Analyzing the temporal variation of wind turbine responses using Gaussian Mixture Model and Gaussian Discriminant Analysis

J. Park[1], K. Smarsly[2], K. H. Law[1] and D. Hartmann[3]

[1]*Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA; email: {jkpark11, law}@stanford.edu*

[2]*Department of Civil Engineering, Bauhaus University Weimar, Coudraystr. 7, 99423 Weimar, GERMANY; email: kay.smarsly@uni-weimar.de*

[3]*Department of Civil and Environmental Engineering, Ruhr University Bochum, Universitätsstr. 150, 44780 Bochum, GERMANY; email: hartus@inf.bi.rub.de*

ABSTRACT

Site- and time-specific wind field characteristics have a significant impact on the structural response and the lifespan of wind turbines. This paper presents a machine learning approach towards analyzing and predicting the response of a wind turbine structure to diurnal and nocturnal wind fields. Machine learning algorithms are applied (i) to better understand the changes of wind field characteristics due to atmospheric conditions, and (ii) to gain insights into the wind turbine loads being affected by the wind field. Using Gaussian Mixture Model, the variations in wind field characteristics are

investigated by comparing the joint probability density functions of selected wind field features. The wind field features are constructed from long-term monitoring data taken from a 500 kW wind turbine in Germany that is used as a reference system. Furthermore, employing Gaussian Discriminant Analysis, representative daytime and nocturnal wind turbine loads are compared and analyzed.

1. INTRODUCTION

Stochastic wind effects can cause fluctuations in the wind turbine responses such as loads, displacement, fatigue damages and power outputs. However, wind flow is a complex phenomenon and is difficult to characterize. Even for a given site, the wind flow characteristics, such as wind speed, wind shear profile and turbulence, vary with the time of day. The variation in the wind field during the diurnal and the nocturnal periods can affect the loading on a wind turbine (Kelley *et al.*, 1999, 2004; Hand *et al.*, 2003; Park *et al.*, 2013a). With the data collected on a wind turbine, this paper presents a machine learning approach to analyze the effect of the wind field on wind turbine responses.

Typically, recorded site-specific wind field data are described by time-averaged statistics, such as mean wind speed, mean wind direction, turbulence intensity and power exponent quantifying the vertical profile of the mean wind speed. In addition to the mean statistics, the distributions of such wind parameters are important factors in evaluating the wind turbine responses. Currently, the wind turbine design standard of the International Electrotechnical Commission (IEC 61400-1, 2005) uses the Weibull probability

distribution to describe 10-min averaged wind speed. To more accurately model wind speed with distribution more than a single mode, bimodal probability distribution function (Jaramillo *et al.*, 2004) and a mixture of probability distribution functions (Chang *et al.*, 2010; Kollu *et al.*, 2012) have been proposed. Furthermore, the current standard does not describe the temporal variations of the turbulence intensity level and the vertical wind profile, which can have an effect on the wind turbine response (Park, *et al.*, 2013). In this study, we describe time-specific wind field characteristics by using multivariate probability density functions (PDFs), in which the relationships among wind field statistical parameters are more explicitly described. We implement Gaussian Mixture Model (GMM), due to its capability of representing complex multi-modal probability density functions, to construct the PDFs for the wind parameters and to study how wind field characteristics vary between the diurnal and the nocturnal periods.

Analyzing wind turbine responses is difficult due to the complex interactions between a wind turbine structure and the stochastic wind flow. It is even more difficult to simulate the realistic wind fields that can be used as inputs for wind turbine analysis. Current standard does not specify the time specific wind field statistical parameters (IEC 61400-1, 2005). Wind field simulators that are based on the current standard are not able to generate time dependent wind field. Without the description of time dependent wind fields, wind turbine analysis codes are not able to analyze time dependent wind turbine responses. To more accurately simulate the time dependent wind field, Large-Eddy Simulation (LES) has been used (Park, *et al.*, 2013). However, realistic simulation of wind field based on LES is computationally expensive. With recent advances in sensor

technologies, however, a wind turbine can be continuously monitored, and the sensor data can be used to gain a better understanding on the dynamic responses of the wind turbine structure. Using the measured wind field as input data and the wind turbine response as output data, the input-output relationship can be studied by applying supervised machine learning techniques. In this study, we investigate the influence of wind fields on wind turbine response by constructing the response classification models using Gaussian Discriminant Analysis. In addition, we elucidate the reason for different levels of wind turbine response at day and night time periods by relating the time-specific PDFs for the wind features and the wind turbine response classification functions.

This paper is organized as follows: First, an integrated life-cycle management (LCM) framework implemented for the 500 kW wind turbine is presented. The overall framework to analyze wind turbine responses during diurnal and nocturnal periods due to the variation in the wind field characteristics is then discussed. The basic features of Gaussian Mixture Model (GMM) are introduced and the application of GMM to construct the multivariate probability density functions for the collected wind field data is presented. Furthermore, the basic features of Gaussian Discriminant Analysis (GDA) are discussed and the application of GDA to study wind field effects on wind turbine responses is presented. The two models obtained from GMM and GDA are then combined to study the temporal variation of wind turbine response. The paper is concluded with a brief summary and discussion.

2. WIND TURBINE LIFE-CYCLE MANAGEMENT FRAMEWORK

This study is based on the monitoring data collected on a 500 kW wind turbine located in Dortmund, Germany (see Figure 1). The wind turbine has a hub height of 65 m and upwind rotor diameter of 40.3 m. A life cycle management (LCM) framework has been developed for the monitoring and operational management of the wind turbine. The LCM framework provides valuable information that can be used not only to monitor the integrity of the wind turbine structure, but also to support new research and to test algorithmic developments. The LCM framework – details can be found in Smarsly *et al*. (2012a,b,c, 2013a,b) and Hartmann *et al*. (2011) – consists of several interconnected components that are installed at geographically distributed locations:

- A *structural health monitoring (SHM) system* comprises of sensors, data acquisition units as well as a local computer for recording structural, environmental, and operational wind turbine data.

- A *decentralized software system*, installed on different computers at the Institute for Computational Engineering (ICE) in Bochum (Germany), is remotely connected to the SHM system to continuously collect, process, and archive the sensor data.

- An agent-based *self-diagnostic system* for detecting sensor malfunctions is implemented to ensure reliability and availability of the SHM system.

- A *model updating component* is included to continuously update the finite element wind turbine models and to support damage detection analyses based on continuously updated monitoring data.

- A *management module* is incorporated to enable online life-cycle management through remote data access and analyses of the structural, environmental, and operational conditions of the wind turbine.

- A 32-node *PC cluster*, also located at the ICE in Bochum, is integrated into the LCM framework to provide high-performance parallel computing capabilities for computationally intensive calculations.

