

# Wind Turbine Condition Assessment through Power Curve Copula Modeling

Simon Gill, Bruce Stephen, *Member IEEE*, & Stuart Galloway

**Abstract**— Power curves constructed from wind speed and active power output measurements provide an established method of analyzing wind turbine performance. In this paper it is proposed that operational data from wind turbines are used to estimate bivariate probability distribution functions representing the power curve of existing turbines so that deviations from expected behavior can be detected. Owing to the complex form of dependency between active power and wind speed, which no classical parameterized distribution can approximate, the application of empirical copulas is proposed; the statistical theory of copulas allows the distribution form of marginal distributions of wind speed and power to be expressed separately from information about the dependency between them. Copula analysis is discussed in terms of its likely usefulness in wind turbine condition monitoring, particularly in early recognition of incipient faults such as blade degradation, yaw and pitch errors.

**Index Terms**— Wind power generation, Energy conversion, Power generation reliability.

## I. INTRODUCTION

WIND power based renewable energy has seen dramatic growth over the past decade, which is set to continue as many countries implement stringent targets for sustainability and emissions reduction. The UK government is aiming to generate 20% of the country's electricity from wind turbines by 2020, up from 3% in 2008 [1] to form part of the European Union's target of producing 20% of all energy from renewable sources [2]. A supporting trend in wind generation is the move towards larger turbines in offshore locations. Notable examples include Danish plans to double offshore wind capacity from 661MW to 1256MW over the coming four years [3] and plans for 25 GW of wind generation as part UK round 3 offshore sites [4].

Maintaining profitability with large offshore wind farms poses a significant challenge as operation and maintenance costs are significantly higher compared to those onshore, whilst turbine availability is significantly less. Onshore turbine availability can be maintained at levels up to 98% [5]; in comparison an offshore farm in the South of England – Scroby Sands – published average availability of 83% for 2007 [6]. Operation

Manuscript received May X, 2011. Dr. B. Stephen is a Senior Research Fellow in the Advanced Electrical Systems Research Group, Institute of Energy and Environment, University of Strathclyde, Glasgow, G1 1XW (phone: +44 (0)141 548 5864, e-mail: [bstephen@eee.strath.ac.uk](mailto:bstephen@eee.strath.ac.uk)). This work was supported in part by the EPSRC through the Wind Energy Doctoral Training Centre, University of Strathclyde.

and maintenance (O&M) costs are also significantly higher offshore: one estimate suggested O&M costs account for approximately 10% of onshore total expenditure, offshore this rises to 30% [7]. Initial work to translate onshore downtime data to the offshore environment suggests that failure downtime can be several multiples of the onshore average. This stems from access issues and potentially limited opportunities for windows in sea and wind conditions to return the turbine to service. Failure rate in onshore turbines can be as high as 6 failures per year per turbines for turbines with capacity greater than 1MW, although most studies find rates between 1 and 3 per year [8-9]. It is suggested in [10] that failure rates of less than 0.5 per turbine per year are a likely requirement of offshore operation.

One means of reducing the cost of maintenance and the effect on availability is to implement condition monitoring and preventative maintenance strategies on machines. These techniques have matured for a range of rotating machines [11], however they have only recently been applied to the specific technical challenges of wind turbines. Wind turbines are typically constructed from a number of subassemblies. Studies of faults in these subassemblies give an idea of the importance of monitoring and maintenance of specific turbine components. Systems where the fault frequency is highest include the electrical system, rotor, converter and yaw systems [10]. Gearbox and generator failure, whilst not the most common failure, do cause the greatest downtime per failure; it has been estimated that a generator fault in an onshore turbine can lead to 7 days downtime on average [9].

The power curve is an important metric of the performance of a wind turbine; it relates the power output to the wind speed. Traditionally, a power curves provide an expected relationship under standard operating conditions, for example with turbulence maintained within specified limits and air-density corrections applied. Measurement for wind speed and power must be made by following international specifications laid down in standard IEC 61400 [12]. Power curves are often used by manufactures as part of the technical specification of a turbine, possibly a performance guarantee.

Power curves can be generated without the need for retrofitted telemetry from operational SCADA data by utilizing wind speeds measured by nacelle mounted anemometers [13]. The variability of nacelle wind speed relative to free wind speed makes it difficult to compare operational data with Original Equipment Manufacturer (OEM) power curves. However, data taken when a particular turbine is assumed operating correctly

allows future performance to be benchmarked against this data accounting for local variables such as turbulence and turbine-wake interactions. This suggests that a probabilistic model may prove useful; such a representation of wind speed and power would allow probabilities for likelihood of pairs of nacelle wind speed and power output to be computed.

The contribution of this paper is to develop a probabilistic model of a power curve for condition monitoring purposes based on copulas; modeling such a relationship is challenging as the form of dependency between wind speed and active power is clearly non-linear; copula statistics are used to deal with the complexity of the relationship. In the section that follows, condition monitoring techniques specific to wind turbines are reviewed. Following this, Copula statistics are introduced with their use illustrated with examples of their application in other domains. In section IV, the application of Copulas to modeling the power curve of a wind turbine is developed. This unobtrusive, economic means of monitoring the condition of the plant is the key contribution of this paper and its practical use is demonstrated on SCADA data taken from a fleet of operational wind turbines.

