Wind resource forecasting using enhanced measure correlate predict (MCP)

Citation for published version:
Zakaria, AA, Fruh, W-G & Ismail, FB 2018, Wind resource forecasting using enhanced measure correlate predict (MCP). in *6th International Conference on Production, Energy and Reliability 2018: World Engineering Science and Technology Congress (ESTCON).*, 040005, AIP Conference Proceedings, no. 1, vol. 2035, AIP Publishing, 6th International Conference on Production, Energy and Reliability 2018, Kuala Lumpur, Malaysia, 13/08/18. https://doi.org/10.1063/1.5075569

Digital Object Identifier (DOI):
10.1063/1.5075569

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published In:
6th International Conference on Production, Energy and Reliability 2018

Wind Resource Forecasting using Enhanced Measure Correlate Predict (MCP)

Firas Basim Ismail[1, a)] As'ad Zakaria[1, b)], Wolf – Gerrit – Früh[2]

[1] Department of Mechanical Engineering, College of Engineering, Universiti Tenaga Nasional(UNITEN),43000 Kajang, Selangor Darul Ehsan, Malaysia.

[2] School of Engineering & Physical Sciences; Mechanical, Process & Energy Engineering, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom.

[b)] Firas@uniten.edu.my
[a)] asadzcm@hotmail.com

Abstract. The enhancement of Measure Correlate Predict (MCP) using Principal Component Analysis (PCA) is a new wind prediction method based on studying the patterns of historical wind data. The method is trained based on past wind data to predict the wind speed using an ensemble of similar past events. The method is tested based on Meteorological Office (MET-Office) wind speed from a reference site that spans from 2000 to 2010. The last two years (2009 to 2010) were used as training years where the MCP – PCA algorithm learns (predicts) the wind patterns between the reference(s) and target(s) site. (the first 8 years learn pattern, the last two years predict and compare with actual which is called training period). The prediction result is then compared to the actual wind speed distribution at the target site of the training years. The method is further tested with an increase in number of reference sites for predictions. The new prediction results show that the prediction error improves to 23.1 % in average in comparison to a standard linear regression method.

Keywords: Prediction, Measure Correlate Predict with Principal Component Analysis (MCP – PCA)

