Development of a stochastic computational fluid dynamics approach for offshore wind farms

Richmond, M, Kolios, A, Pillai, VS, Nishino, T & Wang, L

PAPER • OPEN ACCESS

Development of a stochastic computational fluid dynamics approach for offshore wind farms

To cite this article: M Richmond *et al* 2018 *J. Phys.: Conf. Ser.* **1037** 072034

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Development of a stochastic computational fluid dynamics approach for offshore wind farms

M Richmond[a], A Kolios[a], V S Pillai[a], T Nishino[a] and L Wang[b]

[a] *Offshore Energy Engineering Centre, Cranfield University, Bedfordshire MK43 0AL, United Kingdom*
[b] *School of Mechanical, Aerospace and Automotive Engineering, Coventry University, Coventry, CV1 5FB, United Kingdom*

Corresponding author email: m.richmond@cranfield.ac.uk

Abstract. In this paper, a method for stochastic analysis of an offshore wind farm using computational fluid dynamics (CFD) is proposed. An existing offshore wind farm is modelled using a steady-state CFD solver at several deterministic input ranges and an approximation model is trained on the CFD results. The approximation model is then used in a Monte-Carlo analysis to build joint probability distributions for values of interest within the wind farm. The results are compared with real measurements obtained from the existing wind farm to quantify the accuracy of the predictions. It is shown that this method works well for the relatively simple problem considered in this study and has potential to be used in more complex situations where an existing analytical method is either insufficient or unable to make a good prediction.

1. Introduction

Offshore wind power is growing rapidly, not only within the UK and EU but is beginning to gain traction around the world in other markets [1]. To continue its growth and expansion into other markets the cost needs to be reduced. While capital expense (CAPEX) has dropped significantly, the proportion of operational expense to the total levelised cost has risen [2]. To aid in this area, new methods need to be developed to more effectively predict the loads on turbines not only at discrete points in time, but over the entire 25-30 year life span of a project.

A wide range of methods exist and have been used in literature for modelling the wake of turbines within a wind farm. These methods include Navier-Stokes computational fluid dynamics (CFD) solvers where the turbine is modelled simply as an actuator disk [3, 4] to more computationally intensive models of fully resolved wind turbines [4, 5]. These can provide accurate answers and predict the flow quite realistically but are limited to modelling only relatively short time intervals due to their computationally intensive nature.

Through CFD simulation, it is hard to cover the wide range of variables experienced by a wind farm; the wind speed can range from a turbine's cut-in speed of around 3-4 m/s to its cut-out speed around 24-25 m/s and can flow from all directions. The pitching strategy employed results in a non-linear response of these turbines. In additionally, many other parameters can also vary such as turbulence intensity, yaw misalignment, blade rotational speeds, wind shear and temperature gradients. With so many variables, it is unlikely for any two points of data recorded to ever be the same and hence using a deterministic CFD approach is unfeasible.

Engineering wake models, such as Jensen and Larsen [6] which can provide solutions to wake interaction very quickly, can be used for a Monte-Carlo analysis. However, they are limited in what they can model and cannot be used in transient simulations to evaluate time-variant phenomenon and also struggle with wake redirecting.

Therefore, a more adaptable stochastic method is needed which can not only predict simple cases but also which can be adapted to transient and less conventional problems. The current research develops such a method which combines CFD model results with approximation models and stochastic inputs to build a probabilistic view of wind turbine array loading.

2. Methodology

The current research proposes a novel approach for extrapolating computational fluid dynamics results of wind turbine arrays to a stochastic analysis. The first step is to conduct a series a CFD analyses at a discrete set of variables such as different wind directions and freestream velocities. The results from these analyses are then used to train regression or surrogate models from which predictions for other input values can be generated. The Input values for the CFD studies should be sufficiently close to each other to allow for accurate prediction from the approximation models but should cover the range from which predictions will be made. The current study reveals that the location and selection of independent variable values is a question of compromise between resource availability and resource utilization with the accuracy of results needed.

2.1. Actuator Disk model.
This method is applied to an existing wind farm which contains 25 turbines and a meteorological mast located on the outer edge of the wind farm. Data obtained from this wind farm are used for constructing the CFD model. This data includes SCADA data from four of the turbines as well as the met mast.