## II. WIND TURBINE CONDITION MONITORING

Condition monitoring telemetry may be expensive or impractical to retrofit so most modern wind turbines integrate some condition monitoring systems into their design. These may include sensors for drive train vibrations, analysis of particulates in gearbox oils and blade strain gauges [14]. A review of commercially available condition monitoring systems is provided in [15].

Numerous methods have been employed to carry out system-wide monitoring. Neural networks are the most common and have been used in a number of studies to ‘data-mine’ relationships between variables recorded in SCADA data and faults in the wind turbines [18]. SCADA data typically includes active power output, wind speeds from nacelle mounted anemometers, component temperatures, electrical currents and power factors all with a frequency of 5 to 10 minutes. Data-mining algorithms can be used together with historic SCADA data and fault information to estimate fault likelihoods [19]. Data from neighboring turbines can be combined to estimate the expected operation of one particular turbine, this can be used to spot ‘soft’ faults that do not cause the turbine to shut down, but can reduce performance [20-21].

Other methods proposed include Physics-based modeling [16] where a model of the physical law operating in a gear box is developed. The wind speed time-series seen by a particular turbine is then applied to the model to estimate fatigue damage and the likely condition of the gearbox. Petri-net analysis has been applied to the cooling and lubrication systems of a wind turbine gear box [17]. This method allows non-deterministic modeling of processes that can lead to fault development.

Condition monitoring is now being combined with fault likelihood estimation and automated maintenances scheduling [22]. The ultimate aim is to notify operators when a component is showing early signs of failure, maybe months

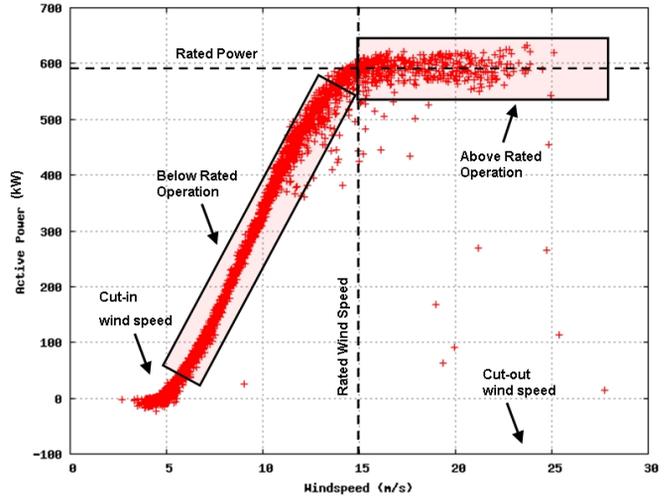


Fig 2: Key features of a wind turbine power curve [29].

before a catastrophic failure occurs, allowing preventative maintenance to be planned around weather and other O&M activities.

## III. WIND TURBINE OPERATION AND EFFICIENCY

The performance of a wind turbine can be visualized by its power curve which relates the power output to the wind speed observed thus giving a measure of performance. The key features of a wind turbine power curve are illustrated in Fig. 1.

The power curve illustrates the operational regimes of a wind turbine. Turbines do not operate at low wind speeds; if the wind falls below a cut-in speed for a specified period of time the turbine will switch off. Around the cut-in speed, it is often necessary for the turbine to draw power from the electrical grid to start up or maintain rotation during short lulls, which can result in negative power production. At higher wind speeds power increases approximately as the cube of the wind speed until rated power is reached, which occurs at the rated wind speed. In above rated operation the turbine control system limits the power extracted from the wind attempting to maintain a constant power output while at very high wind

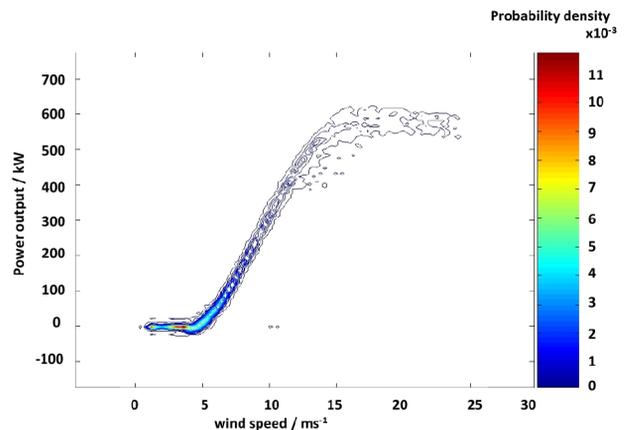


Fig 1: Kernel density estimate of the joint probability of wind speed and power output.

speeds – often 25m/s, the turbine shuts down to avoid incurring damage.