1.0 Introduction

Wind Energy has been growing rapidly in the timeframe of the past 10 years [1] with the total installed capacity worldwide increased significantly from circa 23.9 GW in 2001 to approximately 486.8 GW in 2016 [2], [3]. The highest capacity of wind energy was installed in China (168.7 GW), followed by the USA (82.2 GW), and Germany (50.0 GW) [3]. Through advancements and researches, the wind energy generation is financially attractive as the cost per kWh has significantly decreased over the last years [4]. Although the energy is abundant in most parts of the northern Europe [5], wind energy is always fluctuating, thus making it hard to predict its behavior. Wind directions and wind speed are the most important components in generating the energy [1]. Energy generated from turbines depends greatly on the cube of the wind speed. Slight changes in the speed will cause a huge change in output, making a robust wind forecasting method very desirable especially for the wind farm operators. Furthermore, turbine often has delays in their response due to adjustments to the wind fluctuations [1], hence increasing the importance of having a reliable wind forecasting method for wind farm developers and operators. Wind resource assessments are normally carried out on site for a period of several years depending on the techniques of forecasting that will be applied. As a general rule of thumb, wind farm developer would need a wind resource assessment of approximately 20 years which is costly [6]. However, assessing the resource for a short period of time would prove to be insufficient in order to have low errors in the predictions [6]. This is why a Measure – Correlate – Predict (MCP) has been the most used method in the wind farm industry as this method requires a significantly shorter wind data sets from the proposed site in order to predict the long-term wind behavior at the site [7]. Nonetheless, it needs a reference site which has similar properties (common dynamical systems) which has a long and reliable wind record [7]. For the dissertation, a more sophisticated version of the MCP called Principal Component Analysis (PCA) is used. Wind forecasting is normally divided into two sections e.g. statistical and physical approach [8]. Each of them are divided into three commonly used time horizons which is immediate short-term (0 – 8 hours ahead), short – term (8 – 24 hours ahead), and long-term (24 hours – several days ahead) [1]. The MCP – PCA falls under these categories of commonly used horizons. The statistical method analyzes the historical time series of wind data of a certain reference and target site [9].The correlations between the proposed and reference site are then determined. In addition, the statistical approach does not include any meteorological data as its inputs [8]. It usually involves the artificial intelligence and time series analysis approach. There are several popular statistical wind forecasting techniques aside from MCP that is widely used. The most common ones are Auto Regressive Moving Average models (ARMA) that uses linear methods for forecasting [1]. A hybrid (statistical and physical) of the model is used according to [10] that produces a better result in predicting the wind direction. However, worse result is recorded for the prediction of the wind speed. The other popular statistical method is the Artificial Neural Networks (ANN) [4] [11]which uses an artificial intelligence system which acts like a brain to perceive input data and figuring out its behavior. This system learns the pattern of the wind resource over a period of time and a prediction is made from its 'learnings'. A case study in the Canary Island shows that the ANN performs better (87 % of the cases) than the traditional MCP – Variance Ratio Method (VRM) [12]. But MCP – VRM outperforms ANN (with a single target site as a reference) when the correlation between target and reference site is high. ANN only outperforms VRM in most of the cases where the reference site is more than one. A forecast approach is normally deployed based on the characteristics, whether it is linear or non-linear, the forecasting horizon, and the time window [1]. According to Zhang et. al [9], statistical approach is good in predicting the wind speed in an immediate-short-term (0 – 8 hours) time window. They proposed based on their study in several wind farms in China, using a hybrid of Singular System Analysis (SSA) with Autoregressive Integrated Moving Average (ARIMA), ANN, and Support Vector Machine (SVM) reduces the Root Mean Squared Error (RMSE) obtained for most of the wind farms (some up to 40% of reduction in RMSE) [9]. According to Skittides and Früh [1], MCP – PCA outperforms the so called 'naïve' persistence method [13] where persistence is normally used as benchmarks in forecasting windows for up to more than 10 hours. This study also shows that the MCP – PCA is reliable in forecasting the wind speed hours ahead and days ahead. Short term wind prediction of the MCP – PCA can be minimized with the hybrid of the MCP – PCA with persistence as it outperforms the MCP – PCA in a short time window in the mentioned case study. According to Sun et al [3], the combination of PCA method such as Kernel – PCA outperforms several popular forecasting methods including the typical integrated approach. Focusing on the strength of PCA, MCP – PCA has a unique ensemble method that gathers the similar past events together and bases the forecasting upon these events [1] . Therefore, a forecast accuracy is high due to the fact that events are separated when the forecast is made. Case studies in this work are made based on pairings of reference and target site to evaluate the result of the MCP – PCA. Furthermore, the approach can also be applied to other forecasting problems such as stock trend forecasting, temperature forecasting, electricity demand forecasting, as well as solar irradiance forecasting [3]. Previously, tests on MCP – PCA are done on a one to one basis which means one reference to one target only [1]. By adding several references into the algorithm of the MCP – PCA against a target, the change (improvement or deterioration) in prediction error is quantified. The main objective is to test the prediction quality by introducing more references into the algorithm. The results from the additional references are compared to the standard linear – regression method and the one to one (one target against one reference) cases of MCP – PCA.

2.0 Methodology

A time delay method is used in the PCA methodology to define equal variables to the phase space's variables [1] . This method aids in reconstructing the phase space from a given time series in dynamical systems. The reconstruction process of a time series can be useful in a way that it can concentrate on bringing out the most significant pattern that characterize the whole system. It is often associated that the challenge arises from lack of data i.e. having data from only one site or considerably less data from another sites. Is it still possible to predict properly for all the sites if the PCA algorithm has multiple inputs e.g. multiple source of reference and target data? More precisely, are the data from several references and targets help the prediction as a lot more data is present or is it going to 'confuse' the system due to significant increase in data processing. Firstly, the variables (wind speed and wind direction) of the data set need to be rescaled so that they contribute to the system equally. It is accomplished by subtracting the mean from each variable and then dividing it by the variance. This will produce the 0 to 1 scale for each variable that is suitable for the analysis. PCA is then applied to the wind speed and the wind direction where they are expressed in vectors as can be seen in equation (1) to avoid any discontinuity across all wind directions.

