Wind forecasting using Principal Component Analysis

Skittides, Christina; Fruh, Wolf-Gerrit

Published in:
Renewable Energy

DOI:
10.1016/j.renene.2014.03.068

Publication date:
2014

Wind Forecasting using Principal Component Analysis

Christina Skittides and Wolf-Gerrit Früh

School of Engineering and Physical Sciences,
Heriot-Watt University, Edinburgh EH14 4AS, UK
tel: +44-131-451 3593; fax: +44-131-451 3129

Corresponding author: Christina Skittides, cs261@hw.ac.uk ,tel: +44-131-451 3593; fax: +44-131-451 3129

Abstract

We present a new statistical wind forecasting tool based on Principal Component Analysis (PCA), which is trained on past data to predict the wind speed using an ensemble of dynamically similar past events. At the same time the method provides a prediction of the likely forecasting error. The method is applied to Meteorological Office wind speed and direction data from a site in Edinburgh. For the training period, the years 2008–2009 were used, and the wind forecasting was tested for the data from 2010 for that site. Different parameter values were also used in the PCA analysis to explore the sensitivity analysis of the results.

The forecasting results demonstrated that the technique can be used to forecast the wind up to 24 hours ahead with a consistent improvement over persistence for forecasting more than 10 hours ahead. The comparison of the forecasting error with the uncertainty estimated from the error growth in the ensemble forecast showed that the forecasting error could be well predicted.

Key words

Wind energy resource, Principal Component Analysis (PCA), forecasting

1. Introduction

Wind energy is one of the most established renewable energy forms. It has been the world's fastest renewable energy resource in growth for the past 7 years [1]. Wind energy has also the characteristic of a strongly variable form of energy. To achieve a high level of performance, a good quality of wind speed or generation forecasting is vital. Wind speed and direction are the most important factors that determine the power output and they can vary at all time scales. Different cycles with time scales ranging from daily to seasonal and interannual can be observed in addition to turbulence and gusts. For example, for mean daily or hourly wind speed forecasts, the underlying atmospheric dynamics become of great importance [2]. In addition, the turbines have to adjust to the wind fluctuations at all time but often have a delay in their response. Hence, the methods of analysis and prediction of wind behaviour are indeed of extreme importance for a good operation of wind turbines and wind farms.

1.1 Forecasting methods

Because the wind variability can be characterised by slow cycles (daily and longer), fast (unpredictable) turbulence, and synoptic weather changes which tend to change only slowly, the forecasting horizon can be divided into the three following categories: 1: immediate-short-term (up to 8 hours ahead), 2: short-term (8 to 24 hours ahead), and 3: long-term (multiple-days-ahead) forecasting [3,4,5]. It is more common to use hourly forecasts in order to determine daily forecasts of hourly winds [6].

Several forecast models have been created which can be categorised into physical, such as the Numerical Weather Prediction systems (NWPs) [3], statistical, including linear methods such as Auto Regressive Moving Average models (ARMA) or methods coming from artificial intelligence and machine learning fields such as Artificial Neural Networks (ANNs), or even by hybrid approach methods which are a combination of statistical and physical methods with a use of weather forecasts and analysis of time series [4]. Erdem and Shi [7] used four ARMA approaches in order to obtain wind speed and direction forecasts and found that the ARMA model based on the decomposition of wind speed into lateral and longitudinal components was better in predicting direction in comparison to the traditional ARMA model. However, that was the opposite case for wind speed. De Giorgi et al. [8] used ARMA models in combination with different types of ANNs and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) for several testing period models but also time horizons. For all the attempts it was found that the forecast was progressively worse as the prediction length was increasing.

An integration of ANNs with NWPs for forecasting purposes was undertaken again by De Giorgi et al. [9]**.** The neural network was initially based on the statistic model of wind power time series and was later integrated with NWPs which indicated a significant improvement on the performance. Specifically, pressure and temperature as NWP parameters seemed to improve the forecasting model. Früh [10] explored a simple a linear predictor and based on the observed mean daily cycle model with wind speed or power output data as inputs and noted that increased sophistication in the forecasting methods surprisingly seemed to deteriorate the predictive ability.

Hybrid approaches typically employ an ARIMA model for the linear characteristics and an ANN or SVM (Support Vector Machine) model for the nonlinear characteristics. Wang et al. [5] found that depending on forecasting horizon, hybrid methods or ARIMA method perform better in forecasting than the ANN and SVM methods. They also concluded that hybrid methods add significantly in the short-term forecasting modelling for wind speed and power generation, but in general, they do not outperform the other methods [11].

1.2 Principal Component Analysis of Time Series

Underlying all statistical and empirical approaches is the need to separate the predictable component from the turbulent component in an effective and efficient manner. For example, for mean daily or hourly wind speed forecasts, i.e., short-term horizons, the underlying atmospheric dynamics become of great importance [12]. The wind related data could be treated as dynamical systems so that cycles and random unusual behaviours that often characterise them can be identified, explained and understood. Based on this understanding, we propose to use a time series analysis technique based on the dynamical systems theory which was devised to separate coherent dynamical information from noisy experimental data, known as Singular Systems Analysis [13,14], which is effectively the standard Principal Component Analysis (PCA) [15] from Statistics applied to a suitably formatted time series. It is also known as Empirical Orthogonal Function (EOF) Analysis in the Meteorological and Oceanographic community to identify the main circulation patterns in the Atmosphere and oceans, e.g., [16,17]. This technique is now widely used for time series analysis of nonlinear dynamical systems in general, e.g. [18,19] as the analysis is very powerful to separate coherent dynamics from noise.

The principle in terms of a dynamical system is that the dynamic evolution of the system takes place on a time-invariant object, called ‘attractor’, after initial transients have decayed. This attractor is a geometric object in the phase space defined by the dynamic variables of the dynamical system. In the example of a harmonic oscillator, the phase space is defined by the position and momentum of the oscillating object, and the motion of it takes place on a limit cycle. This cycle is the attractor, and the trace drawn by the oscillation, or its ‘orbit’, would draw repeating copies of that cycle over and over.

In complex systems, where the phase space is not fully accessible from measurements, one can use Takens' method of delays [20] to create a space equivalent to the phase space but this phase space reconstruction cannot separate the important dynamics from measurement noise or turbulence. Applying PCA to the set of delay time series is a method to redefine the phase space to concentrate the coherent information in a few directions (or dimensions) of the phase space, which then allows to 'delete' the weaker and uncorrelated dynamics from the description of the system. The creation of this system based on a training set of wind data defines the model for the forecasting. New measurements can then be mapped onto the cleaned-up attractor to find previous measurements which are, in dynamical terms, similar to the current measurements. Finding one or more 'similar' previous measurements, then allows us to the evolution of those measurements as equivalent to predicting the current measurements. In addition to a prediction, however, this method predicts a number of similar events and following how their distances change over the lead time of the prediction also provides a measure of how sensitive the system is to uncertainties in measurements or out-of-system perturbations. Hence, it provides a measure of the uncertainty of the prediction at the same time.

1.3 Aims and Outline

The aim of this paper is to develop a wind speed forecasting tool which, by being based on PCA, provides a forecast based on the slow dynamics of the atmosphere alone and also provides an intrinsic measure of the quality of each forecast.

To develop the tool, we will in Section 2 first introduce the formalism of PCA applied to a time series of wind speed and direction, and then the forecasting method. Section 3, introduces the main data set, the parameter settings, and the error measures used to develop and evaluate the approach. The results of this analysis are presented in Section 4.

2. Principal Component Analysis for forecasting

This section contains background information regarding phase space reconstruction as well as PCA and explains in detail how they will be used for the forecasting purposes. The stages for the training of the predictor are preparation of the phase space using the training set of data (e.g., wind speed and direction), Principal Component Analysis of the phase space to optimise the phase space and truncation of the phase space to the relevant components only to define the predictor.