2.1 Structural health monitoring system

The SHM system, representing an essential component of the LCM framework, is installed inside and outside of the steel tower as well as at the concrete foundation of the wind turbine structure (Lachmann *et al*. 2009). The SHM system consists of a network of sensors, data acquisition units (DAUs) and an on-site server located in the maintenance room of the wind turbine. Six three-dimensional accelerometers are mounted on five different levels on the inner surface of the wind turbine tower. The sensitivity of the accelerometers is 700 mV/g with a measurement range of ± 3.0 g and a frequency range between 0 and 100 Hz. Three additional accelerometers are placed at the foundation of

the wind turbine. Since very small accelerations are expected at the foundation, single-axis piezoelectric seismic ICP accelerometers with a sensitivity of 10,000 mV/g are used; their measurement range amounts to ± 0.5 g. Furthermore, six inductive displacement transducers, having a range of 5 mm, are mounted inside the tower at 21 m and 42 m height. To determine temperature effects on the displacement measurements, the inductive displacement transducers are complemented by RTD surface sensors Pt100 capable of measuring temperature from –60 °C to +200 °C. Additional temperature sensors are placed inside and outside the tower to gauge temperature gradients.

For measuring wind speed, wind direction and air temperature, two anemometers are deployed. The first one, a cup anemometer that is directly connected to the integrated supervisory control and data acquisition (SCADA) system of the wind turbine, is installed on top of the nacelle at 67 m height (Figure 1). The second one, a three-dimensional ultrasonic anemometer, is mounted on a telescopic mast adjacent to the wind turbine at 13 m height; it continuously monitors the horizontal and vertical wind speed (0...60 m/s), the wind direction (0...360°) and the air temperature (–40...60°C). In addition to the structural and environmental sensor data collected by the DAUs, operational wind turbine data, such as power production and rotational speed, is taken directly from the internal SCADA system of the wind turbine.

All recorded data sets are continuously forwarded from the data acquisition units to the on-site server for temporary storage and periodic backups. Using a permanently installed DSL connection, the on-site server transfers the monitoring data to the decentralized

software system of the LCM framework. All data, after being synchronized and aggregated, is stored in a monitoring database being remotely available for further processing (Smarsly and Hartmann 2009a, 2009b, 2010).

2.2 Data sets used in this study

"Learning" from the existing input-output patterns, the machine learning approaches to be discussed employ various inputs (i.e. wind field statistics) and the corresponding outputs (i.e. structural response data) in order to understand the wind characteristics and to make reliable predictions on the structural wind turbine behavior.

As input, wind field statistics are computed from the monitoring data stored in the database of the LCM framework using remote connections (Figure 2). Specifically, the wind field feature vector $\boldsymbol{x} = (x_1, x_2, x_3) = (\overline{U}_{67}, \overline{U}_{13}, (Q_3 - Q_1)_{13})$ consists of three statistical features, where $\overline{U}_{67}$ is the mean of a 20-minute wind speed time series recorded at 67 m height by the cup anemometer, $\overline{U}_{13}$ is the mean of a 20-minute wind speed time series obtained at 13 m height by the ultrasonic anemometer, and $(Q_3 - Q_1)_{13}$ is the interquartile, the difference between the 75 percentile and the 25 percentile for the data distribution over the 20-minute wind speed time series recorded at the 13 m height. Since the monitoring data stored in the database system includes only the statistical mean and quartile data for each 20-minute period, the interquartile is used in this study as a measure to quantify the level of turbulence. The turbulence level is usually quantified by the standard deviation of wind speed time series (IEC 61400-1, 2005; Pope, 2004). Assuming

the wind speed time series data follows a Gaussian distribution, the interquartile value is a reasonable measure analogous to the standard deviation of the wind speed time series. In wind engineering, 10-min time window is widely used to estimate the statistical parameters of wind field (IEC 61400-1, 2005). The time window of 20 minutes for the recorded data is sufficient to ensure that wind speed time series is stationary for estimating the statistical parameters.

For the output response, the acceleration of the tower recorded by one of the accelerometers located at 21 m height is used. Specifically, the interquartile values for the 20-minute acceleration time series are employed as an indirect measure for the averaged magnitude of the wind turbine tower acceleration. Assuming the acceleration time series are of zero or constant mean values, the interquartile value is a reasonable measure analogous to the RMS value of the acceleration time series, which is commonly used for quantifying the averaged magnitude of acceleration response of on a structure. For classification, the interquartile values are evenly divided and categorized into N_C discretized classes based on their magnitude; a response class is denoted by y_C, $y_C \in \{1, \dots, N_C\}$. In total, 7,960 pairs of $\boldsymbol{x}$ (input feature vector) and y_C (the interquartile output class), provided by the LCM framework, serve as the basis for this study.

3. A FRAMEWORK FOR ANALYZING TEMPORAL WIND FIELD EFFECTS ON WIND TURBINE RESPONSES

Atmospheric conditions vary with the time of day and affect wind turbine response due to the changes in the wind field characteristics. Figure 3 depicts the overall probabilistic framework designed to understand the temporal variation of wind turbine response. First, we strive to understand how wind field characteristics vary with the time of day. We then study the effect of wind field on wind turbine response. Finally, we investigate wind turbine response due to temporal variation of wind field. The basic methodologies can be summarized as follows:

- **Characterization of wind fields:** We describe the time-specific wind field characteristics by constructing the probability density function (PDF), $p(\boldsymbol{x}|t_P)$, for wind field feature vector $\boldsymbol{x}$ given time period t_P. To construct $p(\boldsymbol{x}|t_P)$, we apply Gaussian Mixture Model (GMM) to the wind monitoring data.

- **Wind turbine load classification:** We describe the underlying relationship between the wind field and the wind turbine response by constructing the probability distribution, $P(y_C|\boldsymbol{x})$, for the wind turbine response class y_C given the wind input feature $\boldsymbol{x}$ using Gaussian Discriminant Analysis. The probability distribution $P(y_C|\boldsymbol{x})$ describes how a certain level of wind turbine response is likely caused by certain wind field characteristics.

- **Temporal variation of wind turbine responses:** The changes in the wind field characteristics over different times of day would induce different wind turbine responses. We describe the influence of time on wind turbine response class by constructing the probability distribution, $P(y_C|t_P)$, of the wind turbine response class y_C for a given time period t_P. The probability distribution $P(y_C|t_P)$ describes how a certain level of wind turbine response is likely occur at the time period t_P. The two probabilistic models, $p(\boldsymbol{x}|t_P)$ and $P(y_C|t_P)$, obtained using the GMM and GDA analyses are employed to construct $P(y_C|t_P)$.

The following sections describe in detail the methodologies and their applications to study wind field characteristics, wind turbine load classification and the variation of wind turbine response between diurnal and nocturnal periods.

4. CHARACTERIZATIONS OF WIND FIELDS

We characterize time dependent wind field characteristics using multivariate probability density functions constructed by Gaussian mixture model (GMM). GMM is used, in this study, because of its capability of constructing multivariate probability distribution with multi-modes from unsupervised data set. Gaussian mixture model (GMM) can be regarded as a statistical unsupervised learning to extract the underlying hidden structure by estimating the probability density of features from a statistical data set. GMMs have been widely used in computer vision and speaker recognition (Stauffer and Grimson 1999; Reynolds *et al*., 2000). GMMs have also been applied to describe the stochastic load

variations in a distributed energy grid (Singh *et al*., 2010; Gonzalez-Longatt *et al*., 2012). As a case study, GMM is applied to the wind monitoring data collected during the diurnal and the nocturnal periods to construct the time specific PDFs. The constructed PDFs are compared to understand the difference between the diurnal and the nocturnal wind field characteristics.