Comparison between published power curves and operational data is difficult for a number of reasons. OEM curves are created under standard conditions, and are recorded using a specific methodology that is not possible to reproduce within an operating wind farm. Operational power curves created from SCADA data use wind speeds measured by the nacelle mounted anemometer. Results from these will differ from the standard power curve due to local turbulence, averaging period and turbine condition. Wind speeds measured at the nacelle are significantly different from the upstream speeds from which the OEM power curves are constructed. Comparisons can be made between operational power curves created using data from different months. This provides a direct method of benchmarking the performance of a specific turbine. Terrain and therefore wind conditions remain the same between the two sets meaning the changes to the power curve will come from changes in the condition of the turbine its self. If the turbine can be assumed to be operating correlly during a particular month, this data can be used as a benchmark power curve for comparison with future data set. In [13], this approach is used to build turbine specific power curves models.

An extension of this is to consider this data as a sample from an underlying bivariate probability distribution; this interpretation is already implied by the scatter plot of active power and wind speed seen in Fig. 1. A probabilistic model has several attractive attributes that could not be achieved with a simple curve fitting analysis. In the first instance, the model would be a direct replacement for the power curve by creating a contour density plot of joint probability density of measured wind speed and active power output.

The example shown in Fig. 2 is produced from a kernel density estimator [23], which has no parametric form but provides an accurate representation of the density it is intended to approximate by summing kernel functions placed on every observed data point. Although computationally expensive, this approach does produce an estimate in situations where a parametric distribution is not a good fit.

#### IV. COPULA STATISTICS

There are circumstances when dependent variables are related by more than linear dependence or correlation. Even if the marginal densities of these variables have a known distribution their joint distribution may not be known. In complex situations, it is not guaranteed that the joint distribution is the multivariate equivalent of the known marginal distributions, even if such a multivariate distribution exists.

Originating from Sklar’s Theorem [24], copulas are a way of describing complex dependency structures and how to relate them to the marginal distributions within a single function.

The formulation of a copula for a bivariate distribution is as follows [25]: Given two random and continuous variables  $X$  and  $Y$ , the probability distribution,  $H(x,y)$ , is defined as:

$$H(x, y) = \Pr\{X \leq x, Y \leq y\} \quad (1)$$

The marginal distributions of  $H$  are given by:

$$F_x(x) = \Pr\{X \leq x, Y \leq \infty\} \quad (2)$$

$$F_y(y) = \Pr\{X \leq \infty, Y \leq y\} \quad (3)$$

The marginal distributions can be used to transform the original random variables  $X$  and  $Y$  to new variables  $U$  and  $V$  with uniform marginal densities on the range  $[0,1]$ :

$$u = F_x(x); v = F_y(y) \quad (4)$$

For any continuous bivariate distribution, Sklar’s theorem states that there exists a bivariate function,  $C$ , such that [24]:

$$C(u, v) = H(x, y) \quad (5)$$

and  $C$  is a copula. To clarify,  $C$  is the bivariate distribution joining  $u$  and  $v$  which are transformed variables with uniform distribution:

$$\begin{aligned} H(x, y) &= C(F_x(x), F_y(y)) \\ &= \Pr\{U \leq u, V \leq v\} \end{aligned} \quad (6)$$

Inverting equation (4) provides a way of estimating the copula:

$$x = F_x^{-1}(u); y = F_y^{-1}(v) \quad (7)$$

And substituting into equation (5):

$$C(u, v) = H(F_x^{-1}(u), F_y^{-1}(v)) \quad (8)$$

If  $H$ ,  $F_x$  and  $F_y$  are known or can be estimated from data they can be used to construct the copula.

Copulas are useful because they allow the marginal distributions and the dependency structure to be specified separately. Copulas have been applied extensively over the past decade in economics to analyze situations where relationships are non-linear and the dependency between the variables is non-symmetric, for example the relationship between variations in three or more exchange rates [26]. Recently, copulas have been applied in a biological context to analyze the firing rate relationships between neurons [27].

Copula estimation can be achieved by fitting parameterized copula families to data or alternatively by a number of non-parametric techniques such as kernel density estimation [28]. In this paper a simple non-parametric method of copula estimation is used. The marginal distributions,  $F_x$  and  $F_y$ , and the full bivariate distribution  $H$  are estimated from a large baseline data set (approximately 6300 data points).

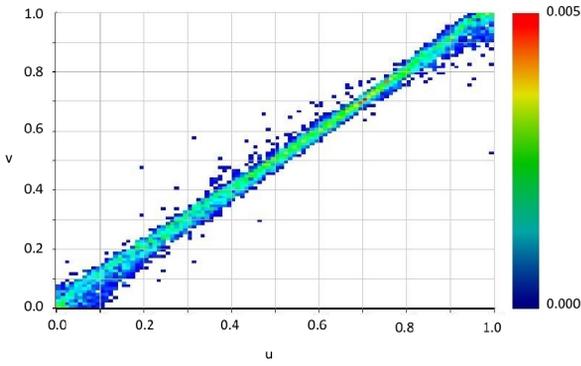


Fig 3: Transformation of a Power Curve into Copula space.