The application of the predictor goes through the preparation of the test data to the same specifications as the training set, mapping the test data onto the truncated phase space, finding an ensemble of nearest neighbours on the attractor as defined by the test data, tracing the evolution of that ensemble for the lead period of the prediction, and finally re-transforming the ensemble of predictions into the original variables (e.g., wind speed and direction). A summary of the forecasting algorithm is presented in Table 1, and the remainder of this section will describe each of these steps in turn.

Table 1. The forecasting algorithm

Training:	
1) Normalise measurements	
2) Create time-delay matrix; eq.(1)	$Y^{i,j+(jo-1)M_w} = y_{j_o}\left[j+\left[i-1 \right]\tau \right],$
3) Perform PCA to optimise; eq. (2)	$Y = P\Lambda S$
4) Truncate to the relevant components to define predictor; eq.(3)	$Y_r = P_r\Lambda_r S_r$
Forecasting:	
5) Normalise new measurements using Training normalisation	
6) Create time-delay matrix using same parameters as for Training	
7) Map time-delay matrix onto attractor coordinates; eq.(5)	$P_n = Y_n S_r^T \Lambda_r^{-1}$
8) Find number of similar events in training period and follow evolution of past events i.e. nearest neighbours; eq.(6)	$d_i = \frac{1}{n_x}\sum_j \left\| P_n^j - P_r^{i+j-1} \right\|$
9) Find distance vector due to n. neighbours; eq.(7)	$D_j = P_r^{k_j} - P_n^{n_x}$
10) Use ensemble prediction based on n.neighbours; eq.(8)	$P_f^j\left[T \right] = P_r^{k_j+T} + D_j$
11) Map back to delay matrix and return predicted wind speed; eq.(9)	$Y_f^j = P_f\Lambda_r S_r$
12) Re-scale back to proper units	

2.1 Phase space preparation

Furthermore, a method is needed so as to define equivalent variables to the phase space ones which is the *time-delay* method [19]. It is a practical implementation of the dynamical systems since it aids in reconstructing the phase space of a dynamical system from an observed deterministic time series. The reconstruction of a phase space is indeed significant since it can extract useful information about the time series that characterise the system. Using previous measurements is equivalent but not practical with data containing noise or turbulence [20]. The challenge that arises then is that if we have measurements from only one site, can we use similar analysis concepts to identify the state of a combined system on the phase space? If so, can we then predict for the second site, for which we have no data obtained? More precisely the question that arises is by taking the defined points of the combined two sites system and adding the new measurements can we project them to the existing attractor and predict from the nearby points?

If the training data set consists of N_O variables, $y_{jo}(t)$, for example wind speed and wind direction with N_O= 2, covering N_t time steps, the first step is to rescale them in such a manner that they both contribute equally to the analysis. This is achieved by rescaling them both to

time series of zero mean and unit variance, i.e., subtracting the mean from each variable in turn and then dividing by the variance.

The phase-space equivalent variables can then be constructed using Takens' Method of Delays [19], which postulates that the dynamic variables not directly measured have influenced the evolution of the measured variables and are therefore somehow represented by the previous measurements. Thereby, a sufficient representation of the state complete phase space at time t is given by the delay vector $(y_1(t), y_1(t-\tau), y(t-2\tau), \ldots, y_1(t-M_w\tau)$, where M_w is the number of time lags, τ, used, and the same can be done for further variables measured, e.g., $y_2(t)$.

With a time series of N_o variables of length N_t, the delay matrix will have $N = N_t - M_w\tau$ rows and $M = N_o M_w$ columns with

$$Y^{i,j+(jo-1)M_w} = y_{j_o}(j + (i-1)\tau), \quad (1)$$

with the row index $i = 1 \ldots N$, the column index $j = 1 \ldots M$, and the observable index, $j_o = 1 \ldots N_o$ [20]. In this matrix, a row m is equivalent to a complete phase-space description of the system at time t_m as long as M is sufficiently large.

2.2 Principal Component Analysis

Since the time-delay method is sensitive in the choice of the parameters, Principal Component Analysis (PCA) is used to optimise the phase space reconstruction. It is a non-parametric statistical method and by that is not limited to be of a certain distribution or linear relationship. PCA can separate noise from useful information applied to time-delay series [19]**.** It can identify the number of needed time-delays and give a picture of their shape. Its goal is to explain important variability of the time series data and to extract useful information (i.e. hidden structures of the data) from its more relevant components in a reduced number of dimensions.

The mathematical procedure to carry out a PCA is through the Singular Value Decomposition (SVD) of the delay matrix. In terms of the linear algebra of the SVD, it is a transformation of the basis vectors of the phase space which finds orthonormal basis vectors to maximise the variance described by as few basis vectors as possible. The three SVD/PCA outputs are the *singular vectors* which are the basis vector for each dimension (they are also the eigenvectors of the covariance matrix of *Y*), the *singular values* which measure the time-averaged contribution of each dimension to the total variance**,** and the *principal components* (pc's) which form an attractor and describe the system's time series. In matrix notation, the Singular Value Decomposition is written as

$$Y = P\Lambda S \quad (2)$$

where $Y(n,m)$ is the *time-delay* matrix with $n = 1 \ldots N$ the time point within time series and $m = 1 \ldots M$ the index of the dimension. $P(n,m)$ is the principal component matrix, $\Lambda(m,m)$ is the diagonal matrix of singular values, and $S(m,m)$ contains the singular vectors.

The singular values represent a measure of the variance, more specifically the square root of the variance of the time series in the corresponding dimensions and they can pick out the important variability of the data. The singular vectors have the property of being orthonormal, i.e. orthogonal and of unit length and they span the dimensions of the phase space. They represent a measure of those dimensions that define a dynamical system, for instance they can replace position and momentum, two variables which can form a dynamical system. The principal components are the time series of the system in the coordinate system defined by the singular vectors. This means that plotting the principal components against each other draws the orbit of the measurements and thereby provides an estimate of the underlying attractor.

2.3 The Forecasting Model

The singular values are a key measure on which the determination of the best predictor is based, since our initial assumption was that the wind conditions several hours ahead is better predicted by the slower atmospheric dynamics than the short-time fluctuations. The PCA has separated the coherent (slower) dynamics from the temporally uncorrelated short-term fluctuations, such that uncorrelated fluctuations are visible as a noise floor in the singular value spectrum. Persistent variance from the atmospheric dynamics is concentrated in the leading singular values of much higher magnitude. For that reason, the phase space can now be truncated to a much smaller dimension than the original delay matrix

By creating a reduced set of M_r principal components, $P_r^{N,Mr}$, singular values, $\Lambda_r^{Mr,Mr}$, and singular vectors, $S_r^{Mr,M}$, one can produce a filtered time series of the original data by

$$Y_r = P_r \Lambda_r S_r \tag{3}$$

There, the filtered time series of the first observable, y_1, is contained in the first column of Y_r, the filtered time series of the second observable in column M_w+1, and so on. However, due to the method of delays, those columns only cover the time steps M_w to M and one has to append the bottom row to the end of that variable:

$$y_{Jo}\ t_1, t_2, \dots, t_{Nt} \ = \ Y_r\ 1 \dots N, 1 + \ j_o - 1\ M_w\ , Y_r\ N,\ 2 \dots M_w \ + \ j_o - 1\ M_w \tag{4}$$

The forecasting model therefore consists of the truncated dynamical system P_r, Λ_r, , S_r and the principle is to interpolate the current measurements to 'close' examples of the filtered training data, where 'close' is in terms of dynamic behaviour rather than time.

2.4 Preparing new data for the forecasting model

It is possible to project a new time series onto this reduced set of singular vectors by creating a delay matrix following the same procedure as for the training set, including using the mean and standard deviation from the training data set to rescale the new data. This projection will then give principal components, P_n, to place the new data in this phase space as

$$P_n = Y_n S_r^T \Lambda_r^{-1} \tag{5}$$

To generate a single point in this phase space, the new time series must contain $M_w\tau$ measurements. Conversely, if the new time series contains $M_w\tau + n_x - 1$ points, its time delay matrix contains n_x columns for that observable and its projection onto the singular vectors results in a section of orbit containing n_x points.

