Copula based Model for Wind Turbine Power Curve Outlier Rejection

Yue Wang, David G. Infield, Bruce Stephen and Stuart J. Galloway
Institute for Energy and Environment, Electrical and Electronic Engineering
University of Strathclyde, 204 George Street, Glasgow, G1 1XW

Abstract

Power curve measurements provide a conventional and effective means of assessing the performance of a wind turbine, both commercially and technically. Increasingly high wind penetration in power systems and offshore accessibility issues make it even more important to monitor the condition and performance of wind turbines based on timely and accurate wind speed and power measurements. Power curve data from Supervisory Control and Data Acquisition (SCADA) system records, however, often contain significant measurement deviations, which are commonly produced as a consequence of wind turbine operational transitions rather than stemming from physical degradation of the plant. Using such raw data for wind turbine condition monitoring purposes is thus likely to lead to high false alarm rates, which would make the actual fault detection unreliable and would potentially add unnecessarily to the costs of maintenance. To this end, this paper proposes a probabilistic method for excluding outliers, developed around a Copula-based joint probability model. This approach has the capability of capturing the complex nonlinear multivariate relationship between parameters, based on their univariate marginal distributions, through the use of a Copula; data points that deviate significantly from the consolidated power curve can then be removed depending on this derived joint probability distribution. After filtering the data in this manner, it is shown how the resulting power curves are better defined and less subject to uncertainty, whilst broadly retaining the dominant statistical characteristics. These improved

power curves make subsequent condition monitoring more effective in the reliable detection of faults.

Index term: wind turbine, power curve, outlier rejection, SCADA, Copula Model.

1. Introduction

Wind energy has an essential role in meeting electrical power demand in an environmentally sustainable manner. The considerable UK offshore wind resource and the need for the UK to reduce carbon emissions from electricity generation is driving policy to install 33GW of new wind generation capacity offshore by 2020 [1]. Costs and operational difficulties involved in offshore maintenance, resulting from poor plant accessibility, lead to a substantially lower turbine availability offshore than onshore where availability can be as high as 98% [2]. Consequently, preventative, condition based wind turbine maintenance is expected to be more economically efficient than responsive and/or scheduled maintenance.

Significant efforts have been made to develop condition monitoring of wind turbines. The applied techniques and algorithms can be classified into two general categories, i.e. physical model based methods and data driven algorithms. The development of a physical model for a specific turbine component requires detailed physics which is not always readily achievable [3]. In contrast, data driven methods can facilitate the required analysis with the aid of artificial intelligence techniques. A physical model based application to wind turbine gearbox fault detection using the Physics of Failure methodology is presented in reference [3]. Reference [4] implements the condition monitoring of gearbox based on signals from both the SCADA system and a conventional vibration based Condition Monitoring System (CMS). Wavelet transforms, which are capable of providing good frequency resolution at low frequencies and time resolution at high frequencies, have been applied to measurements from the generator for condition monitoring purpose [5]. References [6, 7] employ the commonly

used Artificial Neural Networks (ANN) to construct the normal behaviour modelling based on the SCADA data, which is then used to detect the anomalies in the corresponding turbine subcomponent.

A fundamental but important metric for monitoring wind turbine performance is the power curve which relates turbine power output to the wind speed experienced by the turbine rotor. Data suitable for power curve determination are generally available from the SCADA systems installed with most modern wind turbines. The SCADA system logs general turbine operational and meteorological data in a 10-minute averaged form for each individual wind turbine and any meteorological masts within the wind farm, before communicating them to a remote central computer [8]. SCADA system errors, such as those from errors in the communications system [9], or measurement sensor errors can result in data loss and hence null entries in SCADA records. Other spurious measurements that can deviate significantly from the power curve supplied by the original equipment manufacturer (OEM) may be caused by the 10-minute averaging period used in power curve determination, including a mixture of normal turbine operation and a period of non-production when the turbine has been stopped by the control system for one reason of another. Turbine start/stop decisions do not necessarily coincide with the bounds of the ten minute averaging periods. These measurements (referred to as outliers in the rest of this paper) are facets of the data collection and are not an indication of faults or anomalous turbine operation. Such misleading data should therefore be removed before further analyses are undertaken. Both references [10] and [11] claim the necessity of elimination of the power curve outliers, where Kusiak et al. acknowledge that outliers do exist in power curve measurements and they will affect the accuracy of the associated analysis, and they employ an analysis of residuals together with control charts to filter potential outliers. Identification of blade and yaw system faults based on the monitoring of power curves are documented in reference [12]; however the authors

removed the outliers by visual inspection, which is neither accurate nor efficient. The authors of reference [11] mention the difficulties in checking the data validity and deciding between normal and anomalous turbine operation. The main target of the work presented here is to reject potential outliers whilst broadly retaining the statistical characteristics of the power curve, in particular the mean values of the measurements. Despite its simplicity such an approach to condition monitoring is relatively unexplored.