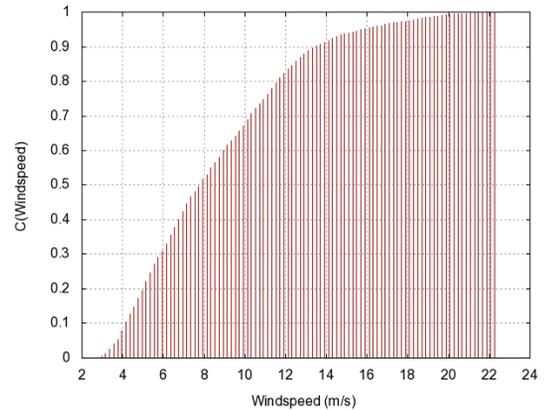
### V. COPULA POWER CURVE MODEL

A copula representation of a power curve can be constructed if the power curve is considered as a bivariate joint distribution [29]. Correct estimation of the wind speed and power marginals are required to ensure the transformed variables have a uniform distribution. Parameterized models such as Weibull or mixture-of-normal distributions can be fitted. However, unless a suitable fit is found this method may lose the specifics of the operational condition of individual turbines. Estimation of the marginal distributions is used to form the estimate of the copula – the distribution of the transformed variables  $u$  and  $v$ . Here non-parametric estimation is used. This non-parametric function is used to make the transformation. An example of an estimated power curve copula is shown as a non-parametric probability density estimate in Fig. 3.

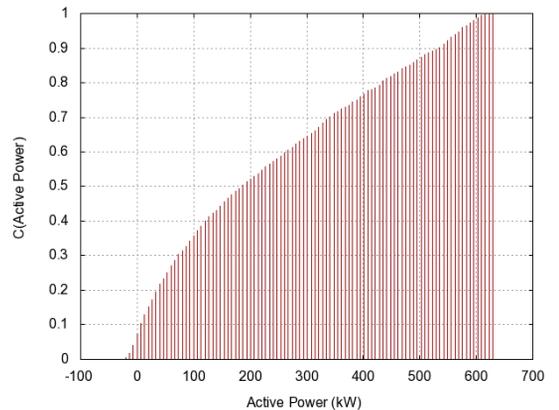
This estimated copula power curve follows a straight line between (0,0) and (1,1) and illustrates the dependency of the data across the range  $[0,1]^2$  showing correct estimation of the marginals. Features include a larger spread of data at the tails compared to the center, this corresponds to a lower correlation between wind speeds and powers around cut-in and above rated. This highlights that the dependency between wind speed and power changes with the operating regime. This precludes the use of a single parameterized copula. The copula model therefore needs to be either a piecewise ensemble of copulas or an empirical copula.

The underlying copula is estimated here using a simple empirical method. Baseline data, obtained from the normal operation of the wind turbine, is transformed into copula space, the region  $[0,1]^2$  is divided into 100x100 bins corresponding to percentiles. A count is made of baseline data in each bin to estimate copula density in that region. This method is used as a simple first pass and a more sophisticated estimator may be based on a kernel density method.

To demonstrate the resulting model in use, SCADA data for wind speed and active power averaged from two wind turbines at a site in central Scotland is used. The turbines are Bonus 600kW Mk IV machines based on a fixed-pitch, fixed-speed, stall-control design. Five minute averages data is used.



(a)



(b)

Fig 4: Cumulative marginal distributions for (a - top) wind speed and (b - lower) active power.

### VI. APPLICATION OF COPULAS TO POWER CURVE ANALYSIS

Exemplar or baseline operational data is taken from the first two months of data for each turbine – April and May – and is used to estimate the marginal distributions and empirical copula. A total of 6300 data points cover the full range of wind speeds and power outputs during this period. Visual inspection of the power curve is used to ensure that the data is without outliers and is nominally consistent with a turbine

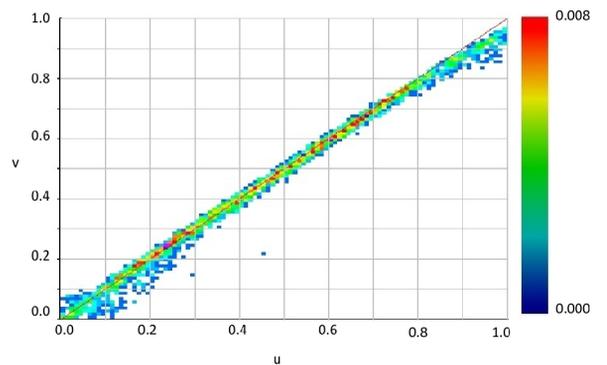


Fig 5: August power curve data mapped into copula space using marginals from figure 4.

???

operating correctly. Records with no output data are removed as it is assumed that the turbine is not operational at these times. The cumulative probability density functions are estimated for the marginals by locating the percentile values and using linear interpolation. The cumulative density functions for turbine 1 are shown in Fig. 4 (a) and (b).