2.5 Finding nearest neighbours

Ensemble forecasting in dynamical forecasting making several forecasts, each initialised with a slightly different initial condition but within the measurement accuracy of the initial point to predict a large sample of possible future outcomes. The results are then evaluated by examining the distribution across all ensemble members of the forecast variables. A useful feature of ensemble forecasting is that it also provides an estimation of the reliability of the forecast. The idea is that when the different ensemble members differ widely, the actual event we try to forecast could shadow any of the modelled ensemble members. This then means that the forecast is affected by a large uncertainty; when there is a closer agreement between the ensemble member forecasts, the uncertainty in the prediction is lower [12]. This principle can also be applied to PCA forecasting where the attractor represents the model. Now, current measurements can be mapped onto the attractor and previously observed wind states close to the current measurements can be found. They can then be taken as an ensemble of initial conditions close to the current state and thus be used for prediction.

The two key stages in the forecasting part of the method are, firstly, to find a number of 'similar' events in the training period, which is done by finding a chosen number of nearest neighbours in the attractor and, secondly, to follow the evolution of those past similar events. From that evolution one can calculate an expected mean evolution which is the prediction, and one can also calculate by how much the evolution of the ensemble of similar past events either stayed close (giving confidence in the mean forecast) or diverged over the forecasting horizon (indicating that the currently measured wind comes from a part of the attractor which is unstable and not well predictable).

The nearest neighbours are found by calculating the Euclidean distance between the new point, or the mean distance of each point of the section of orbit, to all other points or sections of the training attractor; for a single point: $d_i = \left| P_n - P_r^i \right|$ or for a section of orbit with n_x points

$$d_i = \frac{1}{n_x} \sum_j \left| P_n^j - P_r^{i+j-1} \right| \qquad (6)$$

From this complete set of distances to all points of the training attractor, a specified number of nearest neighbours is selected, subject to a constraint that they do not come from adjacent points on the training orbit but from different passes of the orbit through the neighbourhood. This can either be done by sorting all distances and rejecting those which come from adjacent points of the training time series, or by stepping through all distances, and skipping a set number of time points after having identified a local minimum of the distances. If entry *k'* of the training Principal Components has been identified as one of the nearest neighbours, then the entry $k = k' + n_x - 1$ is the neighbour to the latest measurement.

The number of nearest neighbours, n_n, to use for the forecasting depends on the dimension of the reduced system and how densely the phase space is covered by the training attractor. If too few neighbours are chosen, the ensemble prediction might not capture the divergence or convergence of the attractor and hence may not give a good estimate of the forecasting error. If too many neighbours are chosen, the nearest neighbours may not be that near and no longer be a good representation of the local dynamics, hence introducing errors into the forecasting.

2.6 Predicting using nearest neighbours

Once the nearest neighbours have been identified, each can be moved forward in time by the lead time or forecasting horizon while sampling all intervening time steps. A key assumption in the implicit forecasting here is that the current point will evolve alongside the identified nearest neighbours from the training data. This means that the relative position of the point from the Training attractor at time $k = k' + n_x - 1 + T$ will have a similar position relative to that of the current measurement predicted a lead time *T* ahead. If the current distance vector to nearest neighbour *j* is

$$D_j = P_r^{k_j} - P_n^{n_x} \qquad (7)$$

then the prediction based on this nearest neighbour is

$$P_f^j \left| T \right| = P_r^{k_j + T} + D_j \qquad (8)$$

The ensemble of $P_f^j \left| T \right|, j = 1 \ldots n_n$ is then the ensemble prediction, each member of the ensemble is mapped back onto the delay matrix space by using

$$Y_f^j = P_f \Lambda_r S_r \qquad (9)$$

Each of the Y_f^j returns the predicted wind speeds for the next T time steps as the entries $u_p^j|+1...T| = Y_f^j|N-T+1...N, M_w|$. This ensemble of predicted wind speeds can then be used to calculate the expected velocity as their average, and an estimate of the uncertainty based on the standard deviation: $\sigma_p|t| = \langle u_p^j|t| \rangle_j$.

Likewise, if wind direction is used as a second observable, this can be reconstructed by $\theta_p^j \ +1 \ldots T \ = Y_f^j \ N-T+1 \ldots N, \ 2M_w$.

3. Data and Methodology

3.1 Dataset

The data used for this analysis originated from the Gogarbank surface station in Edinburgh provided through the UK Met. Office – MIDAS Land Surface Station record [21]. The site used an anemometer 10m high above ground and the data records used spanned from 1998-2010 with hourly mean wind readings with the wind speed stored to the nearest knot (1 kn=0.5144 m/s) and the wind direction in degree to the nearest 10°. For this analysis, wind speed and wind direction data were used with the wind speed converted to m/s. An illustration of the data, the wind speed is shown in Fig. 1 for the 2-year period covering 2008 and 2009. The data not used as the training data set were then used for testing the method. A section from the test period was used to apply the prediction model, and the predictions for the 24 hours following that section were then compared against the actual data for the 24h period following that section.

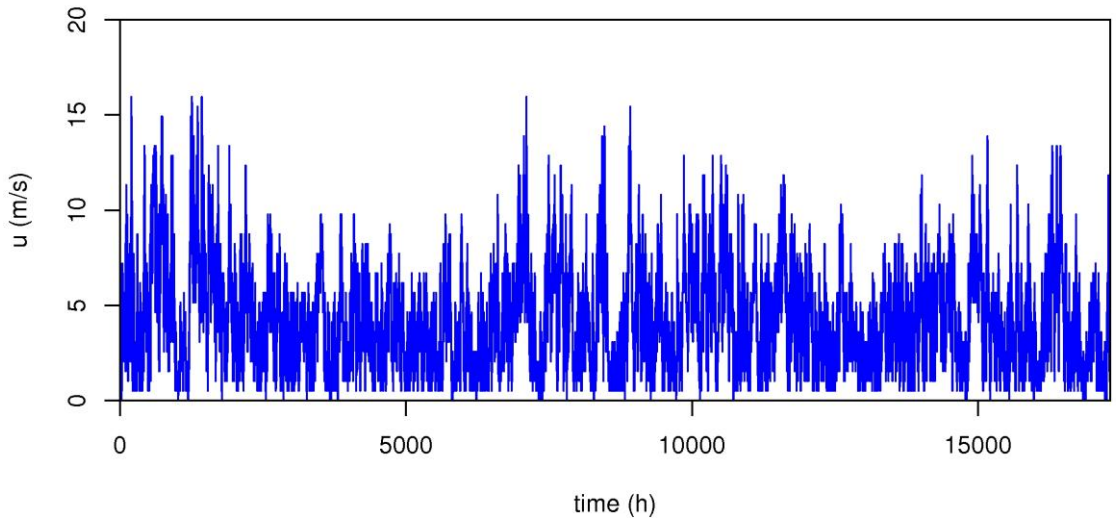

Fig. 1. Wind speed time series for Gogarbank 2008 and 2009.

Table 2. Summary of data used for training and forecasting, with parameter settings used.

	$\mathbf{M_w}$	$\mathbf{M_r}$	$\boldsymbol{n_x}$	$\boldsymbol{n_n}$	**Forecast**	**Forecasting horizon**
Reference values	1 day	16	1	5	2010	24h
Range	1 day - 2 weeks	5-35	1-3	2-10	1999-2007	1-24h

3.2 Analysis setup

From the available record, two 2-year records were chosen as the training period, either the years 2008 – 2009 or 2000 – 2001. For all examples discussed in section 4, the time lag chosen to create the delay matrix was equal to the sampling period of the data, τ= 1 h, but a range of delay window lengths, M_w, ranging from 1 day (i.e., 24 readings) to 2 weeks (336 readings were used. The reference case for the discussion in the results section is the window length of 1 day for the training period 2008–2009 but two days for the training period 2000–2001, as indicated in Table 1 which also summarises the other parameter chosen for testing the method. For the case of a 2-year training period (17520 hours), a 2-week window (336 hours) of wind speed and direction, the delay matrix will have 672 columns and 16848 rows, leading to a principal component matrix of the same dimension, 672 singular values, and 672 singular vectors of length 672 each.