4.1 Gaussian Mixture Model

A Gaussian mixture model (GMM) can be regarded as a multivariate probability density function written in terms of the weighted sum of the Gaussian probability density functions. For the statistical wind input feature vector $\boldsymbol{x} = (x_1, x_2, x_3) = (\overline{U}_{67}, \overline{U}_{13}, (Q_3 - Q_1)_{13})$, the multivariate PDF $p(\boldsymbol{x}|t_P)$ given the time period t_P is expressed in terms of a linear combination of K probability density functions:

$$p(\boldsymbol{x}|t_P) = \sum_{j=1}^{K} \phi_j p_j(\boldsymbol{x}|t_P) \tag{1}$$

where $p_j(\boldsymbol{x}|t_P)$ is the Gaussian probability density fucntion (GPDF) for the jth mixture component of the GMM at time period t_P and ϕ_j, $\sum_{j=1}^{K} \phi_j = 1$, is the corresponding mixture weight. The jth GPDF $p_j(\boldsymbol{x}|t_P)$ can be fully described by its mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$ (Reynolds *et al*., 2000):

$$p_j(\boldsymbol{x}|t_P) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_j|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\right) \quad (2)$$

where n is the dimension of the input feature vector $\boldsymbol{x}$. To construct the PDF $p(\boldsymbol{x}|t_P)$, the parameters $\boldsymbol{\phi} = \{\phi_1, \dots, \phi_K\}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K\}$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K\}$ are determined from the measurement (training) data set $\boldsymbol{x}^{(i)} = \left(\overline{U}_{67}^{(i)}, \overline{U}_{13}^{(i)}, (Q_3 - Q_1)_{13}^{(i)}\right), i = 1, \dots, m_u$, where $\boldsymbol{x}^{(i)}$ is the ith input feature vector and m_u denotes the size (number of input feature vectors) of the training data set.

Assuming the input vectors $\boldsymbol{x}^{(i)}, i = 1, \dots, m_u$, in the training data set are independent, the parameters $\boldsymbol{\phi}, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be estimated as the values that maximize the following log-likelihood estimation from the measured wind input feature data as:

$$\begin{aligned} l(\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \ln \prod_{i=1}^{m_u} p(\boldsymbol{x}^{(i)};\ \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \sum_{i=1}^{m_u} \ln p(\boldsymbol{x}^{(i)};\ \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \sum_{i=1}^{m_u} \ln\left(\sum_{z^{(i)}=1}^{K} p\ (\boldsymbol{x}^{(i)}, z^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\phi})\right) \\ &= \sum_{i=1}^{m_u} \ln\left(\sum_{z^{(i)}=1}^{K} p(\boldsymbol{x}^{(i)}|z^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(z^{(i)}; \boldsymbol{\phi})\right) \end{aligned} \quad (3)$$

where $p\left(\boldsymbol{x}^{(i)}; \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ is the probability of the input feature $\boldsymbol{x}^{(i)}$ given the parameters, $\boldsymbol{\phi}, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The latent random variable $z^{(i)}$ specifies one of the K possible Gaussian

probability distributions from which $\boldsymbol{x}^{(i)}$ is drawn. Because $z^{(i)}$ is a random variable individually assigned to each $\boldsymbol{x}^{(i)}$, it follows different probability distributions over the K possible GPDFs. If $z^{(i)}$ is explicitly known a priori, the summation term $\sum_{z^{(i)}=1}^{K}(\cdot)$ over $z^{(i)}$ in Eq. (3) would result in a single Gaussian term. In this case, Eq. (3) is a quadratic (i.e., a concave) function expressed in terms of the GMM parameters, $\boldsymbol{\phi}, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which can be analytically obtained. However, since the latent variable $z^{(i)}$ is unknown, Eq. (3) cannot be explicitly defined and the GMM parameters are difficult to estimate. In this study, we employ the iterative expectation maximization (EM) algorithm, which has been widely used to estimate statistical parameters given the existence of the latent variables (Dempster *et al.*, 1976; Render *et al.*, 1984).

The key principle of the EM algorithm is that if the latent variable $z^{(i)}$ or its distribution is known, the loglikelihood function as shown in Eq. (3) can be maximized. Therefore, the EM algorithm consists of two iterative steps:

- "E-step" estimates the distribution of $z^{(i)}$ for each feature vector $\boldsymbol{x}^{(i)}$ given the parameters $\boldsymbol{\phi}, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ estimated in the M-step.
- "M-step" optimizes the parameters $\boldsymbol{\phi}, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by maximizing the loglikelihood function given the the distribution of $z^{(i)}$ obtained in the E-step.

It shuld be noted that in the M-step, the maximization is performed with the estimated distribution over $z^{(i)}$ rather than a single fixed value of $z^{(i)}$. As a result, the summation term $\sum_{z^{(i)}=1}^{K} p\,(\boldsymbol{x}^{(i)}, z^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\phi})$ in Eq. (3) involves more than one exponential functions

(Gaussian PDFS); thereby, the loglikelihood function is not necessarly concave. To minimize the possibility of local optimial solutions from gredient-based method, we repeat the optimization procedure with different initial points (i.e., different initial GMM parameters) using MATLAB's statistical learning tool box. The GMM parameters producing the largest logliklihood values are selected and used to construct $p(\boldsymbol{x}|t_P)$.

4.2 Application of GMM to wind field characterization

To study the diurnal and nocturnal wind field characteristics at the wind turbine site, we construct the PDFs from the wind data collected from the anemometers. In this study, we set the time period $t_P = 1$ for 6:00 AM ~ 6:00 PM (diurnal) and $t_P = 2$ for 6:00 PM ~ 6:00 AM (nocturnal). A total number of 7,960 input feature vectors, measured for a period of 110 days, are categorized into the two time periods. Two data sets (each with 3,980 feature vectors) are then used to construct the time specific PDFs, $p(\boldsymbol{x}|t_P = 1)$ and $p(\boldsymbol{x}|t_P = 2)$. The wind input feature data for the two time periods are plotted as shown in Figure 4. It can be seen that the wind field characteristics for the two time periods are quite different because of the difference in the temperature lapse rate over height, for the two time periods. The wind speeds and the levels of turbulence are usually higher during the day due to more active airflow mixing than during the night (Stull, 1988).

Figure 5 shows the PDFs, $p(\boldsymbol{x}|t_P = 1)$ and $p(\boldsymbol{x}|t_P = 2)$, constructed by applying the GMM to the diurnal and the nocturnal wind data sets, respectively. In this study, 3

GPDFs are used to construct the PDFs $p(\boldsymbol{x}|t_P = 1)$ and $p(\boldsymbol{x}|t_P = 2)$. The number of GPDFs is determined such that the Akaike information criterion (AIC) (Akaike, 1974), that quantifies the tradeoff between the complexity of a statistical model and the fitness of the statistical model to data, is maximized. The shapes and the densities of the PDFs, as shown in Figure 5, depict the different wind field characteristics for the two time periods. In Figure 5, the regions with darker color correspond to the regions with lots of measurement data points in Figure 4; as a result, the regions also show a higher probability density. As illustrated in the figure, using only a small number of GMM parameters, the constructed PDFs can effectively describe the density of the wind input feature data, distribution of wind field features and the correlations among the different wind field features.