The baseline data is transformed to the copula variables according to:

$$u = F_w^{-1}(w); v = F_p^{-1}(p) \quad (9)$$

Where  $w$  represents wind speeds and  $p$  is turbine power output.

Data is binned in two dimensions as described in section V; the resulting copula density estimate is shown in Fig. 5. It should be noted that results show high joint probability close to the  $u=v$  line with some variation in spread, the region in which data has greatest spread is at the tails. These areas correspond to operation above rated wind speed at the top end, and the region around cut-in speed at the bottom.

To compare data from subsequent months, the estimated marginals,  $F_w$  and  $F_p$ , are used as transforms to map observed data onto the empirical copula. The data for August is transformed and the resultant frequency distribution is shown in Fig. 5. During August readings corresponding to  $u < 0.8$  appear to follow the estimated empirical copula. For  $u > 0.8$  the data drops below the empirical copula. Fig. 6 shows the August data plotted as a power curve, with shading illustrating the distribution of the baseline data. The drop in the August values of  $u$  and  $v$  can be seen to correspond to a small reduction of power outputs for wind speed measured above rated. This drop-off of data relative to the baseline is observed for all months from July to October. Whilst this effect is relatively small, it can have a significant effect on energy generation across an extended period.

In order to compare operational data with baseline across time periods or plant, three statistical measures of similarity are investigated. Since the original data is distributed about the  $u=v$  line with dissimilar data becoming increasingly distant from this two measured based on the sum-of-squared method are proposed. The third measure based on a chi-squared type statistic investigates the difference between the expected and measured number of data points within each binned region of

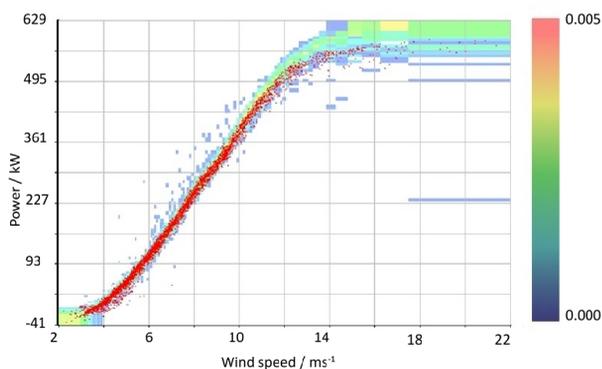


Fig 6: August data plotted as power curve. Shading represents distribution of baseline data.

TABLE I  
ACROSS FULL RANGE OF U

Month	R ( $\times 10^{-2}$ )	R <sup>2</sup> ( $\times 10^{-6}$ )	$\chi^2$ ( $\times 10^{-4}$ )
June	0.201	0.579	2.95905
July	1.32	1.36	1.419221
August	0.931	0.34	2.884236
September	1.60	1.22	2.637837
October	0.845	1.17	2.10833
November	0.459	0.726	2.218387
December	0.571	0.713	4.893074

the copula.

#### A. Sum of Squares Similarity Estimation

As data sampled from the underlying copula should lay, with some spread about the  $u=v$  line a simple statistic for comparison is to sum either the residuals or the square of the residuals relative to this line:

$$R = \frac{\sum_i (v_i - \hat{v}_i)}{n} = \frac{\sum_i (v_i - u)}{n} \quad (10)$$

$$R^2 = \frac{\sum_i (v_i - \hat{v}_i)^2}{n^2} = \frac{\sum_i (v_i - u)^2}{n^2} \quad (11)$$

for analyzing a data set of  $n$  data pairs,  $(v_i, u_i)$  where for each data set the expected value of  $v_i$ , that is  $\hat{v}_i$ , is given by the simple equation  $u=v$ ; this leads to the second equality in each line. The two statistics  $R$  and  $R^2$  have different interpretations:  $R$  will stay close to 0 if the distribution is symmetrical about the  $u=v$  line and will not detect changes in the variability of the data, it will however, detect data sets which deviate from the  $u=v$  line, for example the results in Fig. 5. The  $R^2$  statistic allows detection of changes to the variability of the data whilst symmetry around the  $u=v$  line is maintained. This corresponds to a turbine showing higher variability of outputs across the operating range. Data plotted on the empirical copula for a number of months shows that performance often matches baseline throughout all the operating regimes except above rated. As this corresponds to values of  $u$  greater than approximately 0.8 the  $R$  and  $R^2$  statistics are calculated both for entire range and the range  $u \sim [0.8, 1]$ . Fig. 5 gives an example of this.

#### B. Chi-squared Hypothesis Test

A chi-squared style statistic can be used to test how appropriate a model is for a given data set. In this case how well the test data is modeled by the estimated copula. Test data for a particular month does not require the same marginals as the baseline data, for example in a particularly windy month they may be an excess of data at high  $u$  values.