Because wind direction is a circular variable, one either has to be aware that there is an apparent discontinuity between 360° and 0° or transform the wind speed and wind direction variables into a pair of horizontal velocity components, $u_x = u \sin \theta$ and $v_y = u \cos \theta$. In the present case, we used the direction as a direct input. As there were virtually no cases of the direction jumping across the 0°/360° boundary, we felt that we did not introduce any error. However, for locations with a wider spread of wind directions, it is recommended that the data should be transformed to the velocity components.

Of the singular values (lambda), of which the first 90 are shown in Figure 2, only a few of them have high values which drop off rapidly and then settle to a plateau from the 20th on. From this figure it is clear that at least the leading four dimensions must be retained in the model but that including more than 20 would add increasingly noise to the predictions. For that reason, a truncation of M_r= 5 to 35 was explored.

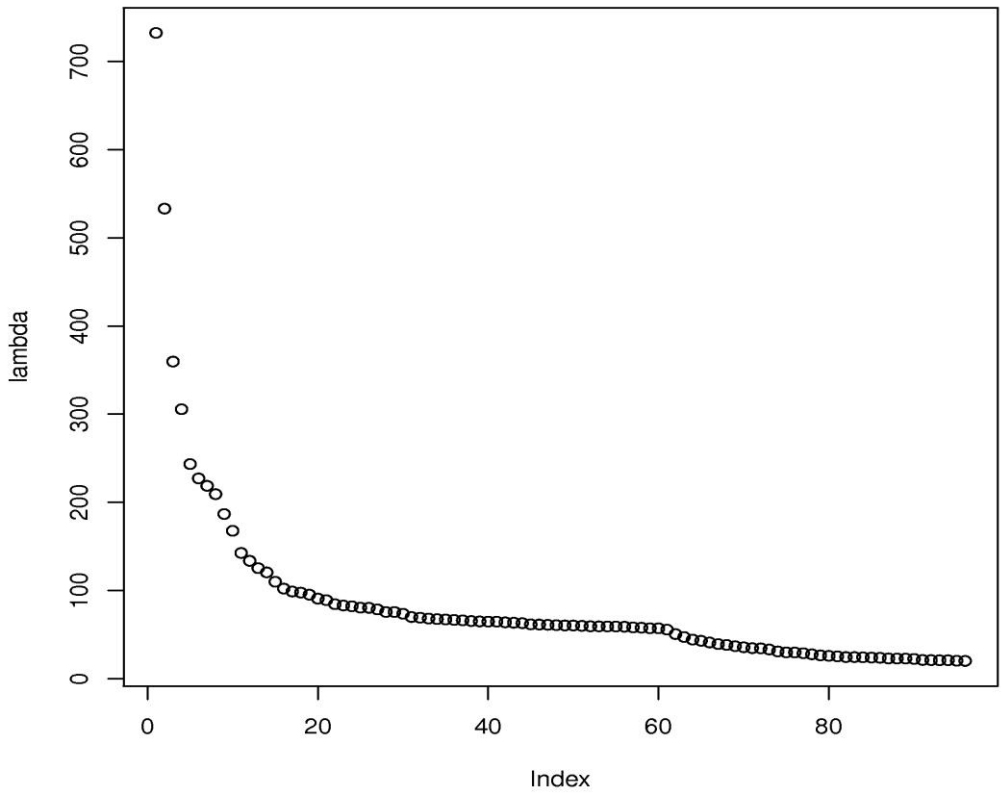

Fig.2. The first 90 Singular values for the PCA of the 2-year training set with window length of 2 weeks.

The first three singular vectors (svec[,1],svec[,2] and svec[,3] respectively) for the PCA applied to the 2008–2009 data using a 48h window are shown in Fig. 3. Since the input data are the wind speed and the wind direction, each singular vector contains two distinct sections, where the first 48 entries correspond to the temporal evolution of the wind speed attributed to that singular vector and the entries 49 to 96 correspond to the wind direction. Fig. 3.a and 3.b show that the first two singular vectors are associated with a slow modulation of the weather, while the third singular vector in Fig. 3.c and the fourth singular vector (not shown) correspond to a daily cycle. The phase space diagram drawn by the first two Principal Components (pc[,1] and pc[,2]), shown in Fig. 4 shows an attractor with a clear structure associated with the prevailing weather conditions in Scotland, and the transition between them.

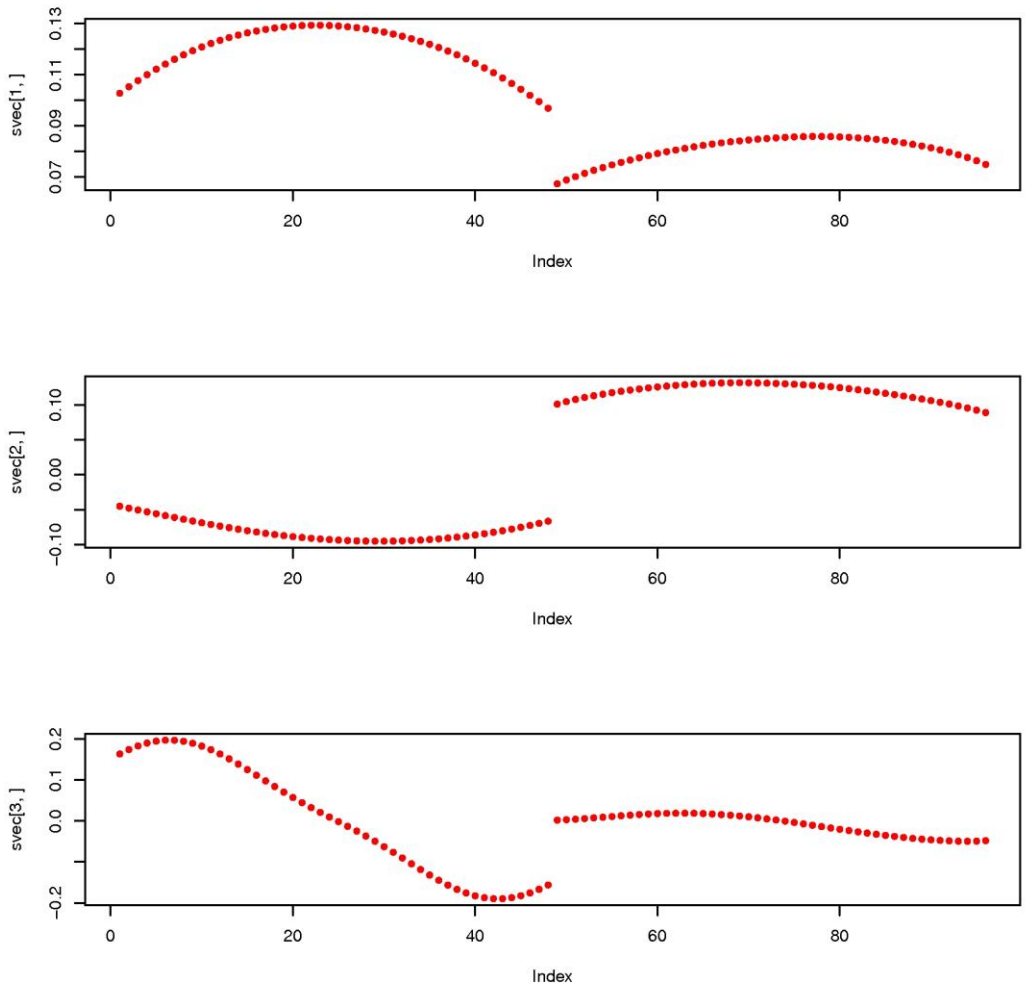

Fig.3. First three singular vectors Fig. 3(a),3(b),3(c). The line between index 48 and 49 separates wind speed on left from direction on the right.