The wind farm is modelled at a range of incoming wind angles between 35 degrees and -35 degrees and the freestream velocity is modelled between 5 m/s and 15 m/s. All combinations from the values presented in *Table 1* are modelled with the CFD analysis, resulting in 35 cases. The limit of incoming wind direction is chosen so that greater accuracy can be attained with the resources available while still adequately evaluating this methodology. The values for the limited range are chosen because accuracy is later compared with values from the met mast and this range shows the most variation at the met mast.

The number of samples used to train the approximation method is a very important consideration and is dependent on the method used. While regression methods typically perform better with hundreds or even thousands of data samples, surrogate models can perform well with far fewer. For example, Queipo et al. [7] performed a case study using 54 designs to optimise 4 parameters using a surrogate model based on radial basis function (RBF). In another example using RBF [8], 200 samples were used to estimate parameters in a 400 km by 700 km area. For the response surface method, for a full factorial design of experiments, it is recommended to have a minimum of either 2^N samples or 3^N samples including center points, where N is the number of variables [9]. The work in this study uses 35 sample points for 2 variables. It can be seen in the results in section 3 that this amount is sufficient for some methods but not for others.

Table 1. Independent variable values from which all CFD cases were based.

Direction (degrees)	0	10	20	35	-10	-20	-35
Freestream velocity (m/s)	5	7.5	10	12.5	15	-	-

The model used is an actuator disk (AD) model implementing the SIMPLE solver method in OpenFOAM 4.0. This is a modified version of Svenning's actuator disk model [10] which represents the turbine not as an infinitely thin disk but as a three dimensional region. This gives more flexibility in

changing the direction of the turbines. The model has been modified to allow multiple instances of the turbine and, more importantly, to measure the flow velocity at the disk and use AD theory to determine an axial induction factor and subsequently a local upstream velocity. This can then be used with turbine thrust and torque curves to determine the correct value for a given wind speed. These curves are generated from the SCADA data from the wind farm through filtering and then using efficiency to find rotor power from generator power and finally using basic equations to find thrust and torque.

Additionally, the wind shear profile is determined from the met mast data which contains velocity values at several different heights. The wind shear profile is then fitted to the met mast values. A grid sensitivity study is conducted to determine suitable cell sizes with a base set of values used being taken from a paper on validating OpenFOAM's actuator disk model [11], though a more fine, structured grid is used in the end.

An initial result from this model is shown in **Figure 1** which shows a two-dimensional cross section of the flow field through the turbine centres. This image shows the velocity magnitude field.
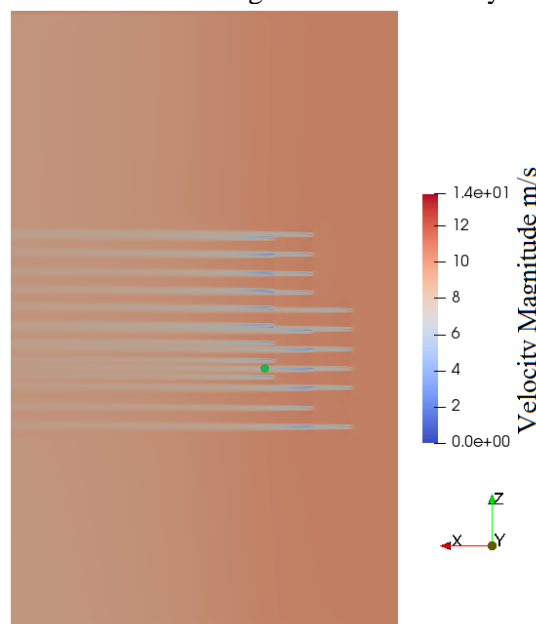
Figure 1. *Velocity-magnitude field from steady state actuator disk model, met mast location denoted in green*

The two turbulence models compared are standard K-Epsilon and K-Omega SST. The results from this comparison are in line with literature which states that the K-Omega SST model is more accurate for this case as the K-Epsilon model under predicts wake deficit [12]. The results from this comparison are shown in **Figure 2**. The K-Omega SST model shows good agreement with the data, except at higher velocities. However, the mean velocity in the data is roughly 7.5 m/s, so the K-Omega SST model is better for most points.

