.b). However, while $\hat{V}^{fld, cgpe}$ is comparable with \hat{V}^{ens} at CH, \hat{V}^{ens} performs considerably better than $\hat{V}^{fld, cgpe}$ at YB. Similar results are discernible in the POD, FAR and ORSS scores for predicting V^{obs} to be above different thresholds as shown in figure 8 for YB and figure 9 for CH. All the scores are much better than the Monte Carlo 1% level but are not quite as good as for predicting large values of ATFE. One exception to this is that \hat{V}^{ens} achieves similar ORSS scores (around 0.7) at YB for each measure of forecast uncertainty (compare figure 8 and figure 3).

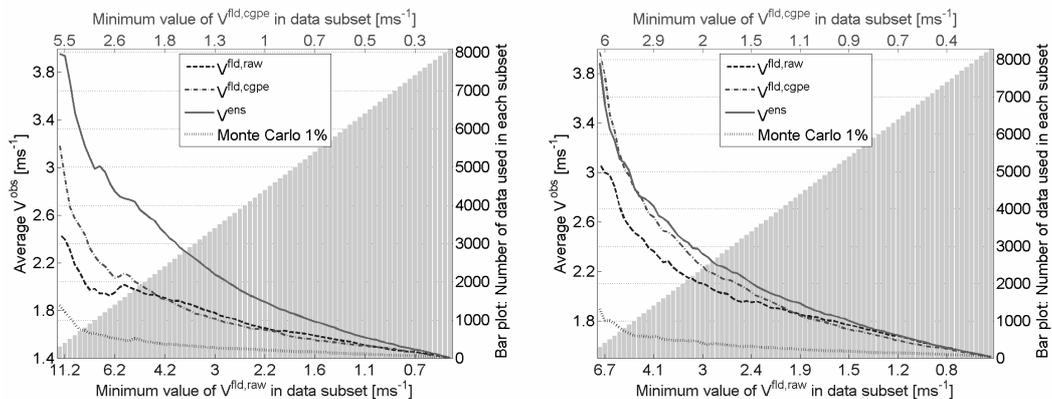


Figure 7: Cumulative V^{obs} plot for Yambuk (left) and Challicum Hills (right) as in figure 2 except the temporal variability V^{obs} , is calculated for each bin instead of V^{fld}

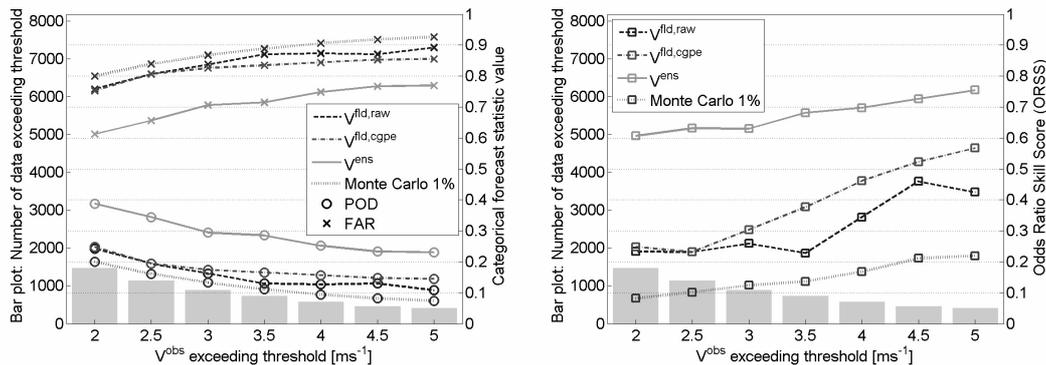


Figure 8: POD and FAR (left) and ORSS (right) for the three predictors and the Monte Carlo 1% level for predicting V^{obs} to exceed various thresholds at Yambuk wind farm