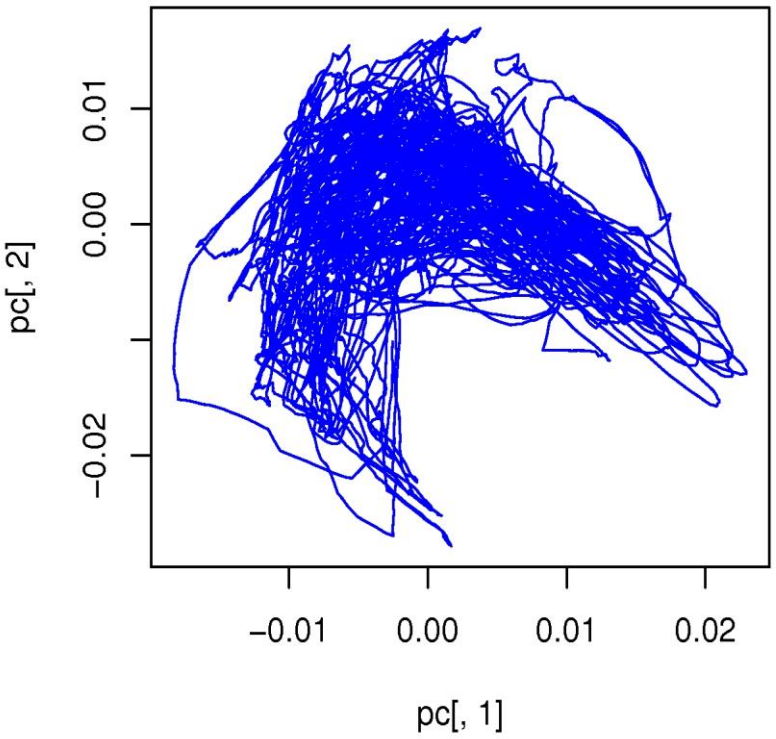

Fig.4. Phase portrait constructed from the first two principal components.

Finally, the parameters for the forecasting component were the length of the orbit section to be projected onto the attractor and the number of nearest neighbours which had to be chosen. For the orbit length a range of 1 to 3 was chosen. That means that, for a window length of, for example 48 hours, a section of 48h, 49h, or 50h, respectively was chosen from the test data to create a delay matrix consisting of 1, 2 or 3 rows, correspondingly. The number of nearest neighbours explored in the analysis ranged from 2 to 10, as summarised in Table 2.

With the model defined by the M_r singular vectors and the past data describing the observed dynamics through the M_r principal components, the new measurements for the forecasting were transformed using the same parameters and then projected onto the observed dynamics. This is illustrated by Figure 5 where the attractor from the training data is the grey object. The blue circle is a single point in the phase space created by a time series section of the window length M_w. In this example, n_n= 5 and the five nearest neighbours on the orbit of the training data are, in order of proximity, identified by the red numbers in Figure 5. These five nearest neighbours can then be traced forward in time over the forecasting horizon, which is shown by the red curves evolving from the numbered positions. Each of these can then be re-transformed to wind speed and direction to produce the ensemble forecast. The final result is then a forecast of the predicted mean wind speed and the uncertainty in that prediction for all lead times from one hour ahead to the specified forecasting horizon, 24 hours in our analysis.

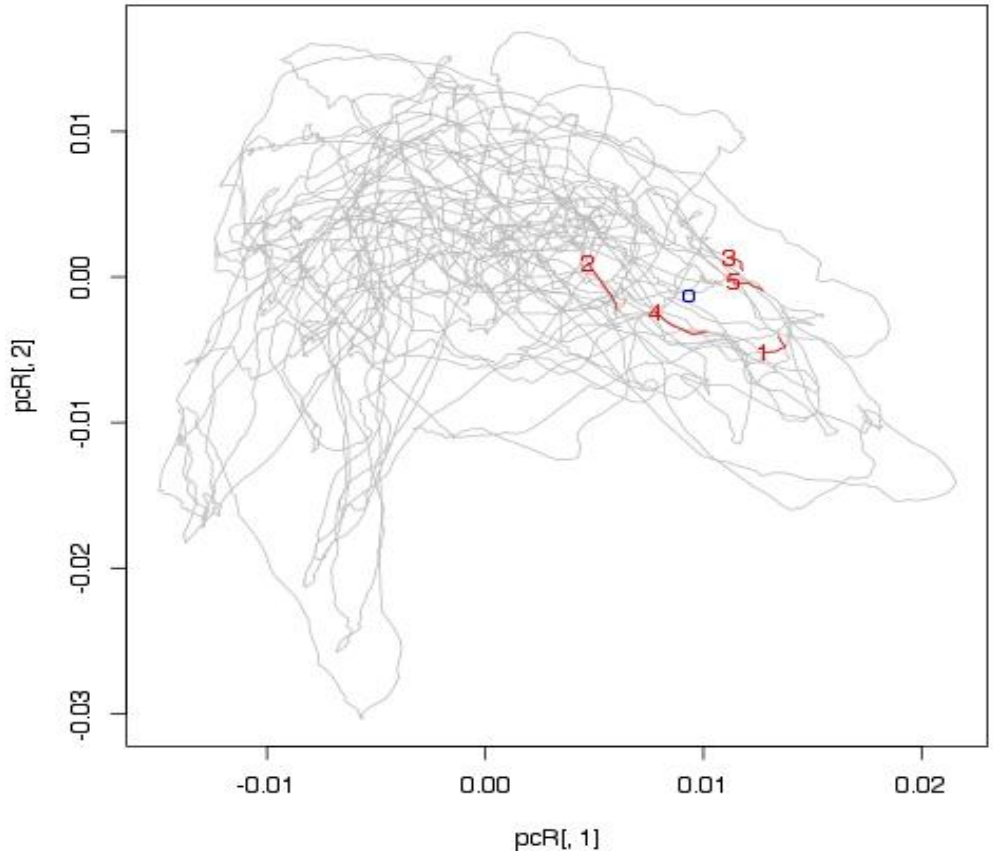

Fig. 5. New data mapped onto training set. The blue circle is the new 'current' observation, and the five red numbers are the nearest neighbours which were then found to evolve for the specified forecasting horizon as shown by the red lines.

3.3 Performance evaluation

To evaluate the performance of the predictions, the predictions are compared against the actual values from the test data, using the three main measures recommended by Madsen et al. [22] albeit for wind speed rather than power output. They are all based on the prediction calculated as the difference between actual observation, u, at time $t+T$ from the test set and the wind speed predicted for that time based on the observation at time t, $\hat{u}$,as

$$e\ t+T\ \ t) = u\ t+T\ - u\ t+T\ \ t) \tag{10}$$

These three measures are the bias

$$BIAS\ T\ = \mu_e\ T\ = e\ T\ = \frac{1}{N}\ \ \sum_{t=1}^{N} e\ t+T \mid t \tag{11}$$

the mean absolute error (MAE), frequently used in the literature, e.g. [9]

$$MAE\ T\ = e(T)\ = \frac{1}{N}\ \ \sum_{t=1}^{N} e\ t+T \mid t \tag{12}$$

and the root mean squared error (RMSE)

$$RMSE\ T\ = \sqrt{\frac{1}{N}\ \ \sum_{t=1}^{N} e\ t+T \mid t\ \ ^2} \tag{13}$$

These errors for the predictions using the PCA forecasting were then benchmarked against the frequently used persistence, $\hat{u}_{ref}(t+T \mid t) = u(t)$. This benchmarking is quantified by an improvement measure as defined [21], e.g., for the BIAS (and likewise for MAE and RMSE) as

$$Imp_{ref,BIAS}\ T\ = \frac{BIAS_{ref}\ T\ - BIAS\ T}{BIAS_{ref}\ T} \tag{14}$$

Since the PCA forecasting intrinsically returns all predicted time steps at the sampling interval until the prediction horizon or lead time T, we also use average of $Imp(T)$ over $T = 1 \ldots T_{max}$. The sensitivity of the PCA forecasting method to different choices of the parameters is here described in terms of the overall improvement of the MAE over persistence:

$$PI = \frac{1}{T_{max}} \sum_{T=1}^{T_{max}} Imp_{p,MAE}\ T \tag{15}$$

where the maximum lead time in our case is 24 hours.