Figure 2. *Comparison of turbulence models with met mast data.*

2.2. *Approximation modelling*

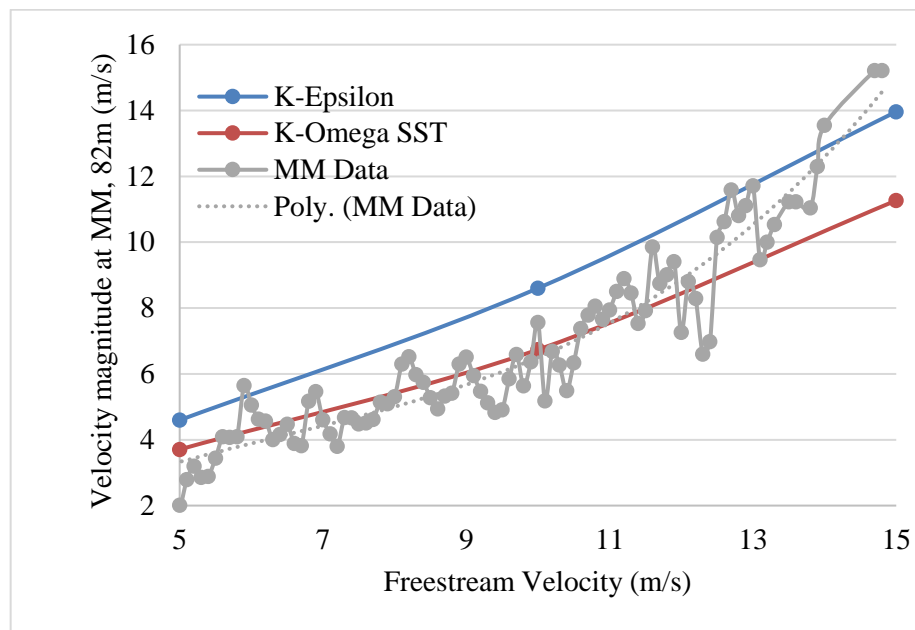A key part of this approach is the ability to quickly predict between the modelled inputs based on the data from the CFD model. To this end, several approximation methods are tested including Support Vector Regression (SVR) as well as artificial neural networks (ANN). The effectiveness of these predictions are compared not only with test models of certain intermediate values but also with the building of probability distributions and probability of exceeding thresholds for data which is also held for the real wind farm. The models compared are: 1. Random Forest Regression, 2. Gaussian Process, 3. Support Vector Regression, 4. Radial Basis Function and 5. Artificial Neural Networks. The differences between these methods and the codes used are presented within this section. The results from this are presented and discussed in later sections.

2.2.1 *Random Forest Regression*

Random forest regression is an ensemble method based on Decision Tree Analysis where the results of multiple Decision Tree models are combined to produce a prediction. This method was first proposed by Ho [13] and improved by Breiman [14]. The method is widely used because it is fast and robust, for example it was used by Microsoft for pose recognition in their Xbox Connect [15]. The algorithm is implemented in python using the Scikit learn library, *sklearn.ensemble.RandomForestRegression* [16].

2.2.2 *Gaussian Process Regression*

When a Gaussian process is used for regression, it is a stochastic process where distributions are defined over a function. After obtaining a predictive distribution, the regression is then applied over a basis function which projects the input onto the feature space [17-19]. The algorithm is implemented in python using the Scikit learn library, *sklearn.gaussian_process.GaussianProcessRegression* using a radial basis function kernel [20].

2.2.3 *Support Vector Regression*

Support Vector Regression (SVR) is where support vector machines (SVM) are used for regression. SVMs, first identified by Vapnik[21], work through linear domain division where the division is made to be as large as possible. This can also be extended to higher order domains and be used for regression

through the use of kernels [19, 20]. The algorithm is implemented in python using the Scikit learn library, *sklearn.svm.SVR* using a radial basis function kernel [24].

2.2.4 *Radial Basis Function*
A radial basis function fits a surface through the measured sample points. The values between the sample points are determined from functions as the radial distance from the point. The method has been used in soil parameter estimation [22, 23]. The code used was developed in Matlab at Cranfield University [27]. **Figure 3** demonstrates the prediction result for the RBF regression method; it is a smooth surface passing through all of the sample points (red dots).
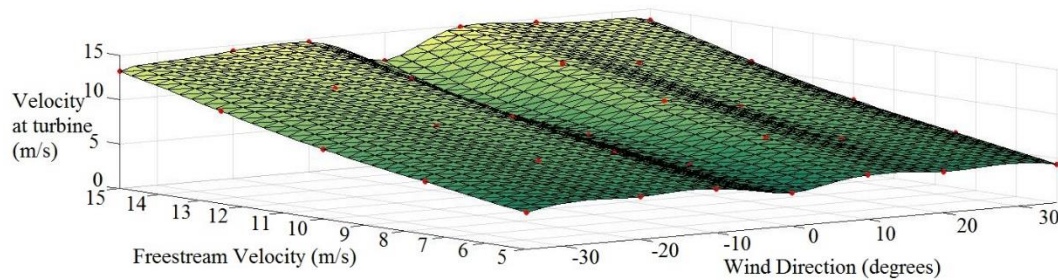