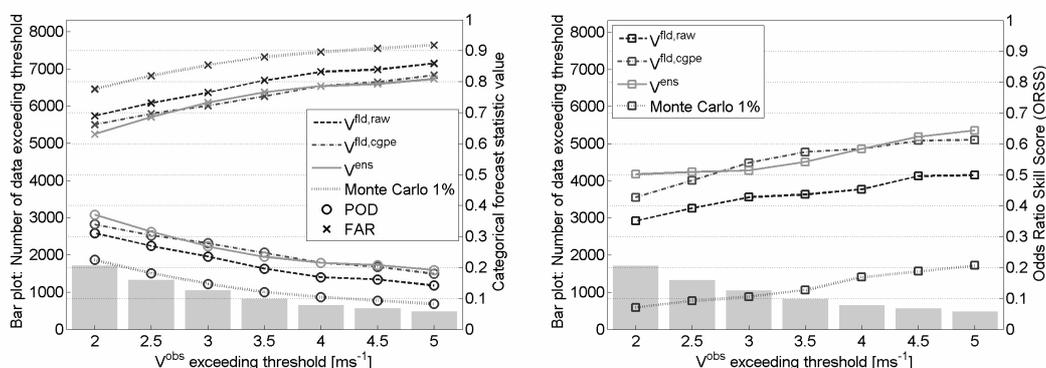


Figure 9: As in figure 8, except for Chalicum Hills

Figure 10 shows the cumulative V^{obs} plots for YB and CH after sorting on the predictors from smallest to largest. As with the equivalent plots for ATFE (figure 6), these show less consistency in the difference between $\hat{V}^{flid,cgpe}$ and $\hat{V}^{flid,raw}$. \hat{V}^{ens} gives a much better performance than the other predictors at YB but they are more similar at CH. However, \hat{V}^{ens} gives better scores for POD, FAR and ORSS (not shown) at both sites. The ORSS scores for \hat{V}^{ens} are mostly less than 0.4, which is not as good as for predicting large values V^{obs} and is a similar result as for predicting ATFE.

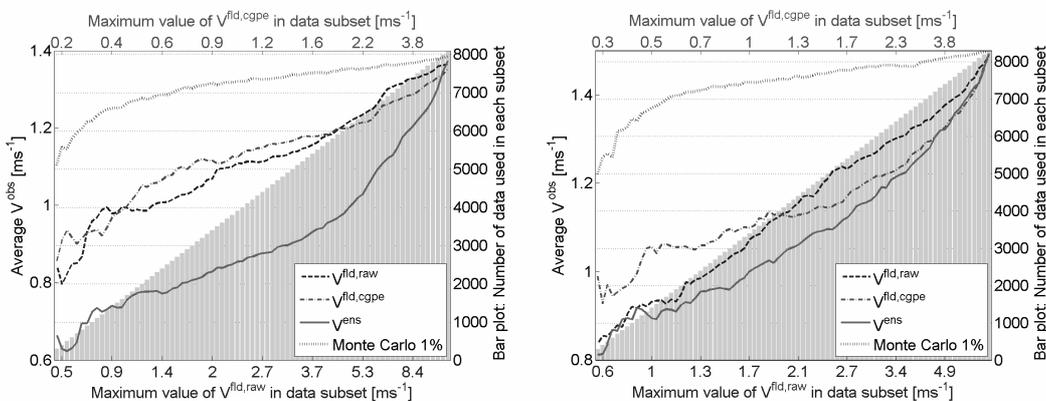


Figure 10: Cumulative V^{obs} plots at Yambuk (left) and Chalicum Hills (right) for after sorting on the predictors from smallest to largest

The above results might have been examined in other ways. One option would be to use a power curve to convert all the forecast wind speeds to wind power for each wind farm and compare the wind power forecasts with observations of wind power. This method was trialed and achieved similar results and conclusions to using wind speed. Wind speed was chosen for the results in this paper since it removes one source of uncertainty for evaluating the new techniques.

We have also tested alternative measures of the field variability (and the ensemble spread). These included using kurtosis instead of the variance and converting the wind speeds to a form of wind energy flux (related to the cube of the wind speed) before calculating the variance. The results using these methods gave smaller categorical forecast scores than the results shown in this paper.

[15] presents a more comprehensive study of the work in this paper and includes results for four sites instead of two.