IEC standard 61400-12-1 [13] specifies the use of the 'method of bins' to form the power curve. SCADA data are grouped and averaged in 0.5 $\mathrm{m/s}$ wind speed bins, with uncertainties (due to both data measurements and sensors) being illustrated by error bars for each bin. The method provides a simple and straightforward way to determine and present a wind turbine power curve, but its accuracy and reliability depends largely on data quantity and their intrinsic spread: a bin with a smaller number of data points will give, all other factors being equal, a greater uncertainty. The nonparametric k-nearest neighbour (kNN) algorithm is utilized in reference [14] to construct a reference power curve for an individual turbine based on exemplar training data, and in [10], the same algorithm has been extended in order to construct a reference power curve for a whole wind farm. In this extended use, Principal Component Analysis, [15], is used in combination with the kNN algorithm to reduce the dimension of the input data by selecting only the most informative wind speed components. The drawback of kNN is not only the computational burden involved, which becomes significant for large training data sets, as implied in [10], but also the fact that the original data is transformed by the k-averaging process, which replaces the original data with the averaged value from the k nearest points. Reference [16] treats the relationship between wind speed and power output across the wind farm as stochastic and develops a probability distribution of wind farm power generation in terms of wind speed and wind direction, based on conditional kernel density estimation [17]. The resulting distribution could be used by

power system operators to model expected power production. Reference [18] also presents a wind farm power predictive distribution based on ensemble probabilistic forecasting. The forecasting method converts meteorological variables into power by using a fitted power curve model, requiring greater accuracy in the power curve measurements.

One effective way of representing the relationship between these two power curve variables, and the corresponding uncertainty, is to investigate their joint probability distribution. The joint probability density of power curve measurements represents a highly non-linear bivariate relationship and is difficult to represent using the common parametric multivariate probability distributions. Copulas provide a means of relating variables with a complex dependency structure. They have been extensively used to solve economic and financial problems [19], where the underlying data show significantly nonlinear features. This paper uses a Gaussian Mixture Copula Model (GMCM), [20], to construct a 2-dimensional joint probability distribution for wind speed and wind turbine power output that can be used to represent a power curve. The model proposed in this paper builds on the ones presented in references [12] and [21]: Gill et al. present a non-parametric approach that serves only to quantify the ability of the Copula to approximate the functional form of the power curve; Stephen et al. propose the use of parametric marginal distributions to be used with Copulas both to approximate the functional form of the power curve as well as to identify particular operating regimes within it as multimodal behaviour. Where the GMCM contributes further is in the unification of both strands of the preceding research: the quantification of anomalous behaviour through low likelihood and the development of parametric models capturing the various modes within the power curve that do not necessarily follow linear Gaussian multivariate distribution assumptions. Reference [22] claims that the kNN is more complex and requires more memory for computation than the GMM model. And the GMCM, which is

based on the Gaussian Mixture Model (GMM) and involves the same parameters as used in GMM, can be assumed to have almost the same computational complexity as the GMM.

The Copulas used as the foundation for the development of power curve model, including the Frank Copula and the GMCM, are introduced in the next section. Particular attention is paid to comparisons of fitness of different Copula models in Section 3, and outlier rejection in Section 4.

2. Expressing Dependency through Copula Statistics

The term Copula was first employed by Sklar to bring together the complex nonlinear dependency structure of a multivariate data set with its one-dimensional marginal distributions [23]. For a set of n marginal probability densities, the n-dimensional joint PDF can be expressed as:

$$f(x_1, x_2, \cdots, x_n) = c(u_1, u_2, \cdots u_n) \times f_1(x_1) \times f_2(x_2) \times \cdots f_n(x_n) \quad (1)$$

where f_i denotes the ith marginal PDF and the corresponding marginal CDF, F_i, is represented by u_i. And c represents the Copula density function that unifies them.

2.1 The Frank Copula

The choice of Copula is governed by the tail dependence implied by the data. In the bivariate case, tail dependence is expressed in terms of the relationship between the extreme values of the two marginal distributions. If one variable has exceeded a particular threshold and the other has also exceeded this threshold with proportional likelihood, then the distribution is tail dependent [24]. Tail dependence can be visualised as a tightening of the scatter of observations around the extremes of the distribution, while low tail dependence will be exhibited as a greater degree of scatter. As has been shown in reference [12], the distribution pattern and the characteristics of the tail dependency of the Frank Copula Model are

consistent with those of the power curve variables, for which reason this particular Copula is selected here. The Frank Copula density function, $c_{Frank}(u_1, u_2, \delta)$, is given by

$$c_{Frank}(u_1, u_2, \delta) = \frac{\delta\eta e^{-\delta(u_1+u_2)}}{[\eta-(1-e^{-\delta u_1})(1-e^{-\delta u_2})]^2} \quad (2)$$

where $\eta = 1 - e^{-\delta}$ [25]. As shown in Figure 1, the larger the value of δ, the stronger the dependence is between the variables related by the Frank Copula [25] throughout their bivariate distribution. The parameter δ can be obtained by optimizing the model's fit for a given bivariate dataset based on some criteria such as maximum likelihood.