With the multivariate probability density function $p(\boldsymbol{x})$ for the three wind field input features, $\overline{U}_{67}$, $\overline{U}_{13}$ and $(Q_3 - Q_1)_{13}$, three marginal PDFs, $p(\overline{U}_{67}, \overline{U}_{13})$, $p(\overline{U}_{13}, (Q_3 - Q_1)_{13})$ and $p((Q_3 - Q_1)_{13}, \overline{U}_{67})$, can be computed in that $p(\boldsymbol{x}\backslash x_k)$ is obtained by integrating $p(\boldsymbol{x})$ with respect to the feature variable x_k. Figure 6 shows the marginal PDFs computed from $p(\boldsymbol{x}|t_P = 1)$ and $p(\boldsymbol{x}|t_P = 2)$. Based on the marginal PDFs, we can gain better understanding of the relationships between each pair of wind statistical parameters. It can be seen that positive correlations exist between the features, although the levels of correlation differ slightly between the day and the night times.

Analogously, we can obtain a univariate PDF $p(x_i)$ for a specific feature variable by integrating $p(\boldsymbol{x})$ with respect to all other feature variables. The probability mass function for a feature variable can then be computed by integrating the univariate PDF over a

discrete interval. Using 30 discretized intervals for each input feature, the discrete probability mass functions, $P(\overline{U}_{67})$, $P(\overline{U}_{13})$ and $P((Q_3 - Q_1)_{13})$, are estimated and plotted as continuous line in Figure 7. That is, the univariate probability mass functions (PMFs) are generated from the multivariate PDF constructed by applying GMM to the wind feature training data set. Comparing the discrete PMF computed by counting the frequency of occurrence directly from the wind data (shown as bar in Figure 7), we can observe that the statistical trends predicted from the GMM are in good agreement with those described by the data.

The univariate PMFs can also be derived individually by applying GMM to each wind feature. As shown in Figure 7, the PMF obtained from the multivariate PDF (3-D GMM) and the PMF from the independent univariate PDF (1-D GMM) are in good agreement. This result indicates that the high dimensional PDFs constructed appear to preserve the statistical trend in the wind input features. In addition, the multivariate PDFs can provide valuable information about the correlation among the wind field features.

5. WIND TURBINE LOAD CLASSIFICATIONS

In this study, we attempt to understand the underlying relationship between the wind field features and the wind turbine response by constructing the response classification function using Gaussian discriminant analysis (GDA). Gaussian Discriminant Analysis (GDA), first proposed by Fisher (1936) as statistical classification analysis, has been widely adopted as a supervised learning algorithm to extract the underlying relationships

between the input (feature vectors) and the output (classes) data. If the input features follow Gaussian distribution, GDA requires the least number of training data to find the parameters for a statistical classification model and has better classification accuracy compared to other non-linear classification algorithms (Pohar *et al.* 2004). Observing from the monitoring data that the distribution of the wind features resembles a Gaussian distribution, GDA is selected in this study to construct the response classification (prediction) function using the wind field features and wind turbine response (training data) collected from the 500KW wind turbine.

5.1 Gaussian Discriminant Analysis

Wind turbine load classification function $g_C(\boldsymbol{x})$ maps the wind input features $\boldsymbol{x} = (\overline{U}_{67}, \overline{U}_{13}, (Q_3 - Q_1)_{13})$, as defined previously, to the output response class $y_C \in \{1, \dots, N_C\}$, where N_C is the number of classes that categorize the different levels of output response feature. In this study, the output response feature is the interquartile values (i.e. the difference between the 75 percentile and the 25 percentile) for a 20-minute acceleration time series recorded by the accelerometer at a 21 m height inside the wind turbine tower. Analogous to the RMS measure, the interquartile is used as measure to quantify the average magnitude of the tower acceleration. Each interquartile data value is then assigned to an output response class y_C.

Using m_s pairs of input and output (training) data, $\left(\boldsymbol{x}^{(i)}, y_C^{(i)}\right), i = 1, \dots, m_s$, GDA first constructs the probability density function $p(\boldsymbol{x}|y_C)$ for the input feature vector $\boldsymbol{x}$

conditional on the given output response class y_C. Assuming multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, the PDF for the jth output class $p(\boldsymbol{x}|y_C = j),\ j = 1, \dots, N_C$, can be represented as (Hastie *et al.* 2008):

$$p(\boldsymbol{x}|y_C = j) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_j|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_j)\right) \tag{4}$$

where n is the number of features in $\boldsymbol{x}$, i.e. the dimension of $\boldsymbol{x}$. Furthermore, the prior distribution on the response class y_C is expressed as a multinomial distribution, i.e. $P(y_C = j;\ \boldsymbol{\varphi}) = \varphi_j$, where φ_j is simply the ratio between the number of data points for class $y_C = j$.

To obtain the mapping function $g_C(\boldsymbol{x})$ between the input and output features, we construct $p(\boldsymbol{x}|y_C)$ and $P(y_C)$ by estimating the parameters, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{N_C}\}$, $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_{N_C}\}$ and $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_{N_C}\}$, for the N_C conditional distributions and the priors. The parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\varphi}$ can be determined by maximizing the log-likelihood function over the m_s pairs of input and output (training) data sets (Hastie *et al.* 2008):

$$\begin{aligned} l(\boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \sum_{i=1}^{m_s} \ln p(\boldsymbol{x}^{(i)}, y_C^{(i)}; \boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \sum_{i=1}^{m_s} \ln\ p(\boldsymbol{x}^{(i)}|y_C^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) P(y_C^{(i)}; \boldsymbol{\varphi}) \end{aligned} \tag{5}$$

The response class $y_C^{(i)}$ denotes the specific PDF from which $x^{(i)}$ is drawn. As the class $y_C^{(i)}$ is explicitly given from the training data, Eq. (5) can be expressed in terms of the previously defined $p(\boldsymbol{x}|y_C)$ and $P(y_C)$. In Eq. (5), since $\ln p(\boldsymbol{x}^{(i)}, y_C^{(i)}; \boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a quadratic function, the loglikelihood function is also quadratic (concave). That is, the loglikelihood function can be analytically optimized with respect to the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\varphi}$ (Hastie *et al*; 2008).