Figure 3. *Prediction surface of RBF method for velocity magnitude at turbine 16*

2.2.5 *Artificial Neural Networks*
Artificial Neural Networks (ANN) were first developed by Rosenblatt as a 'perceptron' in 1957 [28]. This has since been improved upon by the addition of hidden layers of neurons which contribute towards the end result. Matlab's Artificial Neural Network toolbox is used for this study [29]. The best results are found when trained with the Bayesian Regularization method.

2.3 *Validation methodology*
The aim of the method is to make predictions for values in the flow-field, first through conducting a CFD analysis and then training the approximation model on the CFD data. In a conventional validation of a regression fit, the data is split into a test set and a training set; the model is fitted to the training set and the predictions against the test set are used for gauging the accuracy. This method was not chosen for this study as it was not appropriate for this particular case for several reasons: the CFD cases are not randomly distributed but deliberately chosen to maximise their usefulness to the approximation method. Additionally, the fact that there are relatively few sample points, due to the heavy resource requirement in getting the sample points, means that the accuracy of a reduced training set would be significantly impacted when compared to a full training set. Instead, with the aim of predicting real world values, the predictions from the approximation models, trained to the CFD data, were compared against measured values from the real wind farm. The data from the wind farm is recorded in 10-minute intervals between 28th of October 2010 to 31st January 2011.

The available data which can be compared to are wind speed values recorded at a met mast whose location is as shown in both **Figure 1** and **Figure 4**. From this, only wind speed at the met mast at the height of 82 m above mean sea level are used as the data set does not contain any other data which makes for a good comparison. Free-stream wind values of direction and wind speed are also present which are recorded at the same times as the met mast values although there is some limitation due to the fluctuations in the correlation between them.

There are three comparisons made for producing a quantified error. The first method is meant to evaluate the entire method's ability to predict real values. 100 data sample recorded at the met mast are taken and attempted to be predicted by the CFD-approximation process. For each of the 100 samples at the met mast, there are corresponding free-stream conditions which were recorded at the same time;

these free-stream conditions are used as independent variables in the approximation models to predict the values at the met mast, the errors are presented in **Table 4**.

As discussed, potentially a large part of the error is fluctuation in the data, resulting partially from a mismatch between the free-stream conditions recorded at the time and the values recorded at the met mast. Therefore, the second comparison is in the already trained approximation model's ability to predict the same value at the met mast for three new free-stream conditions. To evaluate the ability of the approximation methods to predict the CFD result, three additional CFD cases are performed. The first case is within a previously modelled direction at a non-modelled speed, the second is as a previously non-modelled direction at a modelled speed and the last case is at both a non-modelled direction and speed. The input values are shown in **Table 2**.

Table 2. Independent variables for the three CFD comparison cases

	Wind Direction (degrees)	Freestream velocity (m/s)
Case 1	0	8.25
Case 2	15	10
Case 3	15	8.25

The third comparison is in the stochastic utility of the model. A stochastic analysis was performed on turbine 16 to predict the probability of the power production being at the rated power of 3.6 MW. Turbine 16 was chosen because it is inside the wind farm, not at the freestream side, and SCADA data is present for this turbine. Based on the SCADA data, this turbine is at rated power 10.1 % of the time. For comparison, 10000 predictions are made based on random sampling of the independent variables for the wind speed at turbine 16. The distributions from which samples were taken are a triangular distribution for direction and a Weibull distribution for freestream velocity. These distributions were fitted to filtered data using Palisade @RISK software [30]. The values for the distributions are given in **Table 3**. The assumption is made that if the velocity is at or above nominal velocity then the turbine is at rated power. The nominal velocity is between 13-14 m/s so 13.5 m/s was used. An axial induction factor of 0.3 is used to convert turbine velocity to upstream velocity. This was not conducted for RBF because the code did not allow for it.