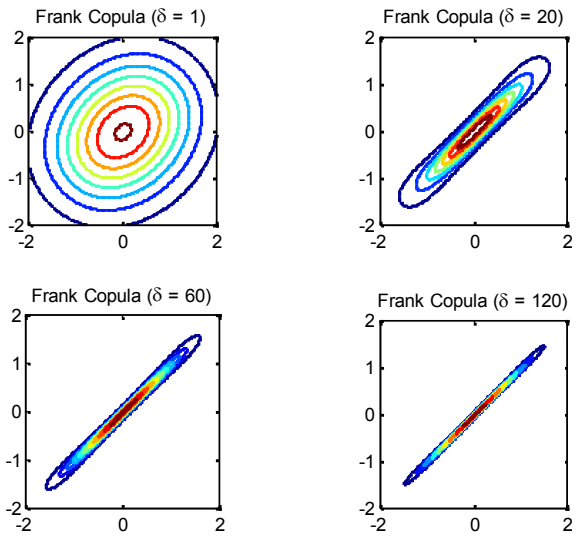

Figure 1: Bivariate distribution with Gaussian marginals demonstrating the effect of δ value on variable dependency

2.2 The Gaussian Mixture Copula Model (GMCM)

Accommodating the complex shape of the power curve joint probability density is beyond the abilities of classical multivariate distributions and would necessitate a mixture model with a large number of parameters, running the risk of over-fitting and losing generalisation

capabilities; where the Copula mixture adds benefit is in identifying the modes without requiring a large number of components to represent their dependency structure.

The GMM probability density function φ, comprises a weighted sum of M Gaussian density components, given by

$$\varphi(x_1, x_2, \cdots, x_n; \Theta) = \sum_{j=1}^{M} \alpha_j N\left(x_1, x_2, \cdots, x_n; \theta_j\right) \quad (3)$$

where α_j are the weights for different components and all the elements of α_j sum to unity. Parameter M indicates the modality number and will be determined in Section 3.2. $_j = \{\mu_j; \Sigma_j\}$ with μ_j representing the mean vector and $_j$ being the covariance matrix for the j^{th} component [26]. And the parameter set, Θ, combines the weight assignment and the statistics in $_j$ for each Gaussian component. Multivariate Gaussian distributions can only express linear dependency, and while the mixture model framework may afford a piecewise approximation of non-linearity, it is therefore clear that mixture components with a more complex dependency structure would allow a superior fit.

A Gaussian Mixture Copula Model (GMCM), [20], derived from a GMM with no implied covariance is capable of characterising multidimensional nonlinear statistics for multimodal data. The GMCM density function, derived from the GMM expression of Equation (3), is defined as:

$$c_{GMCM}(u_1, u_2, \cdots, u_n; \Theta) = \frac{\varphi\left(\Phi_1^{-1}(u_1), \Phi_2^{-1}(u_2), \cdots, \Phi_n^{-1}(u_n); \Theta\right)}{\prod_{i=1}^{n} \varphi_i(\Phi_i^{-1}(u_i))} \quad (4)$$

where φ_i and Φ_i^{-1} denote the marginal density of GMM and the corresponding inverse distribution along each dimension. The parameter set Θ is optimised by maximising the log-likelihood function of the GMCM Copula function as shown in Equation (4).

Equation (1) is used to calculate the joint probability distribution based on the fitted Copula density function: Equation (2) for the Frank Copula model; and Equation (4) for GMCM, with the marginal PDF for each variable in Equation (1) is achieved through kernel density estimation.

3. Power curve density modelling with Copulas

In the specific application to wind turbine power curve analysis, the Copula model links the marginal distribution of wind speed and turbine power output to their two-dimensional joint probability density function.

The basic steps for Copula based outlier removal are as follows:

1) Pre-processing of power curve measurements

   This includes the removal of null entries followed by air density correction of the raw data as will be presented in Section 3.1.

2) Model order determination

   The modality number is derived using the self-organising map in Section 3.2 to facilitate the fitting of the GMCM.

3) GMCM fitting

   In [20] a GMCM parameter optimisation process is proposed that is based on Expectation Maximization (EM) [27] followed by application of a gradient descent optimisation [28]. The reason for this is the non-convex form of the log-likelihood function for the GMCM density function. The solution obtained from the Maximisation step of EM is not guaranteed to find the global optimum, thus necessitating the use of the Gradient Descent algorithm with randomly assigned initial

conditions within an iterative loop for global optimum investigation. This methodology for GMCM parameter estimation is retained in this paper.

4) Outlier rejection

Based on the achieved density distribution, the outliers of power curve measurements are filtered using a probability contour that will be determined in Section 4.

The robustness of GMCM is shown by comparing goodness of fit between GMCM, the Frank Copula, and GMM, using the Bayesian Information Criterion (BIC). Two examples are shown in Section 4 to validate the effectiveness of this outlier elimination method.