$$U = -Usin\theta, \; V = -Ucos \qquad (1)$$

The wind speed measurements at time, t for the reference (U_{ref} , V_{ref}) and target (U_{target} , V_{target}) sites, can then be expressed as follows in equation (2):

$$U_{ref,n}(t), V_{ref,n}(t), U_{target,n}(t), V_{target,n}(t) \qquad (2)$$

Where the index n defines the number of reference(s) or target(s) used in each of the PCA-analysis at a given time, t. This can be extended to form a time delay matrix of $y(t)$ as described below in equation (3) where m is the number of the time lags used:

$$y(t) = U_{ref,n}\,(t, t-1, \ldots\ldots t-m),\; V_{ref,n}\,(t, t-1, \ldots\ldots t-m), U_{target,n}\,(t, t-1, \ldots\ldots t-m), V_{target,n}\,(t, t-1, \ldots\ldots t-m). \quad (3)$$

This creates a time delay matrix, Y in which each of its row contains one of the extended sets. The time-delay relies on the choices of parameters, hence PCA is used to optimize the phase space reconstruction. PCA is performed to the time-delay matrix, Y to separate background noises from the main pattern from the time-delay series. With applied PCA, a depiction of shape of the time series as well as the number of needed time-delays can be identified. Hidden structures from the data (i.e. useful information) can be extracted from its relevant parts and the importance in variability of the time series can be explained. The PCA is carried out in R via the SVD (Singular Value Decomposition) built in tool. SVD transforms the basis vectors of the phase space in a way that it finds the orthonormal fundamental vectors which are then used to maximize the variance. The three outputs from the SVD procedure applied to the time-delay matrix, Y from concurrent period can be expressed in equation (4):

$$Y_{n,m} = P_{n,m}\,\lambda_{m,m}\,S_{m,m} \quad (4)$$

Where $Y_{n,m}$, is the time delay matrix. $n = 1 \ldots\ldots N$ rows, $m = 1 \ldots\ldots M$ columns, $P_{n,m}$ is the principal component matrix, $\lambda_{m,m}$ is the diagonal matrix of singular values, and $S_{m,m}$ represents singular vectors. P is the principal component matrix that describes the time series of the system in which it forms an attractor. Each column in the principal component's matrix contains the normalized amplitude of the pattern at any particular instance in time. λ is the diagonal matrix which contains singular values. The diagonal component of λ contains the square root of the variance in the time series in its reciprocal dimensions. These λ components can identify the significant variability in the data. To express it simpler, the λ diagonal entry in its matrix represents the mean amplitude of the pattern which contributes to the measurements. The singular vectors, S are orthonormal in its character. This means that they are orthogonal with a unit length that spans through the dimensions of the phase space. Each column of S matrix represents a singular vector that describes the empirical pattern's behavior. In R, applying the SVD to the time-delay matrix will return the λ values from the diagonal matrix in a descending order. This separates the strong patterns which contributes the most to the variances from the background noise (uncorrelated short-term fluctuations). Truncating the λ and the S matrices to the r columns of strongest patterns, the background noises can be removed as well as optimizing the coherent patterns across the reference and target sites. New principal components, P_n can therefore be calculated with the truncated λ_r and S_r. The new principal components, P_n is expressed in equation (5):

$$P_n = Y_n\,S_r^T\,\lambda_r^{-1} \quad (5)$$

Where T indicates the transpose of the matrix and -1 indicates the inverse of the matrix. Only the first set of columns of the time-delay matrix, Y_n can be filled due to the data that stems only from the historical information from the reference site. As a consequence, only the first sets of columns from Y_n, in this case, Y_h can be filled. Furthermore, only the first half rows of S_r, in this case, S_h and the first half of λ_r which is in this case λ_h can be used. As every row in the singular matrix, S consists of data from all sites, using equation (5) to the historical data activates the analogous components of the targets. It can be shown in equation (6) below to represent the estimated time-delay matrix from the assumptions made above. The target velocity components can then be excerpted from the estimated time-delay matrix, Y_e:

$$Y_e = P_n\,\lambda_h\,S_h = Y_n\,S_r^T\,\lambda_r^{-1}\,\lambda_h\,S_h \quad (6)$$