4. Results

4.1 Forecasts of wind speed and uncertainty

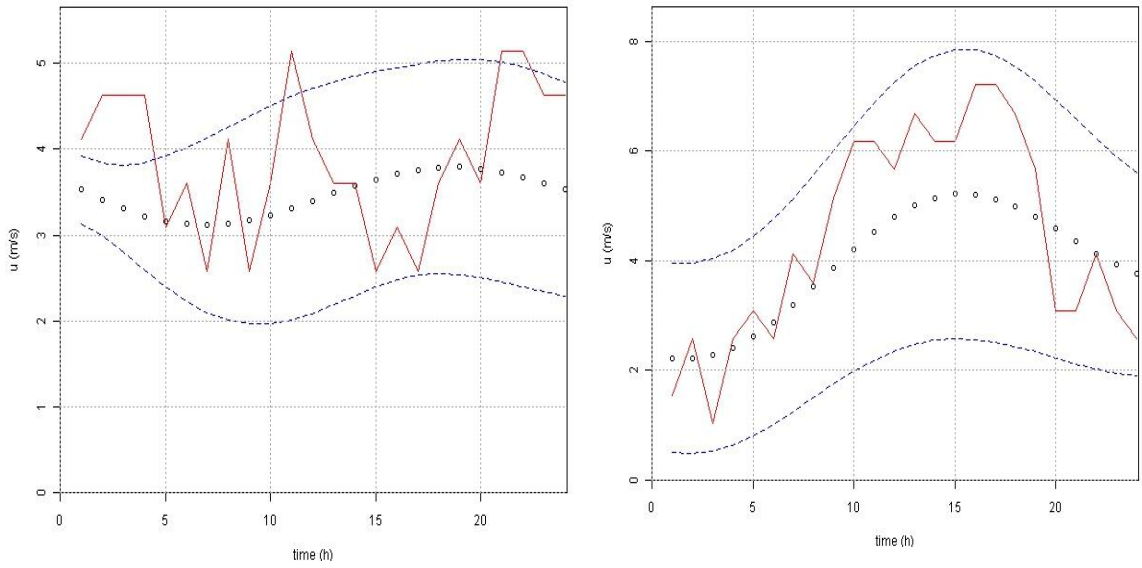

Fig. 6. Comparison of actual wind speed (solid red line), forecasted wind speed (open black circles) and uncertainty of wind speed (dashed blue lines). Fig. 6(a) is a "bad" prediction example whereas Fig.6(b) is a "good" example.

Figure 6 illustrates a comparison of the ensemble forecast representing all 24 hours of lead time for two of the 100 predictions made for this analysis, where the black circles show the forecast and the red line the actual wind speed. As outlined in section 3.2, the predictions made in the phase space were re-transformed to real wind speed and direction. From the ensemble of n_n= 5 forecasts, the prediction was calculated from the mean of the ensemble (open black circles) and the prediction uncertainty was also found with the use of the standard deviation (dashed blue lines).

As both examples in Fig.6 show, the predicted wind speeds form a strongly smoothed curve compared to the actual winds, as the PCA has successfully separated the slow atmospheric dynamics from the unpredictable local turbulence. For a very good prediction at all lead times

from one hour ahead to the forecasting horizon, the black dots would follow the red line very closely, while for an acceptable prediction, the actual wind speed should lie within the band specified by the uncertainty of the prediction. Conversely, wind speeds outside the band would have been poorly predicted.

Fig. 6(a) is an example where the forecast is relatively poor at times due to very large hourly variations in the wind speed. Nevertheless, the prediction is consistent with the observations for most times and, more importantly, the excursions of the wind speed outside the predicted range at the higher lead times are only slightly outside the predicted margin. Fig. 6(b) is a case where the prediction is good. Furthermore, the model predicts a higher uncertainty for lead times between 10 and 20 hours after which the predicted uncertainty suggests a return of predictability for the day-ahead forecast. This is exactly borne out by the actual observations which follow the predicted mean very well but show a persistent error within the 10h to 18 hour lead time.

4.2 Forecasting quality

To quantify the performance of this model we used as the first measure the mean absolute error, *MAE,* as defined in equation (12). The reason for concentrating on this measure is that it gives a direct comparison of the error with the predicted uncertainty. If the *MAE* is less than the uncertainty, the prediction is as good as it can be (and is known to be) but if the *MAE* is much larger than the predicted uncertainty, the model does not work for that data set.

Fig.7 shows the *MAE*(*T*) as the solid red line against the lead time for the case of a 2-week training window (M_w= 336 h), a model predictor dimension of *D*= 15, matching a point on the attractor (n_x= 1) , and using n_n= 5 nearest neighbours. The open black circles are the average of the uncertainties predicted for that lead time and the dotted line is the standard deviation of these predicted uncertainties. Superimposed on this is also the mean absolute error for persistence, $MAE_p(T)$ as the green dash-dotted line. As the figure shows, the actual *MAE* is very close to the predicted uncertainty at short lead times but much higher than the error from persistence. The model performs slightly worse than predicted from its own internal dynamics at lead times between 8 and 20 hours but still within the range of calculated predictions. The key features of the error of persistence compared to that of the PCA model is that persistence is much better than PCA at short lead times up to 6 hours but that PCA outperforms persistence at longer lead times. The fact that persistence is often the best predictor for short lead times was also supported by Madsen et al [22] and can be explained in that the short-term fluctuations affect the local wind at these times more than any slow synoptic weather changes. Based on this, we propose a refinement of the PCA-predictor by merging it with a persistence-based correction at short lead times.

4.2.1 Combining persistence and PCA

After performing this comparison and applying several inputs for the different parameters used by PCA, it was concluded that the respective strengths of persistence and PCA could be exploited in a combined forecast by applying a filter to the PCA prediction [4]. This filter constructs a weighted average of the persistence prediction and the PCA prediction for a filter length long enough to cover the range where persistence outperforms PCA prediction. Over that filter length, the weights of the averaged change linearly from 1 for persistence and 0 for PCA at the 'current' time (lead time = 0h) to the other extreme of 0 for persistence and 1 for PCA at the end of the filter length. The filter is of the form:

$$u_{f,i} = \begin{cases} 1 - \frac{i}{N_f} \;\; u_0 + \frac{i}{N_f}\, u_{PCA,i} & \text{for } i = 0 \ldots N_f \\ u_{PCA,i} & \text{for } i > N_f \end{cases} \qquad (16)$$

where i is the lead time , N_f the filter length, $u_{PCA,i}$ the ensemble forecast and u_0 the current wind speed. By trial and error a good filter length was found to be between 10 h and 15 h, with little change of the results in that range.

The effect of applying such a correction on the performance of the predictor is shown in Fig. 8, where it is clear that the very short term prediction, up to a lead time of 6 h is now as good as for persistence and that the prediction for longer lead times is dominated by the ability of PCA to extract the slower atmospheric dynamics.