3.1 Typical power curve and data pre-processing

Measurements retrieved from the SCADA system consist of 10-minute averaged values of wind speed, turbine power output, ambient temperature and atmospheric pressure. The latter two measurements determine the air density, ρ, to which the turbine power output P is proportional:

$$P = \frac{1}{2}\rho\pi R^2 v^3 C_P \quad (5)$$

where R represents the radius of turbine rotor; v is the wind speed experienced by the rotor and C_P indicates the power coefficient. In order to correct the operational data to standard air density conditions (15 degree Celsius and 101.325 kPa), the acquired power curve measurements are modified following the procedure described in IEC standard 61400-12-1, [13]. Null entries, for wind speed, power or both, can arise in the data record due to a breakdown in data capture, either in the sensors or the communication system of the SCADA system. These should be removed prior to probability density function fitting. All the power curve measurements used in this paper have been corrected for air density and empty entries have been eliminated as described.

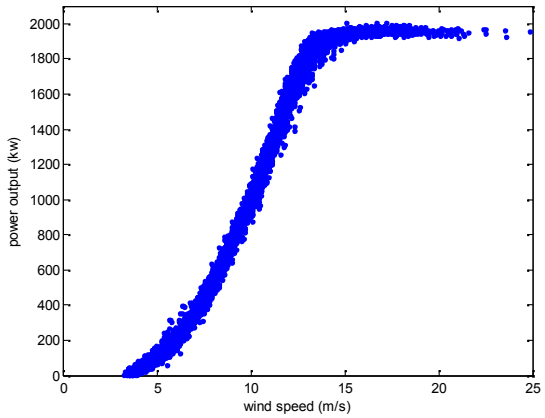

Figure 2: Scatter plot of power curve measurements

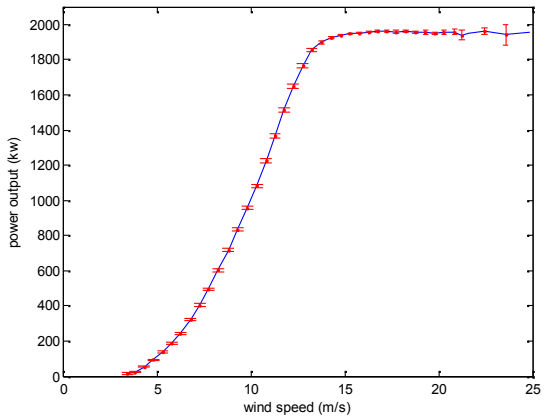

Figure 3: Power curve with error bars showing data uncertainty

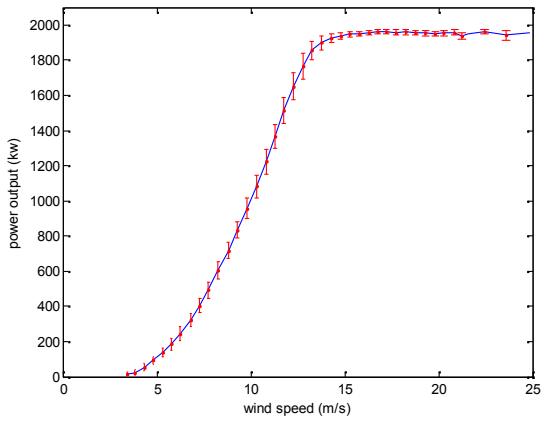

Figure 4: Power curve with error bars showing data dispersion

Figure 2 shows power curve measurements for a two-month period (depicting 7257 pairs of data) of fault free operation from a pitch regulated variable speed wind turbine (hereafter referred to as 'turbine 1') with a nominal rating of 2MW. The corresponding power curve, produced by binning as outlined in section 1, is shown in Figure 3, where the error bars plotted have been calculated from $\frac{\sigma}{\sqrt{K}}$, with σ representing the standard deviation of power output values in the bin and K being the number of points. The term $\frac{1}{\sqrt{K}}$ results from the Central Limit Theorem's measure of uncertainty: error bars calculated in this way give an indication of the confidence in the expected value of the power curve at that point based on the number of observations in the bin. The relatively large error bars at the high wind speed (values over 20 $\mathrm{m/s}$) bins are not important here because it is known that the maximum power generated is well controlled and determined by the turbine control system, [29], as shown in Figure 2. These are due to the insufficient numbers of points in these particular bins reflecting the occasional nature of the very high wind speeds. For the purposes of this paper it is the spread of data that is of more importance in the Copula fitting and thus the power curve has been re-plotted in Figure 4 to show errors bars with a value of the unmodified σ. Note that the largest values of σ occur around and just below the rated wind speed of 13.5 $\mathrm{m/s}$ for reasons that will be discussed in Section 4.

3.2 Model order selection

The optimal data modality is required when using mixture models such as GMCM. While many power curves have three distinct modes, the methodology proposed in this paper is an inherently data driven one: operational, faults or meteorological factors may result in a curve that has a different number of modes, for examples due to anemometer failure [11] or de-rating of the turbine (without a corresponding flag in the SCADA records). The self-organising map (SOM), originally conceived by Teuvo Kohonen, [30], is employed here

because of its ability to cluster the data in an unsupervised-learning manner. The main function of SOM is to construct a nonlinear projection of high-dimensional data onto a low-dimensional (usually 2D) space, in which the clustering of data and its topology are clearly shown and easily interpreted [30].