TABLE II  
ACROSS U RANGE GREATER THAN 0.8 TO 1.0

Month	R (x10 <sup>-2</sup> )	R <sup>2</sup> (x10 <sup>-6</sup> )	$\chi^2$ (x10 <sup>-4</sup> )
June	0.831	7.97	8.91
July	5.0549	9.62	6.00
August	3.3389	12.9	7.68
September	1.7741	3.12	6.73
October	3.2123	16.0	5.74
November	1.7327	6.65	6.20
December	0.0137	4.28	12.7

To account for this the chi-squared statistic is constructed to test how well observed power-data fits the copula give wind speed measurements:

$$\chi^2 = \frac{\sum_u \left[ \sum_v \left( (O_i | n_u) - (E_i | n_u) \right)^2 \right]}{n} \quad (12)$$

Where  $O_i$  is the observed number of data points given the total observed points with the current  $u$  value summed over all  $v(n_u)$ .  $E_i$  is the expected number of data points given  $n_u$ , and  $n$  is the total number of data points.

### C. Application to case study data

The application of the copula analysis method is applied to case study data. The method followed can be summarized by:

- Step 1:** Select baseline data; visual check to ensure consistency
- Step 2:** Estimate marginals for wind speed and power
- Step 3:** Transform baseline data using estimated marginals to produce empirical power curve copula
- Step 5:** Take new turbine data, transform using estimated copula
- Step 6:** Use the statistics developed,  $R$ ,  $R^2$  and  $\chi^2$
- Step 7:** Do statistics show significant changes in dependency from baseline? If yes go to Step 8. Else go to step 9
- Step 8:** A change in power curve dependency has been identified – notify operators of possible fault/anomaly. Continue to Step 9.
- Step 9:** Take the next months data.

Tables I and II list the monthly  $R$ ,  $R^2$ , and  $\chi^2$  values as calculated for turbine 2 and plotted in Fig. 7. As noted above, turbine 2 showed signs of sub-optimal operation during the months of July to October with the copula representation showing a droop for high values of  $U$  (transformed wind speed).

Of the statistics analyzed, high  $R$  values occur during these months. The difference is more pronounced for  $R$  across the restricted range ( $R_{U-[0.8,1]}$ ). The other statistics do not, in this

case, provide a reliable signature that matches the observations. It is anticipated that each statistic will be useful in detecting specific fault modes as each will leave a particular signature on the power curve dependency data. The possible correlation between signatures and fault modes is discussed in the next section.

As an example of a simple warning system, monthly data in which the value of  $R_{U-[0.8,1]} \geq 2 \times 10^{-2}$  would provide a trigger for operators of this turbine suggesting an incipient fault. The particular value of this limit will differ between turbines. The copula method outlined here provides a way of detecting *changes* in the dependency between wind speed and power so whilst  $2 \times 10^{-2}$  provides a useful value here, this is likely to vary between turbines depending on the specific location conditions. For example if a turbine sees a higher turbulence level, the value of  $R$  is likely to be higher.

## VII. DISCUSSION

The results from the previous section showed that the Copula approach has the ability to measure and identify changes in dependency between operationally measured wind speed and power. These changes may be linked to specific faults, or more gradual changes such as blade surface wear which on their own can lead to sub-optimal performance.

The correlation of faults or anomalies to statistical signatures is key to developing copula-power curve condition monitoring. This initially would require detailed fault-logs and

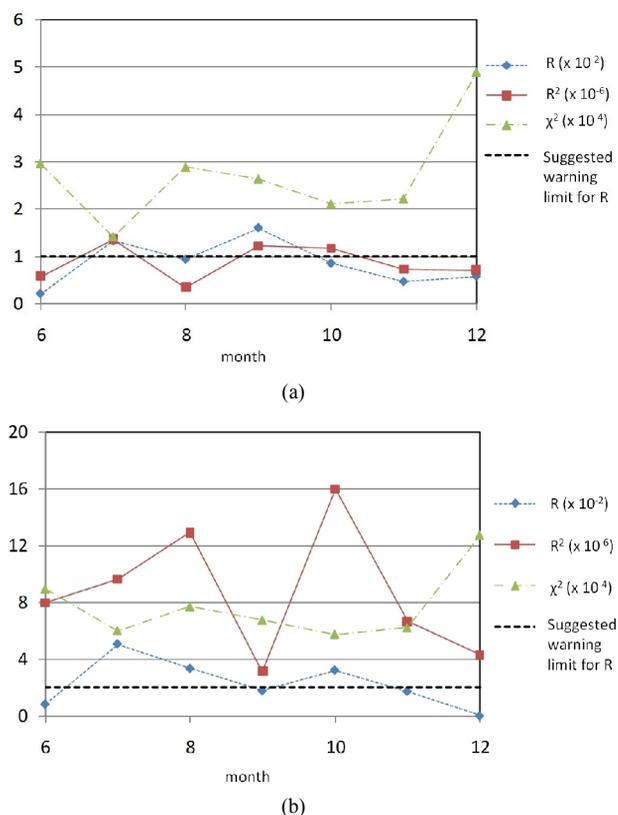


Fig 7: Monthly statistics for case study data including suggested warning limits; upper (a)  $u \sim [0,1]$ , lower (b)  $u \sim [0.8,1]$

regular, detailed manual inspection of turbines to provide information on issues like blade condition. Due to commercial sensitivities fault log data is often difficult to gain access to. In fact, many minor faults may go unnoticed and future work will need to include more detailed observations of turbines.