The error measurement, e as described in equation (7) is used to compare the predictions made for target site from the MCP – PCA method to the actual data from target site for the predicted training years.

$$e = \frac{1}{N}\sum_{t=1}^{N}\left(\left\|\frac{U_{target,n_{pred}}(t) - U_{target_{act},n(t)}}{U_{target_{act},n\,(t)}}\right\|\right).V_{bin}.100 \quad (7)$$

The area of wind probability distribution function for a given time, t until the end of training year, N that is shared between the prediction and the actual wind speed distribution is calculated. It is denoted as the difference in modulus of the wind speed prediction at a given time t, $U_{target,n\_pred}\,(t)$ and the actual wind speed at that given time, $U_{target\_act,n(t)}$. Similar approach is done for $V_{target,n\_pred}\,(t)$ and $V_{target,n\_act}\,(t)$ to avoid discontinuity in all directions. In other terms, the histogram of the prediction(s) and actual target(s) for the training years are created and then the similarity between them is calculated by the absolute difference in each velocity bin multiplied with the width of the velocity bin, V_{bin} of 1 $^{m}/_{s}$. The error, e is then calculated in percentage by multiplying it with 100. If both the prediction and the actual target have the same probability distribution function, the error measurement (error value), e would be zero. However, if the probability distribution between them are completely different, this would yield an error value, e of 1 or percentage – wise, 100 %. Lastly, a performance index, PI as shown in (8) is used to quantify the ratio of prediction error made by the MCP – PCA method in comparison to the standard linear regression method. $e_{MCP-PCA}$, is the error made by the MCP – PCA method using (7) and e_{LR} is the error made by the standard linear regression method using the same equation using the linear regression approach.

$$PI = \frac{e_{MCP-PCA}}{e_{LR}} \tag{8}$$

3.0 Data Source Preparation and Applying MCP-PCA

TABLE 1: Anemometer Locations. *[14]*

No.	Location	Latitude	Longitude
1.	Stornoway airport	58.2138	-6.31772
2.	Edinburgh: Blackford Hill	55.9231	-3.1897
3.	Machrihanish	55.4408	-5.69571
4.	Salsburgh	55.8615	-3.87409
5.	Prestwick: Gannet	55.5153	-4.58343
6.	Edinburgh: Gogarbank	55.9284	-3.34294
7.	Islay: Port Ellen	55.6813	-6.31772
8.	Bishopton	55.9068	-4.53122

Eight data sets containing wind speeds and wind directions from several places in Scotland are acquired from MET-Office [15]. The period of the obtained data equates to 11 years which ranges from 2000 to 2010. The locations of each weather station are summarized in *Table 1.* The weather stations cover a small margin of latitude range but spans from east coast of Edinburgh to the western isle of Islay: Port Ellen. Only the Stornoway airport lies in the northern region which is in the Isle of Harris. The data are recorded with anemometers at a height of 10 meters above the ground with an hourly interval. These sets of data were rounded to the nearest ± 1 knot. Analysis are made from introducing additional reference(s) or target(s) in the MCP – PCA algorithm to all possible weather stations where each of them are treated as reference(s) or target(s) site. The two last years which are 2009 and 2010 from both the reference and target sites are used as a training period in which the algorithm learns the wind pattern between each other (i.e. wind speed and wind direction) in order to be used for predictions. In the analysis, the delay matrix covers the 48 period of hours with a truncation to 20 singular vectors. These are maintained throughout the analysis. The wind resource at each target site is predicted for the year 2008 until 2010 and then compared to the actual data from those years. The wind pattern from the 8 locations according to *Table 1* is distinguished from each other. Firstly, the one to one MCP – PCA method is carried out for the sites that have similar wind patterns. The first sets of tests undergo the MCP – PCA method with similar wind patterns as target and reference site. Then, an error measurement is carried out between the sites by comparing the predicted wind probability distribution function from the target site with the actual ones. The next sets of tests are repeated with sites that consist of different wind patterns between the reference and target sites. All of the predictions made by this method are also compared with the standard Linear – Regression method to prove its reliability. As mentioned, all the significant patterns from the sites are identified (i.e. prevailing winds and wind speed distributions). For the first case, reference sites are chosen from locations that has similar wind patterns to the target site. These are then compared with a standard MCP – PCA of one to one comparison from using one reference and one target site. For the second case, the chosen reference sites differ in wind dynamics from the target site. This means that the wind pattern between the reference sites are similar but differs from the target site. Similar error measurement is carried out for this second case. The third case features different wind dynamics between the reference sites themselves and the target site i.e., If three sites are used for the wind distribution prediction, these three must differ in wind dynamics from each other. Summary of the cases of using multiple references can be described in *Table 2*.