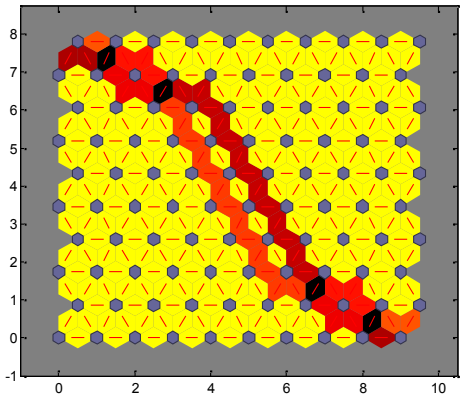

Figure 5: SOFM neighbour weight distances

The data set shown in figure 2 is used to determine the number of modes present in the data which will in turn inform the choice of modality for the optimal model. Three distinct data regions can be observed in this figure: near cut-in, below which the turbine does not operate (3.5 m/s in this example); above rated (13.5 m/s); and the region in between. A 10×8 two-dimensional SOM is used to visualise the data clustering. The learning result is presented in the form of neighbour weight distances as illustrated in Figure 5, where the blue hexagons represent the neurons and neighbouring neurons are connected by red lines. The background colour indicates the distances between the neurons, with darker colours representing larger distances and lighter colours representing smaller distances. Three segments can be observed based on the colour coding scheme introduced: two distinct triangles at bottom left and top right; and a relatively weak segment located approximately along the diagonal. They are separated from each other by dark colour bands. The model order for this power curve data is three, corresponding to the number of distinct regions into which the space is divided, which

matches the original assumption of three parts to the power curve, although of course the plot of weight distances does not directly reproduce the original data.

3.3 Fitness analysis

GMCM is capable of clustering data automatically once the data modality has been identified, as described in the previous section. The same SCADA data as used in Section 3.2 are used here to assess the model's fitness. The Bayesian Information Criterion (BIC), [31], is used here for model selection, with lower values of BIC indicating better models. It is based on the log-likelihood function, $L(\Theta|x_1, x_2, \cdots, x_n)$, sums the log of the probabilities of all data points and provides a convenient and easily calculated metric for goodness of fit, [32]. Over-fitting is avoided by introducing a penalty term, $plog(N)$, which takes account of the model complexity. BIC is defined as:

$$BIC = -2L(\Theta|x_1, x_2, \cdots, x_n) + plog(N) \quad (6)$$

where $L(\Theta|x_1, x_2, \cdots, x_n) = \sum_{i=1}^{N} \log(f(x_1(i), x_2(i), \cdots, x_n(i)))$ (7)

N represents the sample size in both Equations (6) and (7), and has the value of 7257 in this case. And p is the number of parameters. For the Frank Copula $p = 1$ whilst for the GMM and GMCM, it can be calculated using Equation (8), [33].

$$p = m(1 + d + \frac{d(d+1)}{2}) \quad (8)$$

where m denotes the modality number, which is 3, as determined in Section 3.2 and d indicates the data dimension, which is 2 in this paper. This results in a p value of 18.

The different models (GMCM and Frank Copula) can be compared by calculating BIC for identical input data samples. Figures 6(a), 6(b) and 6(c) illustrate the probability density fitting for the GMCM models, Frank Copula and the GMM model respectively. The GMM is

included here due to its capability of dealing with multimodal data, as summarised in Section 2.2. The BIC values of these three models are listed in Table 1, from which it can be seen that the GMCM model outperforms the other models. The GMCM also has the advantage of dealing with multivariate distributions, which would readily accommodate more variables for further applications, whereas the Frank (or Archimedean) Copula could only be used for bivariate data characterisation. In conclusion, the Gaussian Mixture Copula Model is thus chosen for outlier rejection.