Table 3. Distributions used for the Monte-Carlo analysis

	Minimum	Most likely	Maximum
Direction (Triangular distribution)	-25.104	296.8	368.04
	Shape α	Scale β	Shift
Freestream Velocity (Weibull distribution)	1.486	5.8776	1.4761

3. Results

3.1 Prediction accuracy

To evaluate the error of the method as a whole, 100 samples of data at the met mast were randomly taken and predicted with the approximation methods trained from the 35 CFD data cases. It is to be noted that there is considerable noise in the met mast data, as can be seen in **Figure 4**, which shows the deviation from ambient velocity at the met mast by wind direction. The met mast is not in the wake of the wind farm between roughly 120 degrees and 260 degrees and so the deviation should be zero, however this varies considerably. The deviation is understated in the figure as the data points are averaged into bins. The root mean square error of this data is found to be 21.3% and therefore the error of the 100 predictions is inevitably going to be high, which is why, in addition, the error in the approximation model to predict the CFD result is also investigated.
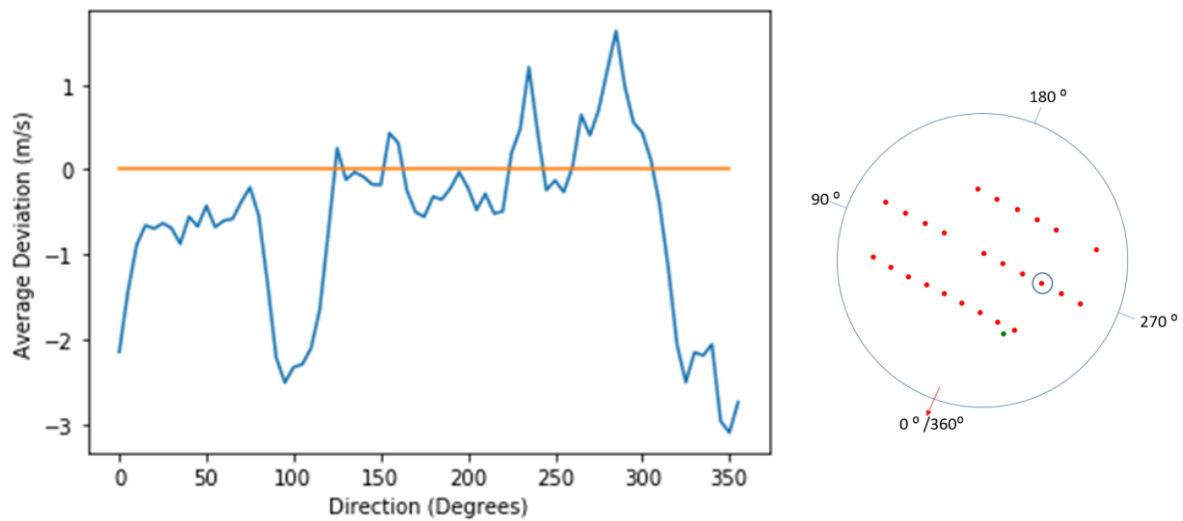
Figure 4. *Wind velocity deviation from freestream velocity by angle at the met mast (left) and a map of the wind farm (right), turbines in red and met mast in green, turbine 16 circled in blue, red arrow shows wind direction of 0 degrees.*

The average magnitude of the errors from each approximation method is shown in **Table 4**. For each of the 100 free-stream points of data used as an independent variable into the regression model, there is a corresponding wind speed values at the met mast as dependant variables. The mean magnitude of the difference between the prediction and the recorded met mast value is shown in the '100 points (%)' column. As discussed earlier, there is considerable fluctuation in the met mast data, which was quantified earlier to be roughly 21.3%, this value was subtracted from the '100 points' value in the 'adjusted error' column to put the error into context. The 'CFD error' is the mean magnitude of the difference between the CFD results and the approximation model prediction for the met mast position. As discussed in Section 2.3, these are wind speed values.