TABLE 2: Cases of using multiple references to predict the wind speed distribution at a target site.

Case	Wind pattern at reference sites	Wind pattern at the target site
1.	Similar to another reference and target site	Similar to reference sites
2.	Similar to each other but differs to target site	Differs from the reference sites
3.	Differs from each other and the target site	Differs from the reference sites

4.0 Major Outcomes

TABLE 3: *Average error distribution in all cases*

Cases	$e_{MCP-PCA}$	e_{LR}	PI
Single reference site to predict a target site (averaged between all the sites)	15.5 %	50.8 %	30.5%
Multiple reference sites to predict a target site (averaged between all the sites)	15.2 %	65.7 %	23.1 %

The summary of results in all sites averaged are given in *Table 3,* where the MCP – PCA method is clearly superior to the standard linear regression method. For one target to predict the wind distribution at one target site, the MCP – PCA method has decreased the prediction error to 30.5 % as indicated by the performance index, *PI*. In using multiple reference sites to predict the wind distribution at a target site, the MCP – PCA has significantly improved its performance in comparison to the standard linear regression method. It is shown that the error has been reduced to 23.1 % as indicated by the *PI* in this case. Even at its worst case (using multiple reference against one site that has different wind patterns from each other), the MCP – PCA yields an error, $e_{MCP-PCA}$ of only 19.2 %. In contrast to the best outcome of the standard linear regression in the best case (using multiple reference against one site that has similar wind patterns with each other), the standard linear regression method yields a staggering error of 61.2 %. Adding a reference to predict the wind behavior at a target site in MCP – PCA cases only improves the error made in average by a slight percentage of 0.3 %. These however differ from case to case. It is found out that in the 2nd case according to *Table 3,* the prediction error decreases by a noticeable amount if multiple reference is introduced. It yields a prediction error of 11.9 % in comparison to the one target to one reference which yields an error percentage of 18.5 %. However, in the 1st case, by introducing more reference sites with similar patterns the result worsen which yields an error percentage of 14.4 % in comparison to 11.7 % which belongs to one target against one reference site which has similar wind patterns as well. In the 3rd case of having both reference as well as target site(s) with different wind dynamics to predict the wind pattern at a target site, it is expected that the error percentage is the most with an error of 19.2 % yielded.

5.0 Results and Discussions

From all the cases and tests that are conducted, it can be concluded that the MCP – PCA method produces a significantly better error measurement value in comparison to the Linear – Regression method. It outperforms the Linear – Regression method in the 1st case (one reference against one target) by an average margin of 40.3 % in error value, in the 2nd case (multiple reference against one target) case by an average margin of 50. 5%. The MCP – PCA method finds the most significant pattern for the target's prediction whereby the Linear – Regression method takes an 'average' between all the sites.

In cases where adding references with similar wind patterns are used across all the reference and target sites, the error value surprisingly increases despite the fact that the wind patterns are similar. This is due to the fact that the algorithm having so much similar data that it truncates the possible outliers from the sites that contribute to the pattern. In other words, the algorithm is focused too much on the major patterns rather than both major and minor patterns that contributes to the pattern as a whole. In cases where introducing different wind patterns as references are used across all the sites, it is expected that the result worsens from every MCP – PCA tests. The significant pattern between the sites are less distinguishable for the MCP – PCA algorithm as these sites have unequal wind patterns between them (less correlation). As a result, the error value is increased as more sites are introduced. When the reference sites share the same wind patterns with each other but differs from the target site, a decrease of error value is observed as more sites are introduced. In this case, the difference in wind pattern is not completely different but at the same time not the same. Therefore, the most significant pattern for the target site can be distinguished and captured to produce a good prediction.