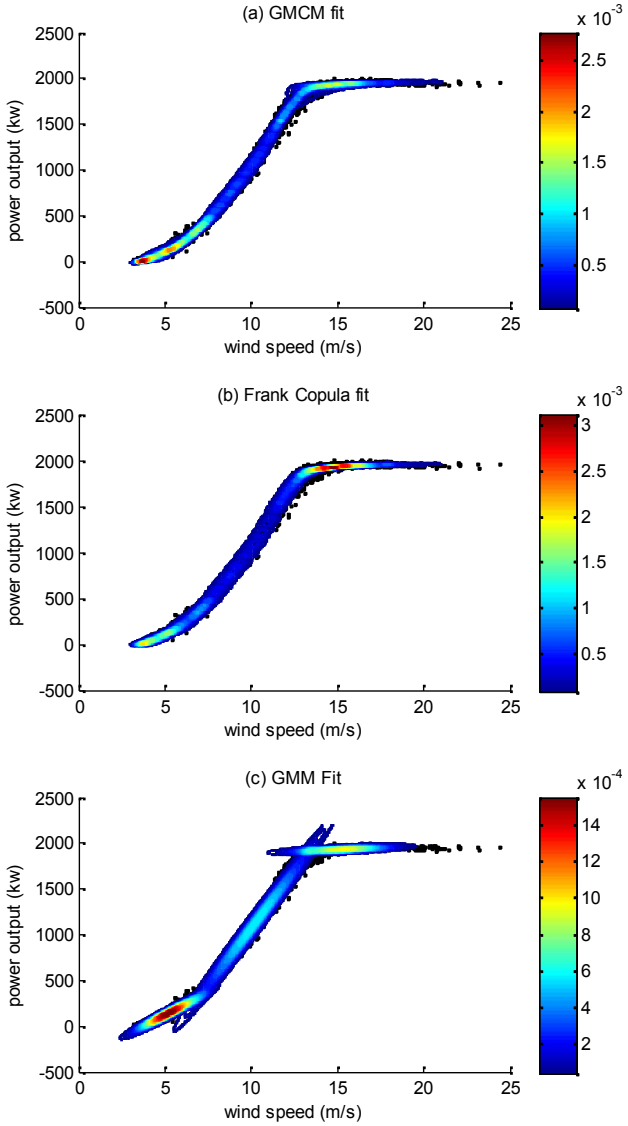

Figure 6: Fitness comparison of three presented models

4. Outlier rejection using GMCM

Power curve deviations (scatter) can result from SCADA system measurement or data transcription errors, anomalous values from measurement intervals that include periods when the turbine is not operating, and finally from anomalies in turbine operation due to some fault or other that is of interest for condition monitoring purposes. Null entries caused by SCADA system issues are eliminated as described in Section 3.1; intermittent operation within the 10 minute measurement time interval will lead to anomalous values that have a random character. This leaves systematic deviations that reflect actual problems in turbine operation or at least in instrumentation.

It is the purpose of the analysis presented here to identify statistical trends associated with these deviations from ideal turbine behaviour. Unambiguously distinguishing between these different situations is not readily achieved without domain expert knowledge. In the absence of this, Copula fitting methods developed can be used to eliminate outliers. Since the joint probability distribution provides a straightforward means of data characterisation, a probability density-based deviation exclusion method should also be effective in the elimination of possible outliers whilst retaining the broad statistical characteristics of the power curve.

For a modern pitch regulated variable speed wind turbine, good power control is available above the rated wind speed. It is shown in Figure 4 that the greatest scatter, as indicated by the error bars of size , occurs at around rated power where the turbine is continually changing between below rated operation where speed is varied to maximise aerodynamic efficiency, and above rated power where electronic control limits current and power from the

generator [34]. The lower variance at the extremes means that the tail dependency is not likely to be a major source of error.

A probability contour level at three standard deviations for data in the 0.5 m/s wind speed bin closest to rated wind speed is judged to be appropriate. Points lying outside this contour are regarded as outliers and are eliminated. The effectiveness of this proposed method is demonstrated on two additional turbines (denoted as turbines 2 and 3), both being pitch regulated.