Once the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\varphi}$ for $p(\boldsymbol{x}|y_C)$ and $P(y_C)$ are determined, the posterior distribution on the response class y_C given the wind input feature $\boldsymbol{x}$ is modeled according to the Bayes' rule as follows:

$$P(y_C|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|y_C)P(y_C)}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|y_C)P(y_C)}{\sum_{y_C} p(\boldsymbol{x}|y_C)P(y_C)} \tag{6}$$

Therefore, for the new wind input feature vector $\boldsymbol{x}_{new}$ representing the newly observed wind field characteristics, the corresponding *estimated* response class $\hat{y}_C \in \{1, \ldots, N_C\}$ can be determined according to the maximum a posteriori detection (MAP) principle expressed as:

$$\begin{aligned} \hat{y}_C = g_C(\boldsymbol{x}_{new}) &\stackrel{\text{def}}{=} \underset{y_C}{\operatorname{argmax}} P(y_C|\boldsymbol{x}_{new}) \\ &= \underset{y_C}{\operatorname{argmax}} p(\boldsymbol{x}_{new}|y_C)P(y_C) \end{aligned} \tag{7}$$

The response classification function $g_C(\boldsymbol{x})$ divides the domain of the input feature $\boldsymbol{x}$ (in a 3D space) into N_C regions, each of which corresponds to a different output response class. We can construct the boundary between the two neighboring classes i and j by setting $P(y_C = i|\boldsymbol{x}) = P(y_C = j|\boldsymbol{x})$ in Eq. (7), which leads to a quadratic equation in terms of the input feature $\boldsymbol{x}$ as follows (Hastie *et al.* 2008):

$$\boldsymbol{x}^T\left(\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_j^{-1}\right)\boldsymbol{x} - 2\left(\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\mu}_j^T\boldsymbol{\Sigma}_j^{-1}\right)\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \ln\frac{|\boldsymbol{\Sigma}_i|}{|\boldsymbol{\Sigma}_j|} = 0 \qquad (8)$$

Eq. (8) represents a quadratic surface that separates two neighboring regions (classes). Note that if the covariance matrices are assumed to be the same for all the classes, i.e. $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ for $i = 1, \ldots, N_C$, Eq. (8) becomes a linear plane surface and the problem is generally known as linear Discriminant analysis.

5.2 Application of GDA to the wind turbine response classification problem.

We construct the wind turbine response classification function by applying GDA to a total of 7,960 pairs wind input feature vectors and output response data collected over 110 days. Because the number of output response classes, N_C, and the size of the training data set, m_t, can affect the accuracies of the response classification function $g_C(\boldsymbol{x})$ shown in Eq. (7), we construct the response classification functions using different combinations of N_C and m_t and compare their accuracy ratios. To define different

number of classes N_C, the interquartile values of wind turbine tower acceleration time series are evenly divided and categorized into N_C discretized classes based on their magnitudes, which range from 0 to 19 mg.

Figure 8 shows the accuracy ratios of the wind turbine response classification function $g_C(\boldsymbol{x})$ with different number of response classes and different size of the training data set. The accuracy ratio is calculated as the ratio between the number of correct predictions by $g_C(\boldsymbol{x})$ and the size of the test data set, m_t. Altogether, 1,000 pairs ($m_t = 1{,}000$) of the input feature vector $\boldsymbol{x}$ and the output response class y_C, which do not belong to the training data set, are used to calculate the accuracy ratio for the classification by $g_C(\boldsymbol{x})$. As shown in Figure 8, it can be seen that the accuracy ratio decreases as the number of the response classes increases. On the other hand, for a given number of response classes, the accuracy ratio rarely changes with the size of the training data set. As the size of training data set exceeds the minimum number of data points necessary to estimate the covariance matrix for each class, the accuracy ratio appears to remain the same, which indicates the effectiveness of GDA in terms of the training data size.

Figure 9 shows the data points for the input feature vectors and the corresponding output response classes for an example case of $N_C = 5$ and $m_t = 3{,}980$. As shown, five response classes are sufficient to distinguish the general difference in the magnitudes of the wind turbine tower acceleration. The ranges of the interquartile values corresponding to $N_C = 5$ are shown in Table 1. In addition, as shown in Figure 8, 3,980 data points provide reasonably high accuracy ratio for the classification (78 ~ 80 %). The boundary

surfaces discriminating two neighboring classes as determined by Eq. (8) are also plotted in Figure 9. It can be seen that the quadratic boundary surfaces discriminate the response classes reasonably well.

6. TEMPORAL VARIATION OF WIND TURBINE RESPONSES

As discussed, the diurnal and nocturnal wind field characteristics are investigated by constructing multivariate probability density functions via GMM, while the influences of the wind fields on the wind turbine response are studied by constructing classification functions with the aid of GDA. Combining these two independently constructed models, we can establish insightful information by establishing a probabilistic framework to study the response of the wind turbine during the day and the night times. Specifically, we compare the probability distribution and the expected value of the wind turbine response classes corresponding to the diurnal and the nocturnal winds.

6.1 Calculating class probabilities and expected vlaues

Using chain rule, the joint probability distribution between the wind input feature $\boldsymbol{x}$ and the output response class y_C for a given time period t_P can be represented as

$$p(\boldsymbol{x}, y_C|t_P) = p(\boldsymbol{x}|t_P)P(y_C|\boldsymbol{x}, t_P) \tag{9}$$

Since the time information t_P is not required to estimate the wind turbine response class y_C given the information about the wind field feature $\boldsymbol{x}$, y_C and t_P are conditionally independent given $\boldsymbol{x}$. Therefore, Eq. (9) can be rewritten as

$$p(\boldsymbol{x}, y_C|t_P) = p(\boldsymbol{x}|t_P)P(y_C|\boldsymbol{x}) \tag{10}$$

where the two probabilistic models, $p(\boldsymbol{x}|t_P)$ and $P(y_C|\boldsymbol{x})$, are constructed by GMM and GDA, respectively, as described in the previous sections. The marginal distribution of the wind turbine response class y_C for the time period t_P can be derived by integrating the joint probability distribution function over (i.e., marginalizing out) the wind field feature vector $\boldsymbol{x}$ as

$$\begin{aligned} P(y_C|t_P) &= \int_{\boldsymbol{x}} p(\boldsymbol{x}, y_C|t_P)d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}} p(\boldsymbol{x}|t_P)P(y_C|\boldsymbol{x})d\boldsymbol{x} \end{aligned} \tag{11}$$

The response class probability $P(y_C|t_P)$, which can be numerically computed, shows the distribution of the wind turbine response classes during the time period t_P.

The expected wind turbine response class for the time period t_P can be calculated by summing the wind turbine output class weighted by its probability as

$$\mathrm{E}[y_C|t_P] = \sum_{y_C=1}^{N_C} y_C \, P(y_C|t_P) \tag{12}$$

The expected wind turbine response class can then be used to compare the response magnitudes for the different time periods.

To study the wind turbine responses corresponding to the diurnal and the nocturnal wind fields, we construct the class probabilities $P(y_C|t_P)$ and the expected response classes $\mathrm{E}[y_C|t_P]$ for the wind turbine tower acceleration during the daytime $t_P = 1$ and during the nighttime $t_P = 2$. Figure 10 shows $p(\boldsymbol{x}|t_P = 1)$ and $p(\boldsymbol{x}|t_P = 2)$ overlapped with the load classification function $g_C(\boldsymbol{x})$, which divides the input feature domain into five different classes. For display purposes, the PDFs are represented by the contour surfaces with $p(\boldsymbol{x}|t_P) = 0.02$. As shown in Figure 10, the PDF for the daytime wind disperses more widely in the region associated with the high load classes ($y_C = 3, 4$ and 5) than the PDF for the nighttime wind. As a result, as shown in Figure 11, the diurnal wind fields have higher probabilities for classes 3, 4 and 5 but lower class probabilities for classes 1 and 2 than the nocturnal wind fields.