Table 4. *Average magnitude of errors for each method and each comparison of wind speed*

	100 points (%)	Adjusted error (%)	CFD error (%)
Random Forest	27.5	6.2	5.4
Gaussian Process	35.4	14.1	17.4
Support Vector Regression	27.8	6.5	14.6
Radial Basis Function	27.0	5.7	9.1
Artificial Neural Networks	22.9	1.6	10.3

Some of these results appear contradictory, but they represent slightly different things. The CFD values are deliberately mid-way between previously modelled inputs and so the error is the maximum that it can be, while many of the data points inevitably are closer to the previously modelled values. All of these errors could be reduced significantly if the approximation models were trained on more CFD data, so this should not be considered the highest accuracy of the method. As an example, referring back to **Figure 3**, which looked at the prediction surface for turbine 16, along the top for directions 10, 0 and -10 there is a noticeable dip in the prediction - it is unlikely that these three points are sufficient to accurately capture the shape of this dip.

3.2 Stochastic Analysis
Each model was used to conduct a stochastic analysis to determine the percentage of time at which turbine 16 was at or above rated velocity. The real value from the data for this particular turbine was 10.1% of the time. The method, outlined in section 2.3, was to randomly sample free-stream values from fitted distributions to produce independent variables for each approximation model which was fitted to

the CFD results. The results from this are presented in **Table 5** which shows the predicted percentage of time at rated power and also the difference from the actual percentage of time.

Table 5. Predictions and errors for percentage of time turbine 16 is at rated power

Method	**RF**	**GP**	**SVR**	**ANN**
Percentage at rated power	11.2	8.5	12.0	10.5
Difference from actual	+1.1	-1.6	+1.9	+0.4

As can be seen in **Table 5**, ANN and RF produced the closest prediction with an absolute difference in values of 0.4 and 1.1 respectively. SVR was the least accurate, over predicting the percentage of time at rated power by a difference of 1.9.
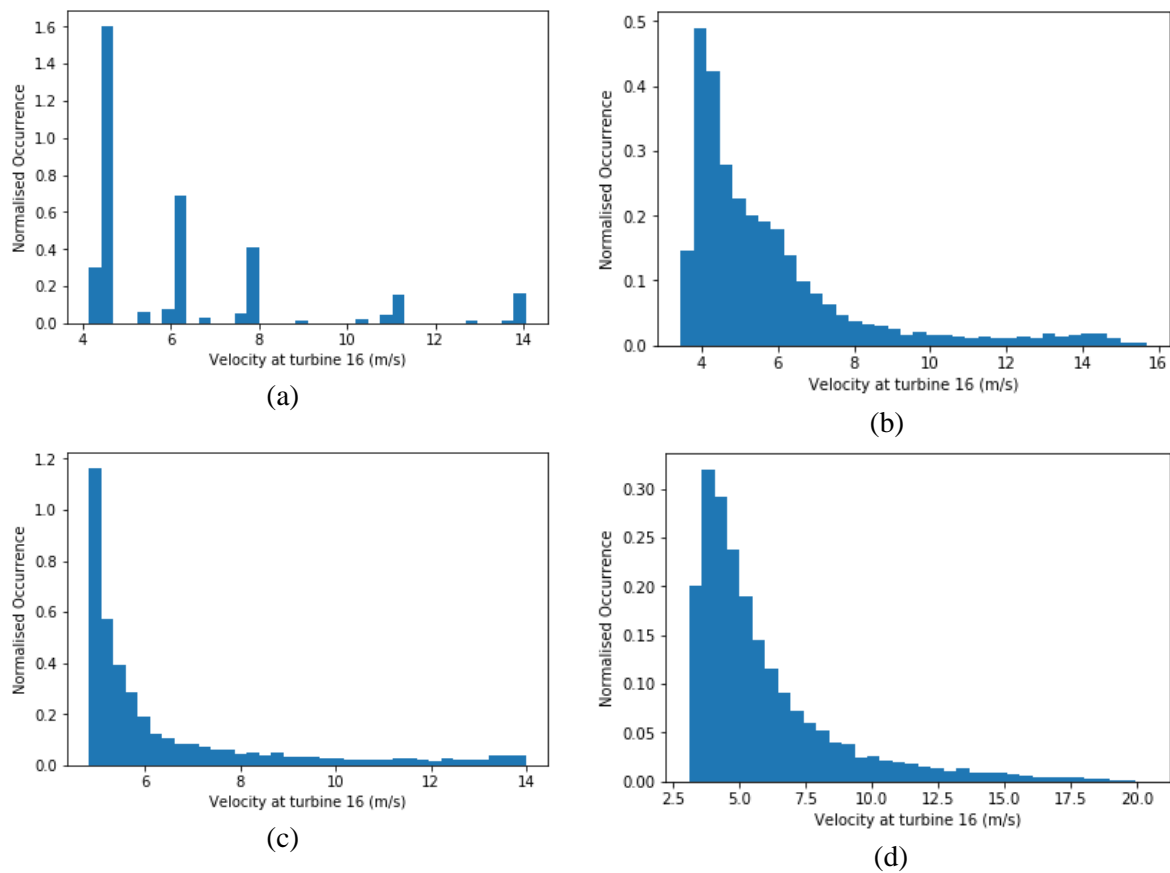