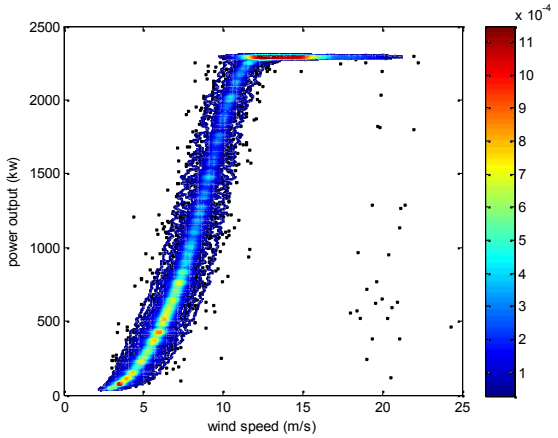

Figure 7: GMCM fit of power curve measurements for turbine 2

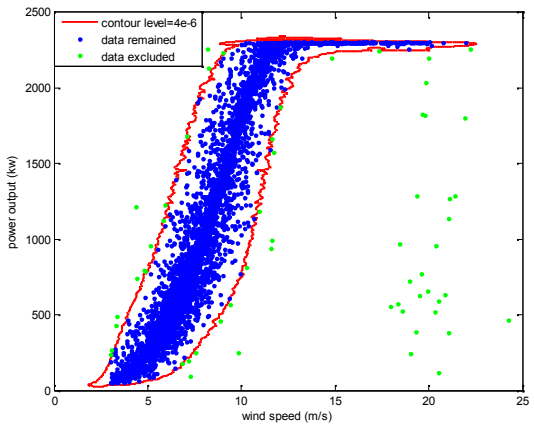

Figure 8: Data exclusion for turbine 2 using density contour

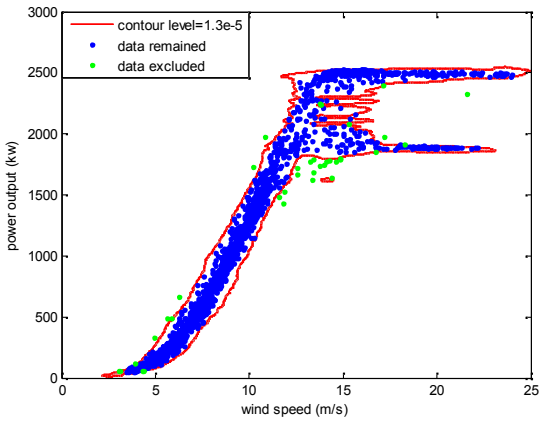

Figure 9: Data exclusion for turbine 3 using density contour

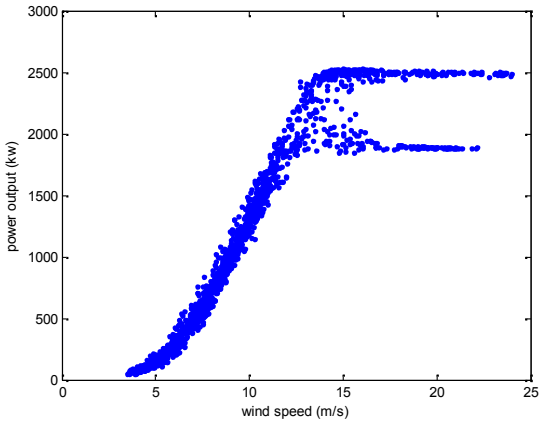

Figure 10: Power curve for turbine 3 after cleaning

Figure 7 illustrates the GMCM fitting of power curve measurement for turbine 2. Figure 8 shows the same power curves after outliers have been identified using the fitted Copulas and the density contour (defined as 3) illustrated by the red line, with green points indicating power curve measurements that are to be excluded. It can be seen from the below rated region of Figure 8 that the turbine control strategy for tracking the maximum C_p shows unsatisfactory performance in that the power curve measurements do not tightly align with the desired trajectory, which is unaffected by the data removal process. Also in figure 8, the data points outside the probability contour and above 15 m/s are firmly believed to be the result of the turbine cutting out for some of the ten minute averaging time, thus giving an

average power somewhere between rated power and zero, depending on how much time the turbine was not operational. It is clearly correct that these data be excluded as they do not reflect the relationship between power and wind speed, but rather the anomalous effect of averaging. Cleaning the power curve through the elimination of outliers makes the operation of turbine 3 clearer, as shown in Figures 9 and 10. Two distinct power levels can be discerned in Figure 10, where the lower level at 2000kW output (even in high winds) results from the turbine being derated. This occurs when the turbine operators are instructed to reduce wind farm output by the network operators, usually due to excess wind.

The power curve cleaning has been demonstrated to be effective in both examples. Subsequent condition monitoring based on these improved power curves will thus result in more reliable detection of faults.

5. Conclusions

Power curves are an established metric for wind turbine performance and have previously been demonstrated to be useful measures of plant condition when constructed from operational monitoring data. To date, the elimination of power curve outliers remains relatively unexplored. This paper has proposed the use of a model capable of modelling the complex stochastic dependency structure inherent in the power curve to allow probabilistic filtering of measurement data as a pre-processing stage to a condition model. Providing an appropriate model order has been identified, a GMCM can capture the complex nonlinear dependency structure between wind speed and power output measurements and can be used to estimate the power curve to a level of accuracy that cannot be matched by parametric multivariate distributions, with limited computational complexity. The probability density-based approach set out in this paper for outlier rejection has been demonstrated to effectively remove the significant outliers whilst retaining the main statistical characteristics of the

power curve measurements. Pre-processing of the power curve will improve the effectiveness of techniques based on power curve anomalies that are increasingly popular for condition monitoring and fault identification in wind turbines. Future work will involve online wind turbine performance assessment based on power curve measurements screened using the presented method, where the GMCM could be applied to the updated power curve measurements on a regular basis, say each month, to take account of any evolution in the power curve. The GMCM could also be adapted to take account of additional variables, such as the wind direction.

Acknowledgements

The authors wish to thank the Energy Technologies Institute for making the wind turbine power curve data available from its Offshore Wind Condition Monitoring Project, and for the support of Yue Wang's PhD research.

References

[1] BWEA, *Actions for 33GW* [online]. Available: http://www.bwea.com/pdf/publications/33GW_08.pdf (accessed 30th July 2012).

[2] E. J.Wiggelinkhuizen et al., *Conmow—Final Report Energy Research Center of the Netherlands*, 2007.

[3] C. S. Gray, and S. J. Watson, *Physics of Failure approach to wind turbine condition based maintenance*, Wind Energy, 13: 395–405. doi: 10.1002/we.360, 2010.

[4] Y. Feng, Y.Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, *Monitoring wind turbine gearboxes*, Wind Energy. doi: 10.1002/we.1521, 2012.

[5] S. Watson , B. Xiang , W. Yang , P. Tavner and C. Crabtree, *Condition monitoring of the power output of wind turbine generators using wavelets*, IEEE Trans. Energy Convers, vol. 25, no. 3, pp. 715 -721, 2010.