6. CONCLUSIONS

This paper has described a theory and methodology for using the variability in forecast wind speeds at multiple grid points in Numerical Weather Prediction (NWP) systems to predict the uncertainty in future wind power generation. The theory is based on the concept that a major source of uncertainty in NWP forecasts arises from potential misplacement of weather features relative to the physical world. In essence, the method produces an ensemble of forecasts from only one run of one NWP system, and therefore has a small computational burden compared to conventional ensemble approaches. The effectiveness of this new approach has been assessed relative to a standard NWP ensemble and Monte Carlo significance tests.

The results in this paper show that the spatial variability of wind speeds in an NWP system can give a useful characterization of uncertainty in future wind power. A weighted variance of chosen-grid-point-equivalent (cgp-equivalent) wind speed forecasts from multiple grid points in an NWP system ($\hat{V}^{fld,cgpe}$) has been shown to have a significant degree of correspondence with forecast errors in single time-series forecasts at a location of interest. This correspondence is comparable to that achievable with an NWP ensemble spread. In particular, $\hat{V}^{fld,cgpe}$ provides a good probabilistic forecast for large forecast errors, again comparable to an NWP ensemble spread. Similarly, $\hat{V}^{fld,cgpe}$ and the ensemble spread were shown to have a significant degree of correspondence with observed temporal variability, although in this case the ensemble spread appeared to significantly outperform $\hat{V}^{fld,cgpe}$ at one site. Both $\hat{V}^{fld,cgpe}$ and the spread from an NWP ensemble spread may have some skill in providing probabilistic forecasts of small errors in single time-series forecasts or when there are small amounts of temporal variability, but the performance is not as good as for forecasting large values of these two quantities. All these results provide plausible evidence of a relationship between misplacement of weather features and forecast uncertainty from an NWP system.

In addition to the above, by comparing the spatial variability in raw and cgp-equivalent NWP fields, this work has shown that the terrain standardization method appears to perform well in removing spatial variability in wind speed forecasts due to the terrain. As such it can enhance the value of assessing wind speeds at grid points around the location of interest in order to better characterise the uncertainty of the forecast at this location.

A useful future study would be to compare the spatial variability and ensemble variability using the same NWP system to allow a more direct comparison. Other further work could be to use the difference between the cgp-equivalent wind speed forecasts at each grid point and

the chosen grid point forecast to predict the sign (as well as likely magnitude) of the single time-series forecast error, if this information would be useful for forecast users. Also, since NWP ensembles and the multiple grid point approach characterize forecast uncertainty from two different perspectives, the results in this paper suggest the possibility of combining these two approaches to make a more superior characterization of the uncertainty in future wind power than from using either method alone.

ACKNOWLEDGEMENTS

This research is partly supported through a grant from the Australian Greenhouse Office (AGO). The Bureau of Meteorology Research Centre (BMRC) provided the NWP data for MALAPS and collaborative support. Pacific Hydro Pty Ltd. provided the observation data for the Yambuk and Chalicum Hills wind farms. Roaring40s Pty Ltd. provided the observation data for Bluff Point. Weprog ApS provided the NWP data for MSEPS and some useful discussions.