Figure 5. Probability distribution functions for the velocity at turbine 16 for (a) RF, (b) GP, (c) SVR and (d) ANN respectively.

For comparison to this, a histogram of the wind speed at the turbine, recorded for turbine 16, is shown in *Figure 6*. This is with the same limits for direction as used in the rest of the research. The shape of *Figure 6* most closely resembles the shape of the prediction from ANN in *Figure 5*(d).
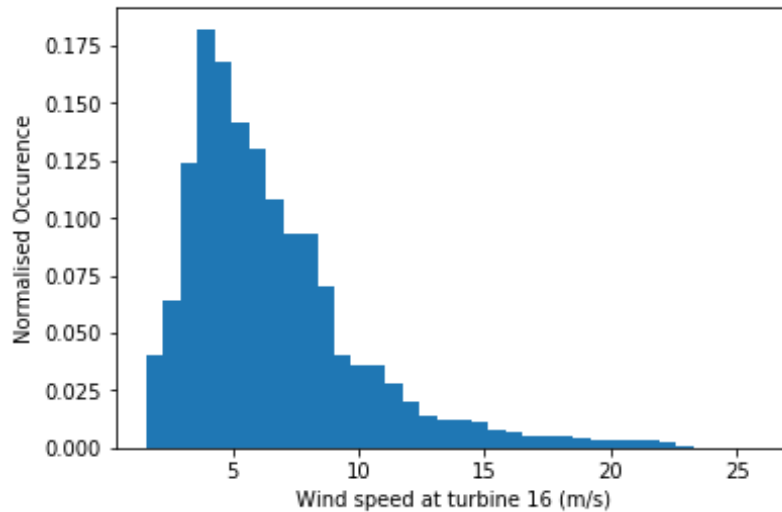
Figure 6. *Histogram of recorded wind speed, turbine 16*

The closest prediction for time at rated power and the most convincing distribution is produced from Artificial Neural Networks. Random Forest is clearly over-fitted to the training data, meaning it is making most of its predictions around the training data and not accurately predicting values not close to the training data, which is why the distribution for RF appears to be discontinuous. SVR shows a rise towards the higher velocities but a sudden cut-off at 14 m/s.

3.3 Practicality
In terms of practicality, the approximation methods are computationally very quick and not at all resource intensive for this particular problem. Training each approximation model on the 35 sample points took no more than a few seconds and then producing the predictions was similarly quick. However, getting the CFD data required terabytes of storage and considerable CPU time.

4. Conclusions
In this study, a method if developed for conducting stochastic analysis of an offshore wind farm. This method takes results from CFD analyses at a set of input values, builds an approximation model of the system from the CFD data and produces predictions between the modelled inputs. These quick predictions are then used to conduct a Monte-Carlo analysis of the system and hence produce distributions for values at the turbines and calculate probability of exceeding a given threshold.

The results are promising with relatively low errors for most methods. For a series of predictions, Random forest, Radial Basis Function and Neural Networks show promise, while Gaussian Process and Support Vector Regression have the highest errors. In the stochastic analysis the two best methods were GP and ANN, the prediction from SVR had a relatively large error and RF appeared to over-fit the training data. On balance, ANN appears to be the best method and RBF is accurate for individual predictions and may have performed well in the stochastic analysis.

Future study in this area will be to apply this method to a transient simulation of a small, highly monitored wind farm with the aim of predicting turbulent fluctuations and hence quantifying cyclic loading on wind turbines in a stochastic manner.

References
[1]   N. Era, W. Power, and U. States, "Wind Vision :," 2016.
[2]   T. Poulsen, C. Hasager, and C. Jensen, "The Role of Logistics in Practical Levelized Cost of Energy Reduction Implementation and Government Sponsored Cost Reduction Studies: Day and Night in Offshore Wind Operations and Maintenance Logistics," *Energies*, vol. 10, no. 4, p. 464, 2017.