This section provides some discussion on likely links between the copula condition monitoring statistics and wind turbine faults/anomalies. These may be difficult or impossible to detect using current techniques and this section provides suggestions which the authors intend to continue in future work.

Faults/anomalies will change the dependency between wind speed and power output and should therefore produce signatures in the  $R$ ,  $R^2$  and  $\chi^2$  statistics. Specific fault modes will affect the turbine system in different ways so combinations of statistical signatures will be able to provide suggestions on the type of fault the turbine is experiencing.

#### A. Blade Faults

Degradation of the blade surface leads to a reduction in power production as the turbine will have reduced aerodynamically efficient. The Copula-power curve will therefore lie below the  $u=v$  line. In stall regulated machines such as those studies here, this is likely to effect the onset of stall and may contribute to the sub-optimal performance identified here. The ‘ $R$ ’ statistic is likely to provide a good characterization of such deviations, and provide information on the direction of power production variation.

Minor blade damage may not be detectable using the current sensor systems attached to wind turbines, however even a small reduction in aerodynamic efficiency can have a significant effect on the profitability of a turbine. Here the changes in dependency caused by gradual changes in the quality of the aerodynamic surfaces can be identified by the proposed copula method.

#### B. Yaw System Faults

The yaw system attempts to maintain the turbine pointing directly into the wind. Misalignment leads to lowered airflow through the turbine and therefore lower power production. This will occur across the wind speed range. The effect is likely to be noticeable with a negative  $R$  value with the signature appearing when calculated across the range  $u \sim [0,1]$ . In addition to this signature the values of  $R^2$  and  $\chi^2$  would be expected to increase.

#### C. Pitch System Faults

Most modern turbines are pitch regulated. Below ‘rated wind speed’ the blades pitch to the angle which allows greatest aerodynamic efficiency. With wind speed above rated the blades pitch to reduce the fraction of power transferred from the wind and to maintain rated power. Faulty pitch mechanisms are likely to show up through a greater variability at all wind speeds, or may lead to over or under production of power at high wind speeds. Increased variability will lead to higher values of  $R^2$  and  $\chi^2$ . The value of  $R$  is unlikely to provide a clear signature as variability will occur above and below the  $u=v$  line.

These three fault modes provide a subset of all possible faults causing sub-optimal performance. The study of larger data sets with more fault information will allow correlations to be built up between copula statistics and faults. This provides a clear aim of future research.

Copula-power curve condition monitoring provides a method of analyzing dependency data that can complement existing condition monitoring methods. Whilst neural-network analysis of SCADA data typically studies correlations between data over the past two or three measurements, copula will allow comparison of dependencies over weeks, months or longer. The use of copula methods to detect either minor faults or performance degradation can be used to trigger more in-depth fault-analysis such as running physics based models to suggest possible faults modes.

### VIII. CONCLUSIONS

The application of copulas to wind turbine power curves has been shown to allow analysis of the underlying dependency between wind speed and power; this paper has presented the example of an empirical Copula tracking the non-linear and non-stationary dependency between wind speed and active power output from the SCADA data of two operational wind turbines. Further developments may include the comparison, using copulas of the dependency from two or more turbines, and transforming data from one turbine using the marginal distributions of another to allow comparison of the variability of data at different levels of production.

This work provides a first pass at Copula modeling for power curves. A more sophisticated method of parametric estimation of marginals and dependency is required for the approach to be maximally useful. This may take the form of a mixture density estimate of the marginals and a cubic spline estimate of the copula which would additionally capture and identify changes in operating regime. Additionally, features within each regime could be identified by the parameterization of the model such as the variance over the gradient of the linear part of the curve which indicates the efficiency of the plant over its useful operating range of wind speeds, including the region between rated wind speed and cut-out where the machine is operating at full power. Piecewise application of Copula models to each of these regions may be sufficient to capture these features of interest as local bivariate probability densities.

### REFERENCES

- [1] DECC, *The UK Renewable Energy Strategy*, DECC, Editor. 2009: London.
- [2] European-Commission, *Renewable Energy Road Map*. Available: [http://europa.eu/legislation\\_summaries/energy/renewable\\_energy/l27065\\_en.htm](http://europa.eu/legislation_summaries/energy/renewable_energy/l27065_en.htm) (accessed 19th May 2011)
- [3] *Renewable energies in the 21st century: building a more sustainable future*. 2006, European Commission. p. 21.
- [4] DECC. *A Prevailing Wind: Advancing UK Offshore Wind Deployment*, June 2009, London
- [5] Wiggelinkhuizen, E.J., et al., *Connow - Final Report*. 2007, Energy Research Center of the Netherlands.
- [6] BERR, *Scroby Sands Offshore Wind Farm 3rd Annual Report*. 2007: London.
- [7] Blanco, M.a.I., *The economics of wind energy*. *Renewable and Sustainable Energy Reviews*, 2009. 13.