The class probabilities and the expected classes can be estimated from $p(\boldsymbol{x}|t_P)$ and $p(y_C|\boldsymbol{x})$ based on Eqs. (11) and (12). Using the available data, the true class probabilities and the expected classes corresponding to the diurnal and the nocturnal wind fields can also be directly calculated by counting the number of measurement data points in each

response class for the day and night time periods. Table 2 compares the class probabilities and the expected classes estimated from probabilistic models and calculated from the measurement data. In general, the estimated class probabilities compare very well with the true (measured) class probabilities with only slight deviation observed for the nocturnal case. The estimated expected class also matches very well with the expected classes calculated from the measurement data. As shown in Eq. (11), the effectiveness of prediction, is the result of both the expressivity of the multivariate, multinomial probability density function $p(\boldsymbol{x}|t_P)$ for wind field features and the accuracy of the response class probability model $P(y_C|\boldsymbol{x})$. It is worth noting that while accurately predicting the statistical trend in the wind turbine output response, the machine learning approach can also elucidate the reason for the statistical trend in the wind turbine output response by relating $p(\boldsymbol{x}|t_P)$ and $P(y_C|\boldsymbol{x})$. That is, the daytime wind fields induce the higher level of the wind turbine tower acceleration due to the higher mean wind speed and turbulence compared to the nighttime wind fields.

7. CONCLUSION

A machine learning approach has been presented to describe time-specific wind field characteristics, to predict the response of a wind turbine given wind field characteristics, and to study temporal variation of wind turbine response. The machine learning algorithms have been tested using the long-term monitoring data, collected by an integrated life-cycle management system installed at a wind turbine in Germany.

A Gaussian mixture model (GMM) approach is applied to the measured wind data to construct the probability density function for wind statistical parameters in an attempt to describe and understand the time specific characteristics of wind fields. We study the characteristics of the daytime and nighttime wind fields by comparing the PDFs constructed based on the diurnal and nocturnal wind data. The PDFs clearly show that the wind fields in the daytime are characterized by faster mean wind speeds and higher levels of fluctuations than the nighttime counterparts; this trend is effectively characterized by a multivariate probability density function for wind field statistical parameters. It is worth noting that the distribution of wind field features can possibly vary over long-term period, i.e., year, due to global climate changes. To address such issue in applying GMM to monitoring data, either an active or a passive remedy can be applied. In an active solution, we can incrementally update the GMM parameters upon receiving new wind monitoring data, which is know as an incremental Gaussian Mixture model. In a passive solution, we can apply GMM to monitoring data collected over different time periods. The variation in the distribution for wind field features then can be tracked by the changes in the GMM parameters.

A Gaussian discriminative analysis (GDA) has been applied to find the relationship between the wind input features and the response of the monitored wind turbine tower. In this study, three input features include the mean wind speeds at 13 m and 67 m height, and the interquartile values of the 20-minute wind speed time series measured at 13 m height. The output response is the interquartile values of 20-minute wind turbine tower acceleration time series. Although only a small number of input features are used, GDA

predicts the wind turbine response class reasonably well. If more wind statistical features are included to the input feature vector by measuring wind speed at additional locations, the prediction accuracy can further improve. In addition, it may be worth noting that although the interquartile values employed in this study is a reasonable measure analogous to RMS values of the acceleration time series, the standard deviation or RMS values could be more ideal features for quantifying the averaged magnitude of the acceleration or other types of wind turbine response. In the future, the wind turbine monitoring system will be updated to calculate and archive the standard deviations of wind speed and wind turbine response time series.

By combining the two probabilistic models constructed by GMM and GDA on the wind input features and wind turbine response class, useful information that are helpful to the operation of the wind turbine can be obtained. This study examines the class probabilities and the expected classes for the levels of wind turbine tower acceleration at daytime and nighttime and used them to elucidate the reasons for different wind turbine tower vibrational response. It becomes transparent that the expected response class for the daytime is higher than that for the nighttime because daytime wind fields have higher mean wind speeds and higher levels of turbulence, both of which are positively correlated to the level of wind turbine tower vibration. The results show that the probabilistic framework can extract the casual relationship among time, wind field characteristics and wind turbine response while providing the quantitative measure (probability) about the observations. This framework can also be used to find casual relationships that may not be obvious and difficult to observe in physical simulations. For instance, the

methodology may be used to study the possible influence of stable boundary layer occurring early in the morning on the fatigue damage of wind turbine blade if wind field data and the corresponding wind turbine response data related to fatigue measurements are available.

The statistical machine learning approach presented in this paper can also be used to study the influence of the daily, monthly, seasonal and yearly variations of wind field characteristics on wind turbine response. If we establish accurate wind turbine response classification function for a specific wind turbine model and construct a site specific probability density function for the wind input statistical features, the approach can potentially be expanded to evaluate how the wind turbine will respond at that site. In future research, we plan to develop an economical and practical monitoring methodology that uses wind input statistical features commonly collected at each wind turbine but employs structural responses measured from only one or a few representative wind turbines equipped with sensors.

ACKNOWLEDGEMENTS

This research is partially funded by the German Research Foundation (DFG) under grants SM 281/1-1, SM 281/2-1, and HA 1463/20-1. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the DFG.

REFERENCES

Akaike, H. (1974). "A new look at the statistical model identification," IEEE Transaction on Automatic Control, 19(6), pp.716–723.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). "Maxumum Likelihood from incomplete Data via the EM Algorithms," Journal of the Royal Statistics Society, Series B, 39(1), pp. 1-38.

Fisher, R (1936). "The use of multiple measurements in taxonomic problems," *Annal Eugen* (7), pp.179–188.

Gonzalez-Longatt, F. M., Rueda, J. L. and Erlich, I., Bogdanov, D. and Villa, W. (2012). "Identification of Gaussian Mixture Model using Mean Variance Mapping Optimization: Venezuelan Case," In: *Proceedings of IEEE PES Innovative Smart Grid Technologies Europe*, Berlin, Germany, October 14, 2012.

Hand, M. M., Kelley, N. D., and Balas, M. J., "Identification of Wind Turbine Responses to Turbulent Inflow Structures," Tech. Rep. NREL/CP-500-33465, National Renewable Energy Laboratory, Golden, Co, 2003.

Hastie, T, Tibshirani, R. and Friedman, J. (2008). "The elements of statistical learning," Springer.

Hartmann, D., Smarsly, K. and Law, K. H. (2011). "Coupling sensor-based structural health monitoring with finite element model updating for probabilistic lifetime estimation of wind

energy converter Structures," In: *Proceedings of the 8th International Workshop on Structural Health Monitoring 2011*. Stanford, CA, USA, September 13, 2011.