REFERENCES

1. Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A. and Nielsen, T. S., Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models, *Wind Engineering*, 2005, 29(6), 475-489.
2. Ernst, B., Oakleaf, B., Ahlstrom, M. L., Lange, M., Möhrlein, C., Lange, B., Focken, U. and Rohrig, K., Predicting the Wind, *IEEE power & energy magazine*, 2007, 5(6), 78-89.
3. AWS Truewind, *AWS Truewind's Final Report for the Alberta Forecasting Pilot Project*, AWS Truewind, 2008. Available at: <http://www.aeso.ca/gridoperations/13825.html>
4. Lange M., and Focken, U., *Physical Approach to Short-Term Wind Power Prediction*. Oldenburg, Springer, 2006.
5. Focken U. and Lange, M., *Final Report - Wind power forecasting pilot project in Alberta Canada*, Energy & Meteo Systems, Oldenburg, Germany, 2008. Available at: <http://www.aeso.ca/gridoperations/13825.html>
6. Jørgensen J. and Möhrlein, C., *AESO Wind Power Forecasting PILOT Project - Final Project Report*, Weprog, 2008. Available at: <http://www.aeso.ca/gridoperations/13825.html>
7. Nielsen, H. A., Madsen, H., Nielsen, T. S., Badger, J., Giebel, G., Landberg, L., Sattler, K., and Feddersen, H., *Wind Power Ensemble Forecasting Using Wind Speed and Direction Ensembles from ECMWF or NCEP*, Risø National Laboratory, Roskilde, Denmark, 2005.
8. Giebel, G., Badger, J., Landberg, L., Nielsen, H. A., Nielsen, T. S., Madsen, H., Sattler, K., Henrik, F., Vedel, H., Tøfting, J., Kruse, L. and Voulund, L., *Wind Power Prediction using Ensembles*, Risø National Laboratory, Roskilde, Denmark Risø-R-1527(EN), 2005.
9. Möhrlein, C., *Uncertainty in Wind Energy Forecasting*, PhD thesis, Civil and Environmental Engineering, University College Cork, Cork. 2004.
10. Focken, U. and Lange, M., Optimal Combination of Different Numerical Weather Predictions for Improved Wind Power Predictions (abstract only), in *European Wind Energy Conference*, Athens, Greece, 2006. Available on CD Rom.
11. Pinson, P., *Estimation of the Uncertainty in Wind Power Forecasting*, PhD thesis, Ecole des Mines de Paris, Paris, France, 2006.

12. Lange, M., On the Uncertainty of Wind Power Predictions—Analysis of the Forecast Accuracy and Statistical Distribution of Errors, *Journal of Solar Energy Engineering*, 2005, 127(2), 177-184.
13. Cutler, N., Kepert, J., Outhred, H., MacGill, I. and Kay, M., Characterising Future Large Changes in Aggregated Wind Power using Numerical Weather Prediction Spatial Fields, available online with *Wind Energy*, 2008.
14. Antoniou I. and Pedersen, T. F., *Nacelle Anemometry on a 1MW Wind Turbine: Comparing the power performance results by use of the nacelle or mast anemometer*, Risø National Laboratory, Roskilde, Denmark Risø-R-941(EN), 1997.
15. Cutler, N., *Characterising the Uncertainty in Potential Large Rapid Changes in Future Wind Power Generation*, PhD thesis, School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney. Forthcoming PhD thesis to be published in early 2009.
16. Vincent, C., Bourke, B., Kepert, J., Chattopadhyay, M., Ma, Y., Steinle, P. and Tingwell, C., Verification of a High-Resolution Mesoscale NWP System, in press with the *Australian Meteorological Magazine*, 2008. Available at:
http://www.bom.gov.au/bmrc/daas/da_staff/kepert_papers.htm
17. National Electricity Market Management Company Limited, *An introduction to Australia's National Electricity Market*, Available at:
<http://www.nemmco.com.au/nemgeneral/000-0187.pdf>, 2004.
18. Weprog, *Homepage of the MSEPS Project*, Access date: 2 December 2008.
<http://www.mseps.net/>
19. Stern, H., *Statistically Based Weather Forecast Guidance*, PhD thesis, Melbourne University, 1999.
20. Stephenson, D. B., Use of the “Odds Ratio” for Diagnosing Forecast Skill, *Weather and Forecasting*, 2000, 15(2), 221-232.
21. Gritmit E. P. and Potter, C., A Prototype Day-Ahead Forecast System for Rapid Wind Ramp Events, in *Windpower 2008 Conference and Exhibition*. Houston, Texas, USA, 2008. Available on CD Rom.
22. Diggle, P. J., *Statistical Analysis of Spatial Point Patterns*, 2nd ed: Arnold, 2003.
23. Abdi, H., *The Kendall Rank Correlation Coefficient*. Thousand Oaks (CA), Sage, 2007.
24. Bradley, C., 68.32 Kendall's Correlation Coefficient Revisited, *The Mathematical Gazette*, 1984, 68(445), 214-215.