[8] Tavner, P.J., J. Xiang, and F. Spinato, *Reliability analysis for wind turbines*. Wind Energy, 2007. **10**(1): pp. 1-18.

[9] Hahn, B., M. Durstewitz, and K. Rohrig, *Reliability of Wind Turbines*, in *Wind Energy*, J. Peinke, P. Schaumann, and S. Barth, Editors. 2007, Springer Berlin Heidelberg. pp. 329-332.

[10] Spinato, F., et al., *Reliability of wind turbine subassemblies*. Renewable Power Generation, IET, 2009. **3**(4): pp. 387-401.

[11] Tavner, P.J., et al., *Condition Monitoring of Rotating Electrical Machines*. 1 ed. Power and Energy Series 2008, Stevenage: The Institute of Engineering and Technology.

[12] IEC 61400: Wind Turbine Generator Systems, Part 12: Wind Turbine Power Performance Testing. 1998.

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

[14] Hameed, Z., et al., Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable Energy Reviews*, 2009. **13**(1): pp. 1-39.

[15] Crabtree, C.J., Survey of commercially Available Condition Monitoring Systems for Wind Turbines. 2010, Durham University; SuperGen Wind.

[16] Gray, C.S. and S.J. Watson, Physics of Failure approach to wind turbine condition based maintenance. *Wind Energy* **13**(5): pp. 395-405. Wiley 2010.

[17] Rodriguez, L., et al. Application of latent nesting method using Coloured Petri Nets for the Fault Diagnosis in the wind turbine subsets. in *Emerging Technologies and Factory Automation, 2008. ETFA 2008. IEEE International Conference on*. 2008.

[18] Zaher, A., et al., Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy*, 2009. **12**(6): pp. 574-593.

[19] A. Kusiak and W.Y. Li, The Prediction and Diagnosis of Wind Turbine Faults, *Renewable Energy*, Vol. 36, No. 1, pp. 16-23, 2011.

[20] Leaney, V., C., D. Sharpe, J., and D.G. Infield, *Condition Monitoring Techniques for Optimisation of Wind Farm Performance*. *International Journal of Comaden*, 1999. **2**(1).

[21] Xiang, Y., et al. Unsupervised learning and fusion for failure detection in wind turbines. in *Information Fusion, 2009. FUSION '09. 12th International Conference on*. 2009.

[22] Garcia, M.C., M.A. Sanz-Bobi, and J. del Pico, SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a wind turbine gearbox. *Computers in Industry* **57**(6): pp. 552-568, Elsevier 2006.

[23] R. Duda, P. Hart and D. Stork, "Pattern Classification" 2nd Edition Wiley-Interscience, 2001.

[24] Sklar, A., *Fonctions de repartition a n dimensions et leurs marges*. *Publ. Inst. Statist. Univ. Paris*, 1959. **8**: pp. 229-231.

[25] R.B. Nelsen. "An Introduction to Copulas", 2nd Edition. Springer Series in Statistics, 2006.

[26] Patton, A.J., *Modelling asymmetric exchange rate dependence*. *International Economic Review*, 2006. **47**(2): pp. 527-556.

[27] Berks, P., F. Wood, and J. Pillow, *Characterizing Neural Dependencies with Coupled Models*. *Advances in Neural Information Processing Systems 2009*. **21**: pp. 129-136.

[28] L, R., On the distribution transform, Sklar's Theorem, and the empirical copula process. *J. Statist. Plan. Infer*, 2009. **139**(11): pp. 3921-3927.

[29] Stephen, B.; Galloway, S. J.; McMillan, D.; Hill, D. C.; Infield, D. G.; "A Copula Model of Wind Turbine Performance", *IEEE Transactions on Power Systems* **26** 2, pp. 965 – 966, 2011.

applications in Power System and Animal Welfare Condition Monitoring and Asset Management.



**Dr Stuart Galloway** is a Senior Lecturer within the Institute for Energy and Environment. He obtained his MSc and PhD degrees in mathematics from the University of Edinburgh in 1994 and 1998 respectively. His research interests include power system optimization, numerical methods and simulation of novel electrical architectures.



**Simon Gill** is currently a PhD candidate within the Wind Energy Doctoral Training Centre at the University of Strathclyde. He obtained a Masters degree in Astrophysics from the University of Edinburgh in 2003, followed by a PGCE in secondary science education in 2005. His research interests include active management of distribution networks and the integration of renewable energy into power systems.



**Dr Bruce Stephen** currently holds the post of Senior Research Fellow within the Institute for Energy and Environment at the University of Strathclyde. He received his B.Sc. from Glasgow University and M.Sc. and PhD degrees from the University of Strathclyde. His research interests include Distributed Information Systems, Machine Learning