International Electotechnical Commission. (2005). "Wind Turbines - Part 1: Design Requirements," IEC 61400-1, Edition 3.0.

Jonkman, B.J. and Buhl, M.L. (2006). "TurbSim User's Guide" Technical Report NREL/TP-500-39797, National Renewable Energy Laboratory, Golden, Co.

Kelley, N. D., "A Case for Including Atmospheric Thermodynamic Variables in Wind Turbine Fatigue Loading Parameter Identification," Tech. Rep. NREL/CP-500-26829, National Renewable Energy Laboratory, Golden, Co, 1999.

Kelley, N., Shirazi, M., Jager, D., Wilde, S., Adams, J., Buhl, M., Sullivan, P., and Patton, E., "Lamar Low-Level Jet Project Interim Report," Tech. Rep. NREL/TP-500-34593, National Renewable Energy Laboratory, Golden, Co, 2004.

Lachmann, S., Baitsch, M., Hartmann, D. and Höffer, R. (2009). "Structural lifetime prediction for wind energy converters based on health monitoring and system identification," In: *Proceedings of the 5th European & African Conference on Wind Engineering*. Florence, Italy, July 19, 2009.

Park, J., Basu, S. and Manuel, L. (2013). "Large-eddy simulation of stable boundary layer turbulence and estimation of associated wind turbine loads," *Wind Energy*, doi: 10.1002/we.1580.

Pohar, M., Blas, M. and Turk, S. (2004). "Comparison of logistic regression and linear discriminant analysis: a simulation study," *Metodološki zvezki*, 1(1), pp.143-161.

Pope, B. P. (2000). "Turbulent Flows", Cambridge University Press, Cambridge, UK.

Render, A. R. and Walker, F. H. (1984). "Mixture densities, maximum likelihood and the eM algorithm," *SIAM Review*, 26(2), pp. 195-239.

Reynolds, A. D., Quatieri, F. T. and Dunn, B. R. (2000). "Speaker verification using adapted Gaussian Mixture Models," *Digital Signal Processing*, 10, pp. 19-41.

Singh, R., Pal, C. B. and Jabr, A. R. (2010). "Statistical representation of distribution system loads using Gaussian mixture model," *IEEE Transactions on Power Systems*, 25 (1), pp. 29-37.

Smarsly, K. and Hartmann, D. (2009a). "Real-time monitoring of wind turbines based on software agents," In: *Proceedings of the 18th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering*. Weimar, Germany, July 7, 2009.

Smarsly, K. and Hartmann, D. (2009b). "Multi-scale monitoring of wind energy plants based on agent technology," In: *Proceedings of the 7th International Workshop on Structural Health Monitoring*. Stanford, CA, USA, September 9, 2009.

Smarsly, K. and Hartmann, D. (2010). "Agent-oriented development of hybrid wind turbine monitoring systems," In: *Proceedings of the 2010 EG-ICE Workshop on Intelligent Computing in Engineering*. Nottingham, UK, June 30, 2010.

Smarsly, K., Law, K. H. and Hartmann, D. (2012a). "A multiagent-based collaborative framework for a self-managing Structural health monitoring system," *ASCE Journal of Computing in Civil Engineering*, 26(1), pp. 76-89.

Smarsly, K., Law, K. H. and Hartmann, D. (2012b). "Towards life-cycle management of wind turbines based on structural health monitoring," In: *Proceedings of the First International Conference on Performance-Based Life-Cycle Structural Engineering*. Hong Kong, China, December 5, 2012.

Smarsly, K., Hartmann, D. and Law, K. H. (2012c). "Integration of Structural Health and Condition Monitoring into the Life-Cycle Management of Wind Turbines," In: *Proceedings of the Civil Structural Health Monitoring Workshop*. Berlin, Germany, November 6, 2012.

Smarsly, K., Hartmann, D. & Law, K. H. (2013a). "A computational framework for life-cycle management of wind turbines incorporating structural health monitoring," *Structural Health Monitoring – An International Journal*, 12(4), pp. 359-376.

Smarsly, K., Hartmann, D. & Law, K. H. (2013b). "An integrated monitoring system for life-cycle management of wind Turbines," *International Journal of Smart Structures and Systems*, 12(2), pp. 209-233.

Stull, R. B. (1998). “An introduction to boundary layer meteorology,” Kluwer Academic Publishers.

Stauffer, C. and Grimson, W.E.L. (1999). “Adaptive background mixture models for real-time tracking,” In: *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Fort Collins, U.S.A., 23 Jun, 1999

Figure 1. Monitored wind turbine and anemometers used.

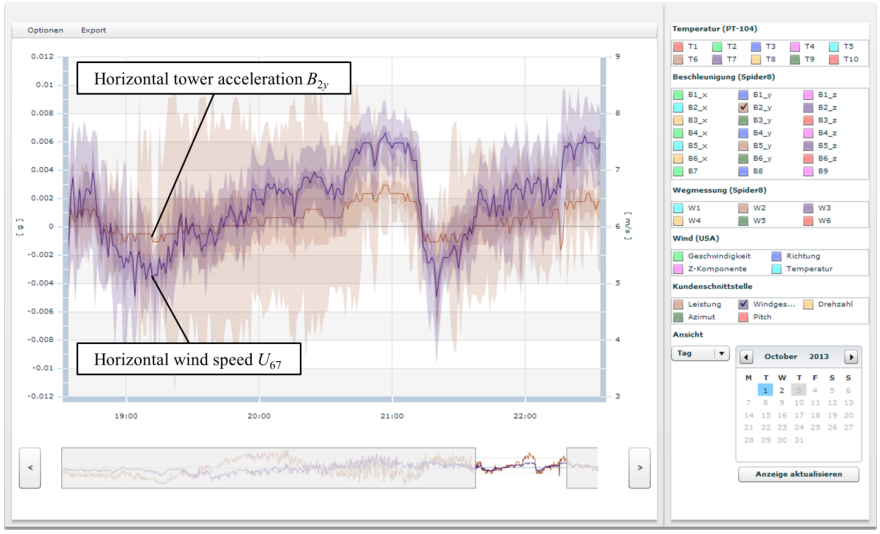


Figure 2. Web interface providing remote access to the LCM framework.

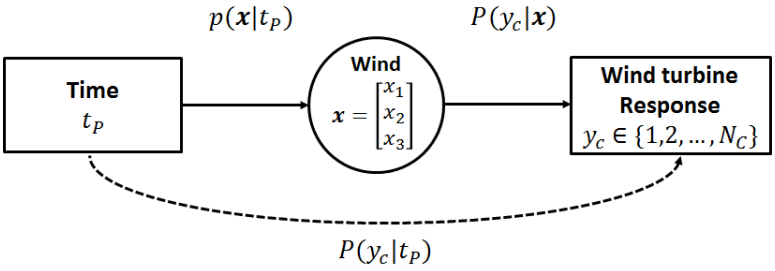


Figure 3. Probabilistic framework designed to understand the temporal variation of wind turbine response. The two probabilistic models $p(\boldsymbol{x}|t_P)$ and $P(y_C|\boldsymbol{x})$ are constructed by GMM and GDA, respectively, and are combined to construct $P(y_C|t_P)$.