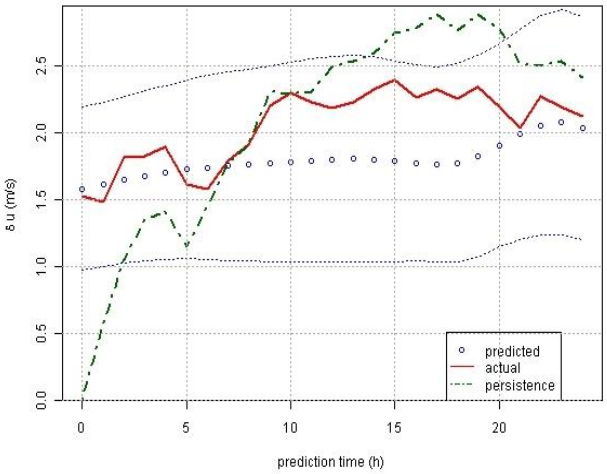

Fig. 7. Comparison of annual mean forecasting error and uncertainty (unfiltered data).

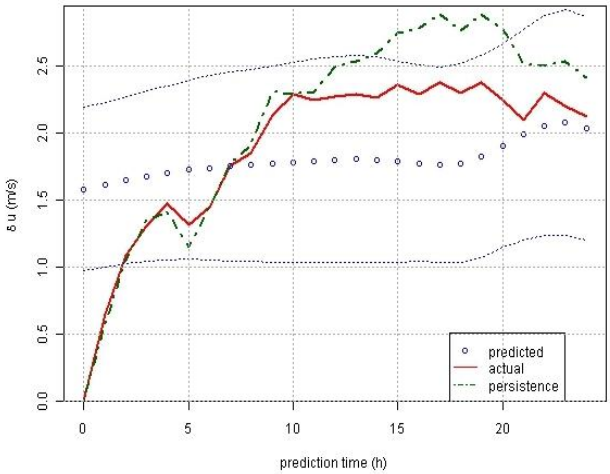

Fig. 8. Comparison of annual mean forecasting error and uncertainty (filtered data).

4.2.2 Other error measures

Following the recommendations of Madsen et. al. [22] the alternative error measures of bias (11) and RMSE (13) were calculated and are shown in Figures 9 and 10. They both indicate that PCA outperformed the persistence method and specifically for the bias error measure, PCA performed substantially better than persistence.

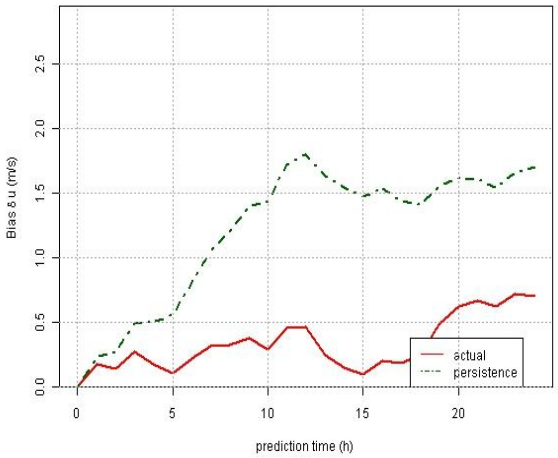

Fig. 9. Comparison of bias between PCA and persistence method.

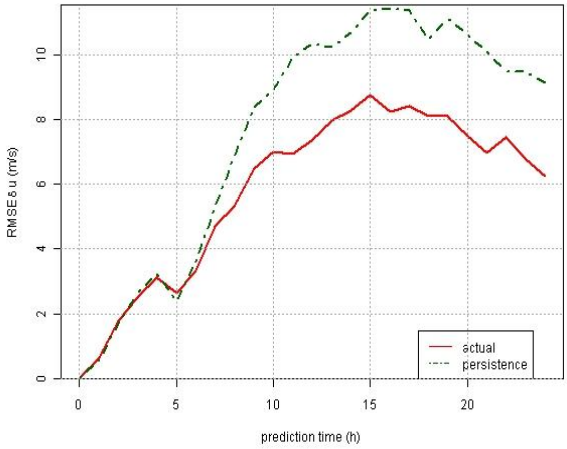

Fig. 10. Comparison of RMSE between PCA and persistence method.

4.3 Sensitivity analysis of parameters

Figures 11 to 13 show the performance index of the results for the different choices of first the length of orbit to use for finding the nearest neighbours on the attractor, n_x, secondly the number of nearest neighbours, n_n, and finally the embedding dimension, M_r, respectively.

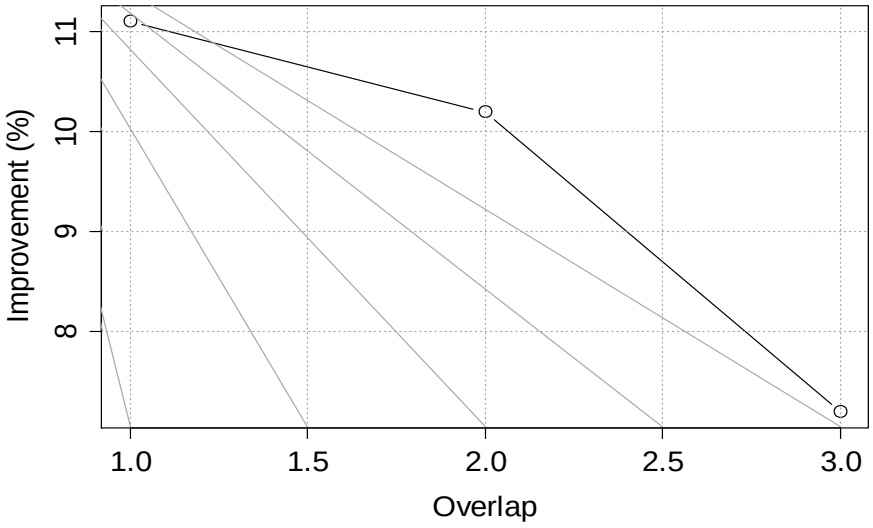

Fig.11. Improvement of PCA results in % for different overlap values.

Figure 11 indicates that using a single point (n_x = 1) rather than fitting a short time series of point (n_x > 1) overlap seems to yield the best improvement (around 11.2%) of the results. This means that the PCA results are 11.2% closer to the actual results in comparison with the persistence method. Using n_x > 3 did not work reliably for our data set as there were not enough

nearest neighbours. After determining that n_x = 1 seems to be the best, this was used for analysing the sensitivity to the number of nearest neighbours, n_n, Figure 12 shows that the overall improvement initially rises substantially from below 8% for only two neighbours to above 11% for five nearest neighbours but then drops again to around 9%. Using too few or too many neighbours might not be appropriate since with too few (i.e. less than 5) the information we use for the analysis might be too little. Conversely, using too many (i.e., more than 5 in our case) requires using information from progressively distant parts of the attractor which are resembling the current observation less and less. There is clearly a distinct optimum which needs to be determined but it is not clear whether it is at or around five nearest neighbours for any data set or whether this must be determined from optimising the parameters through experience at each site individually.

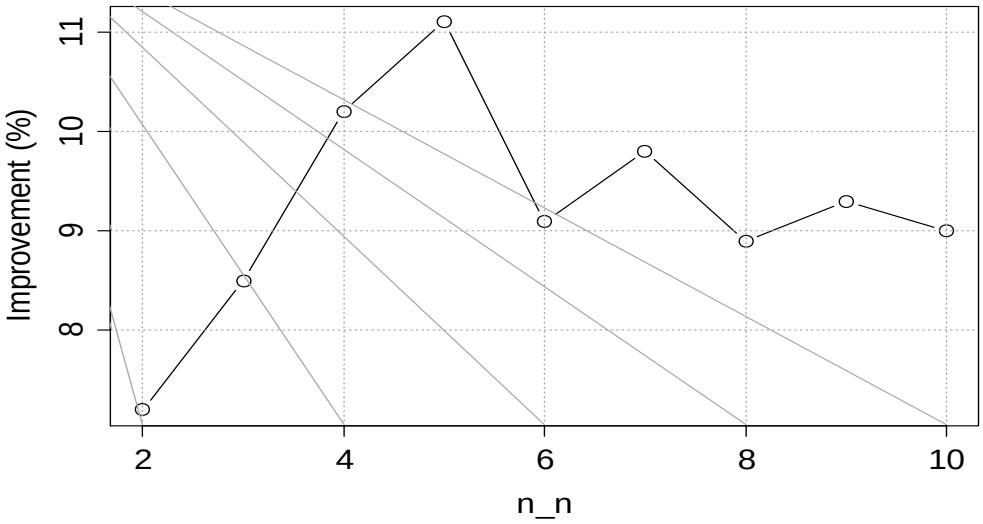

Fig.12. Improvement of PCA results in % for different nearest neighbours values.