[3]   F. Porté-Agel, H. Lu, and Y.-T. Wu, "Interaction between Large Wind Farms and the Atmospheric Boundary Layer," *Procedia IUTAM*, vol. 10, no. 0, pp. 307–318, 2014.

[4]   N. Stergiannis, C. Lacor, J. V Beeck, and R. Donnelly, "CFD modelling approaches against single wind turbine wake measurements using RANS," *J. Phys. Conf. Ser.*, vol. 753, p. 032062, 2016.

[5]   L. Wang, R. Quant, and A. Kolios, "Fluid structure interaction modelling of horizontal-axis wind turbine blades based on CFD and FEA," *J. Wind Eng. Ind. Aerodyn.*, vol. 158, pp. 11–25, 2016.

[6]   M. Gaumond *et al.*, "Benchmarking of wind turbine wake models in large offshore wind farms," *DTU Wind Energy*, 2012.

[7]   N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker, "Surrogate-based analysis and optimization," *Prog. Aerosp. Sci.*, vol. 41, no. 1, pp. 1–28, 2005.

[8]   C. Rusu and V. Rusu, "Radial basis functions versus geostatistics in spatial interpolations," *IFIP Int. Fed. Inf. Process.*, vol. 217, no. 1, pp. 119–128, 2006.

[9]   R. F. Gunst, "Response Surface Methodology: Process and Product Optimization Using Designed Experiments," *Technometrics*, vol. 38, no. 3, pp. 284–286, 1996.

[10]   E. Svenning, "Implementation of an actuator disk in OpenFOAM," pp. 1–39, 2010.

[11]   A. (DEWI) Javaheri and B. (DEWI) Cañadillas, "Wake Modeling of an Offshore Wind Farm Using OpenFOAM," *Dewi Mag.*, no. 43, pp. 15–22, 2013.

[12]   S. Kalvig, E. Manger, and B. Hjertager, "Comparing different CFD wind turbine modelling approaches with wind tunnel measurements," *J. Phys. Conf. Ser.*, vol. 555, no. 1, p. 012056, 2014.

[13]   Tin Kam Ho, "Random decision forests," *Proc. 3rd Int. Conf. Doc. Anal. Recognit.*, vol. 1, pp. 278–282.

[14]   L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[15]   J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," *Cvpr 2011*, pp. 1297–1304, 2011.

[16]   Scikit learn, "3.2.4.3.2. sklearn.ensemble.RandomForestRegression." [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html. [Accessed: 22-Feb-2018].

[17]   C. E. Rasmussen and C. K. I. Williams, *Gaussian Process For Machine Learning*. 2006.

[18]   C. B. Do and H. Lee, "Section notes 9 - Gaussian Processes," *Lect. notes*, pp. 1–14, 2008.

[19]   R. A. Davis, "Gaussian Processes," *Encycl. Environmetrics Stoch. Model. Environ. Chang.*, p. 6, 2006.

[20]   scikit learn, "sklearn.gaussian_process.GaussianProcessRegression." [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html#sklearn.gaussian_process.GaussianProcessRegressor. [Accessed: 22-Feb-2018].

[21]   V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[22]   P.-H. Chen, R.-E. Fan, and C.-J. Lin, "A study on SMO-type decomposition methods for support vector machines.," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 893–908, 2006.

[23]   S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification*, vol. 36. 2016.

[24]   scikit learn, "sklearn.svm.SVR." [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html. [Accessed: 22-Feb-2018].

[25]   Q. Zhu and H. S. Lin, "Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes," *Pedosphere*, vol. 20, no. 5, pp. 594–606, 2010.

[26]   M. Mondrago, "Probabilistic Modelling of Geotechnical Conditions for Offshore Wind Turbine Support Structures," Cranfield University, 2014.

[27]   N. Gatta, "Reliability analysis of coal-fired power plant via Surrogate Modelling," Cranfield University, 2015.

[28]   F. Rosenblatt, "The Perceptron a Perceiving and Recognizing Automation," 1957.

[29]   MathWorks, "Neural Network Toolbox." [Online]. Available: https://uk.mathworks.com/products/neural-network.html. [Accessed: 22-Feb-2018].

[30]   "Palisade @risk." [Online]. Available: http://www.palisade.com/risk/. [Accessed: 09-Mar-2018].