[6] A. Zaher, S.D.J. McArthur, D.G. Infield, and Y. Patel, *Online wind turbine fault detection through automated SCADA data analysis*, Wind Energy, 12: 574–593. doi: 10.1002/we.319, 2009.

[7] M.C. Garcia, M.A. Sanz-Bobi, and J.D. Pico, *SIMAP: Intelligent System for Predictive Maintenance: Application to Health Condition Monitoring of a Wind Turbine Gearbox*, Computers in Industry, Aug. 2006, 6, 57, pp.552-568.

[8] W. F. Young, J. E. Stamp and J. D. Dillinger, *Communication Vulnerabilities and Mitigations in Wind Power SCADA Systems*, American Wind Energy Association WINDPOWER 2003 Conference, Texas, May 2003.

[9] *Supervisory Control and Data Acquisition (SCADA) Systems*, Technical Information Bulletin 04-1, Communication Technologies Inc., National Communications System, Oct. 2004.

[10] A. Kusiak, H. Y. Zheng and Z. Song, *Models for Monitoring Wind Farm Power*, Renewable Energy, vol. 34, pp. 583-590, 2009.

[11] Y. Wan, E. Ela and K. Orwig, *Development of an Equivalent Wind Plant Power Curve*, NREL/CP-550-48146, WindPower Conference, 2010.

[12] S. Gill, B. Stephen, S. Galloway, *Wind Turbine Condition Assessment Through Power Curve Copula Modelling*, IEEE Transactions on Sustainable Energy, vol. 3, no. 1, pp. 94-101, Jan. 2012.

[13] *Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines*, British Standard, IEC 61400-12-1, 2006.

[14] A. Kusiak, H. Zheng and Z. Song, *On-line Monitoring of Power Curves*, Renewable Energy, vol. 34, pp.1487-1493, 2009.

[15] I.T. Jolliffe, *Principal Components Analysis*, 2nd Edition. Springer Series in Statistics. 2002.

[16] J. Jeon, and J. Taylor, *Using Conditional Kernel Density Estimation for Wind Power Density Forecasting*, Journal of the American Statistical Association, to appear, 2011.

[17] R. Duda, P. Hart and D. Stork, *Pattern Classification*, 2nd ed., John Wiley and Sons, New York, 2001.

[18] P. Pinson, and H. Madsen, *Ensemble-based Probabilistic Forecasting at Horns Rev*, Wind Energy, Vol. 12, Issue 2, pp. 137–155 ,2009.

[19] R. T. Clemen and T. Reilly, *Correlations and Copulas for Decision and Risk Analysis*, Manage. Sci., vol. 45, pp. 208–224, Feb.1999.

[20] A. Tewari, M.J. Giering and A. Raghunathan, *Parametric Characterization of Multimodal Distributions with Non-gaussian Modes*, 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), vol. 11, no. 11, pp. 286-292, Dec. 2011.

[21] B. Stephen, S. J. Galloway, D. McMillan, D. C. Hill, and D. G. Infield, *A Copula Model of Wind Turbine Performance*, IEEE Transactions on Power Systems, vol. 26, no. 2, pp. 965–966, May 2011.

[22] M. Shi and A. Bermak, *An Efficient Digital VLSI Implementation of Gaussian Mixture Models-Based Classifier*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol.14, no.9, pp.962-974, Sep. 2006.

[23] R. B. Nelsen, *An Introduction to Copulas*, ser. Springer Series in Statistics, 2nd ed., New York: Springer, 2006.

[24] R. Schmidt, *Tail Dependence*, In Statistical tools in finance and insurance (eds P.Čížek, W.Härdle & R.Weron), 65–91. Springer Verlag, Heidelberg, 2005.

[25] H. Joe, *Multivariate Models and Dependence Concepts*, 1st ed., New York: Chapman & Hall, pp. 139-166, 1997.

[26] G. J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics, pp. 80-85, Nov. 2000.

[27] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, 1997.

[28] Tewari Copula Mixture Model code [online]. Available: http://code.google.com/p/copula-mixture-model/downloads/list (accessed 20th May 2012).

[29] J. F. Manwell, J. G. McGowan and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application*, 2nd edtion, John Wiley & Sons, pp.397-399, 2010.

[30] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Sciences, 3rd edition, Springer, Berlin, 2001.

[31] S. Konishi and G. Kitagawa, *Information Criteria and Statistical Modeling*, Springer New York, 2008.

[32] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Wiley, New York, pp. 7-10, 1989.

[33] W.D. Penny, *Variational Bayes for d-dimensional Gaussian Mixture Models*, University College London, July 2001.

[34] A. D. Hansen, C. Jauch, and P. Sorensen, *Dynamic Wind Turbine Models in Power System Simulation Tool DIgSILENT*, Risø-R-1400(EN), National Laboratory, 2003.

	Bayesian Information Criterion value
GMCM	110597
Frank	112415
GMM	114993

Table 1: BIC values of three models indicating goodness of fit