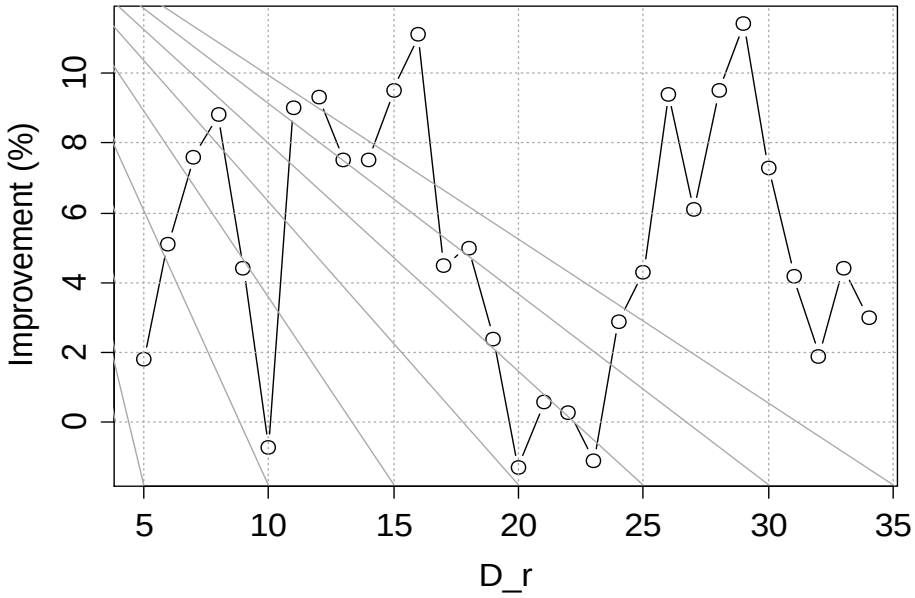

Fig. 13. Improvement of PCA results in % for different reduced dimensions values.

Finally, Figure 13 shows the sensitivity of the model to the choice of the model dimension. Here, it can be seen that different choice of reduced dimensions results in a big variation of the percentage of improvement. The amount of dimensions which the improvement seems to be more consistently high for (5.6%) is around 16. It should be noted that adding more dimensions results in adding more information but whether this information is useful or not is another issue which should be of further investigation and of course depends on the site and wind dynamics used for the analysis.

5. Conclusions

The main conclusions of this research that can be made are firstly that PCA is capable of identifying weather regimes by being able to represent the wind measurements in the form of an attractor with a clear structure. Furthermore, it was demonstrated that this can be done both, by just using wind speed measurements and by using multivariate measurements, such as wind speed and wind direction combined.

Applying the PCA to wind forecasting demonstrated that the method is a reliable forecasting method for forecasting wind speeds hours ahead to day ahead. By combining the PCA prediction with persistence prediction at very short time scales, it was possible to eliminate the weakness of applying PCA to a coarsely sampled wind record.

One of the most useful aspects of PCA over some other forecasting techniques is that it is based on an ensemble forecast using ensembles of similar past events. This allows an estimation of the forecast accuracy at the time when the forecast is made. The analysis showed that this estimated forecast uncertainty is a reliable predictor of the actual forecasting error. This knowledge will be useful for the wind farm operators to evaluate their forecasts and will help with their decision making.

Acknowledgments

The authors would like to thank the UK Meteorological office for providing access to the MIDAS record wind data through British Atmospheric Data Centre http://www.badc.ac.uk/. Also Miss Skittides is grateful to SgurrEnergy Ltd. and Energy Technology Partnership (ETP) for their financial support.

References

[1] Renewable UK (formerly the BWEA British Wind Energy Association), http://www.bwea.com/ (accessed 16 April 2012).

[2] L. Landberg , Short-term prediction of the local wind conditions, Journal of Wind Engineering and Industrial Aerodynamics. 89 (3-4) (2011) 235-245.

[3] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, Wind power forecasting: state –of- the- art 2009. Argonne National Laboratory ANL/DIS-10-1, November, (2009), http://www.dis.anl.gov/projects/windpowerforecasting (accessed 20 May 2013).

[4] G.Giebel, R.Brownsword, G. Kariniotakis, M. Denhard, C. Draxl, The state-of-art in short-term prediction of wind power, A literature overview, second ed., ANEMOS.plus, (2011).

[5] X. Wang , P. Guo, X. Huang, A Review of Wind Power Forecasting Models, Energy Procedia. 12 (2011) 770-778.

[6] J. Zhou, J.Shi, L. Gong, Fine tuning support vector machines for short-term wind speed forecasting, Energy Conversion and Management 52 (2011) 1990-1998.

[7] E. Erdem,J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction, Applied Energy 88 (2011) 1405-1014.

[8] M. G. De Giorgi, A. Ficarella, M. Tarantino, Error analysis of short term wind power prediction models. Applied Energy 88 (2011) 1298-1311.

[9] M. G. De Giorgi, A. Ficarella, M. Tarantino, Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods.,Energy 36 (2011) 3968-3978.

[10] W. G. Früh, Evaluation of simple wind power forecasting methods applied to a long-term wind record from Scotland, International Conference on Renewable Energies and Power Quality (ICREPQ'12), Santiago de Compostela, 2012; paper 715.

[11] J. Shi, J.Guo, S. Zheng . Evaluation of hybrid forecasting approaches for wind speed and power generation time series, Renewable and sustainable Energy Reviews 16 (2012) 3471-3480.

[12] A.M. Foley, P.G. Leahy, A. Marvuglia, E.J. McKeogh, Current methods and advances in forecasting of wind power generation, Renewable Energy 37(1 – 8) (2012).

[13] D.S. Broomhead , R. Jones, J.P. King, E.R.Pike, Singular system analysis with application to dynamical systems in ER Pike and LA Lugiato (eds.). Chaos, noise and fractals. Adam Hilger, Bristol,1987, pp.15-27.

[14] N. Golyandina and A. Zhaglijavsky, Singular Spectrum Analysis for time series, Springer Verlag, Heidelberg, 2013.

[15] B. Everitt, G. Dunn, Applied Multivariate Data Analysis, second ed.,Arnold, (2001).

[16] N.T. Palmer, A nonlinear dynamical perspective on climate change, Weather 48 (1993) 314–326.

[17] F. Bakalian, H.Ritchie, K. Thompson, W. Merryfield, Exploring atmosphere-ocean coupling using Principal Component and Redundancy Analysis, Journal of Climate 23 (18) (2010) 4926–4943.

[18] M. R. Allen, L. A. Smith, Investigating the origins and significance of low-frequency models of climate variability, Geophysical Research Letters 10 (1994) 883-886.

[19] F. Takens, Detecting strange attractors in turbulence. In Rand DA and Young LS, editors, Dynamical systems and turbulence, Warwick ,1980: Lecture notes in Mathematics 898, pp. 366–381. Springer-Verlag, Berlin, 1981.

[20] P.L. Read, Phase portrait reconstruction using multivariate singular systems analysis, Physica D 69 (1993) 353-365.

[21] British Atmospheric Data Centre, Met Office – MIDAS Land Surface Station data. http://www.badc.ac.uk (accessed 16 April 2012).

[22] H. Madsen , P. Pinson,G. Kariniotakis ,H.A. Nielsen,T.S. Nielsen, Standardizing the performance evaluation of short-term wind power prediction models, Wind Engineering 29 (2005) 475 – 489.