6.0 Conclusions

A wind resource assessment using MCP – PCA is carried out in this paper using data spanning from the year 2000 to 2010. The results obtained from the enhanced MCP clearly outperforms the standard linear regression method in all cases. In cases of adding multiple reference sites to enhance the performance of the method, a better selection criterion must be used in future studies to clearly capture the effect of reference site(s) to the error in predictions made. The next stage of work is to further verify the predictions made by making more tests i.e. using more pairings of data sets not only limited to one weather climate but focusing on other climates as well. The predictions reliability can also be further tested with the new and upcoming future data.

7.0 Acknowledgement
The author would like to thank UK Meteorological office for providing access to the British Atmospheric Data Centre.

References

[1] C. Skittides and W.-G. Fruh, "Wind Forecasting using Principal Component Analysis," *Renewable Energy,* no. 69, pp. 365-374, 2014.

[2] A. Foley, "Current Methods and Advances in Forecasting of Wind Power Generation," *Renewable Energy,* vol. 1, no. 37, pp. 1-8, 2012.

[3] S. Sun, H. Qiao, Y. Wei and S. and Wang, "A New Dynamic Integrated Approach for Wind Speed Forecasting," *Applied Energy,* no. 197, pp. 151-162, 2017.

[4] S. Velazquez, J. Carta and J. and Matias, "Comparison between ANNs and Linear MCP Algorithms in the Long-Term Estimation of the Cost per kWh Produced by a Wind Turbine at a Candidate Site: A Case Study in the Canary Islands," *Applied Energy,* vol. 11, no. 88, pp. 3869-3881, 2011.

[5] European Environment Agency, "Europe's Onshore and Offshore Wind Energy Potential - European Environment Agency," 2009. [Online]. Available: http://www.eea.europa.eu/publications/europes-onshore-and-offshore-wind-energy-potential. [Accessed 5 March 2015].

[6] J. Carta, S. Velazquez and P. Cabrera, "A Review of Measure-Correlate-Predict (MCP) Methods Used to Estimate Long-Term Wind Characteristics at a Target Site," *Renewable and Sustainable Energy Reviews,* vol. 0, no. 27, pp. 362-400, 2013.

[7] A. Dinler, "A New Low-Correlation MCP (Measure-Correlate-Predict) Method for Wind Energy Forecasting," *Energy,* no. 63, pp. 152-160, 2013.

[8] X. Wang, P. Guo and X. Huang, "A Review of Wind Power Forecasting Models," *Energy Procedia,* vol. 0, no. 12, pp. 770-778, 2011.

[9] W. Zhang, "Hybrid Wind Speed Forecasting Model Study Based on SSA and Intelligent Optimized Algorithm," *Abstract and Applied Analysis,* 2014.

[10] E. Erdem and J. Shi, "ARMA Based Approaches for Forecasting the Tuple of Wind Speed and Direction," *Applied Energy,* vol. 4, no. 88, pp. 1405-1414, 2011.

[11] O. Shukur and M. Lee, "Daily Wind Speed Forecasting through Hybrid KF-ANN Model Based on ARIMA," *Renewable Energy,* vol. 0, no. 76, pp. 637-647, 2015.

[12] M. De Giorgi, A. Ficarella and M. Tarantino, "Assessment of the Benefits of Numerical Weather Predictions in Wind Power Forecasting Based on Statistical Methods," *Energy,* vol. 7, no. 36, pp. 3968-3978, 2011.

[13] N. Golyandina and A. Korobeynikov, "Basic Singular Spectrum Analysis and Forecasting with R," *Computational Statistics and Data Analysis,* vol. 0, no. 71, pp. 934-954, 2014.

[14] C. Skittides and W.-G. Fruh, "A New Measure-Correlate-Predict Wind Resource Prediction Method," 2015.

[15] British Atmospheric Data Center, "MET Office - Midas Land and Marine Surface Station Data," 2011. [Online]. Available: http://catalogue.ceda.ac.uk.