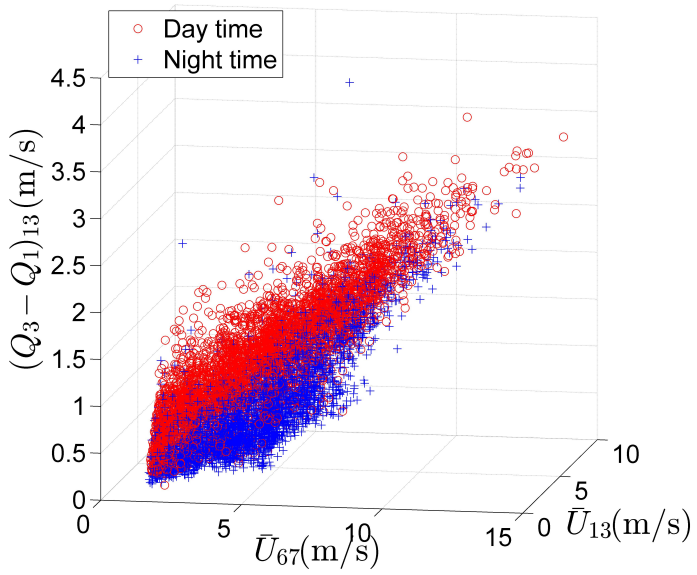
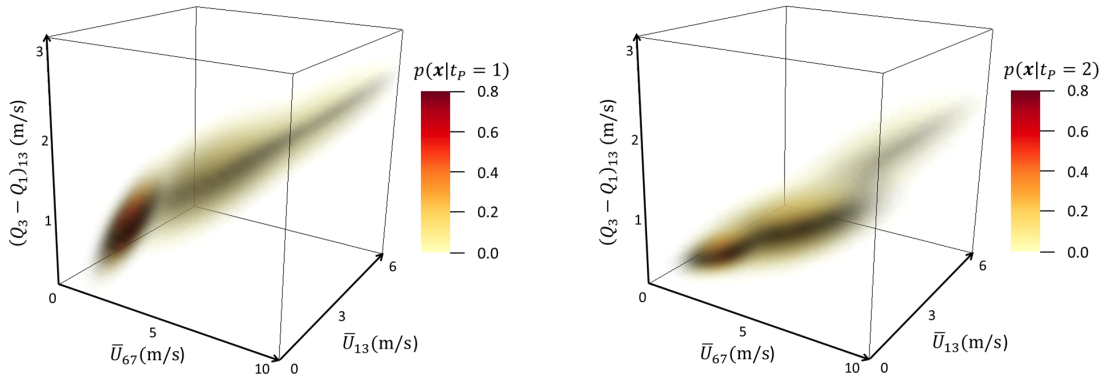


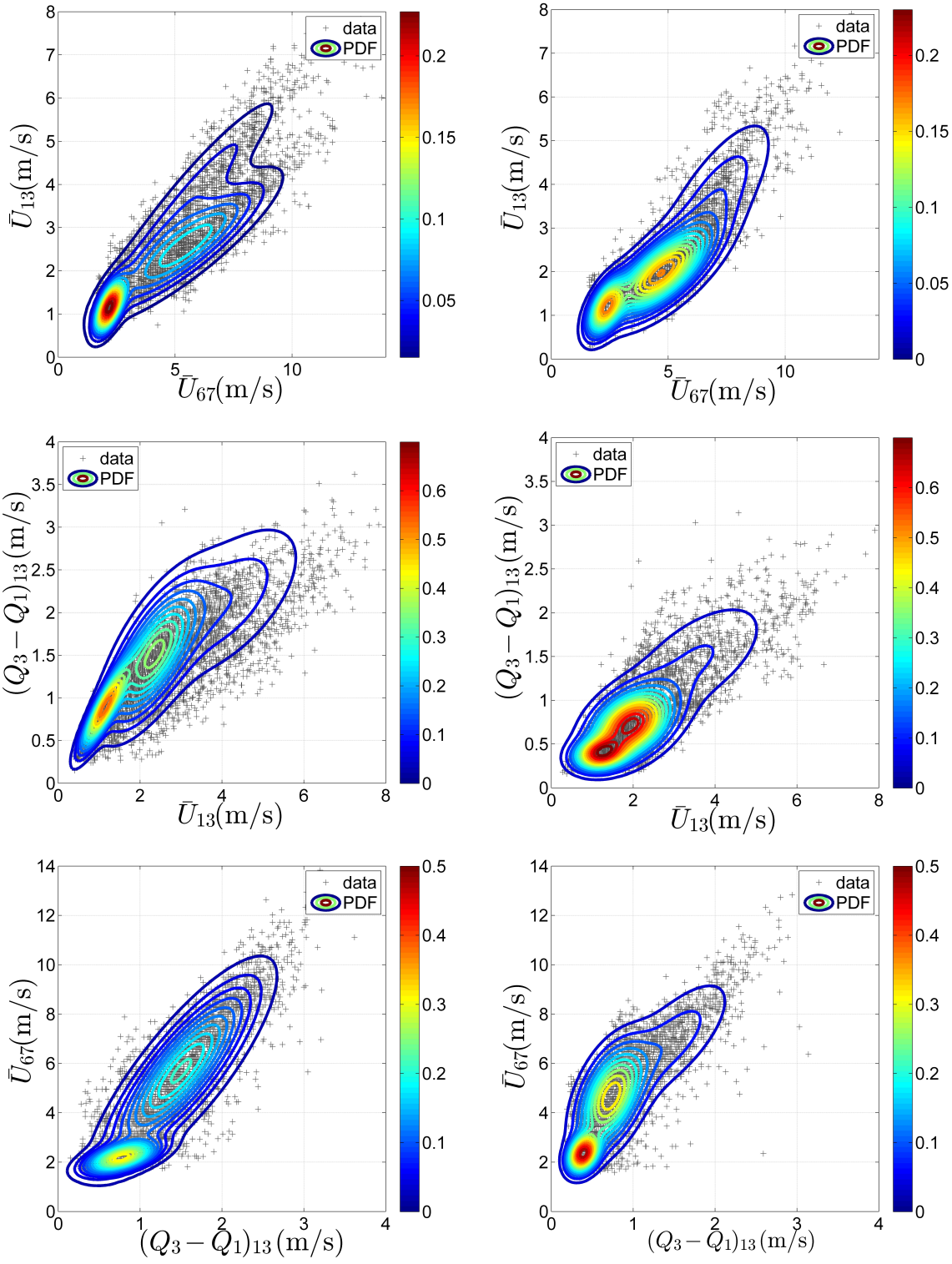
Figure 4. Wind data collected during day (circle) and night (cross) times. $\overline{U}_{67}$ and $\overline{U}_{13}$ denote the mean of a 20-minute wind speed time series data at 67 m height and at 13 m height, respectively. $(Q_3 - Q_1)_{13}$ denotes the interquartile of a 20-minute wind speed time series at 13 m height.



(a) $p(\boldsymbol{x}|t_p = 1)$ (daytime)

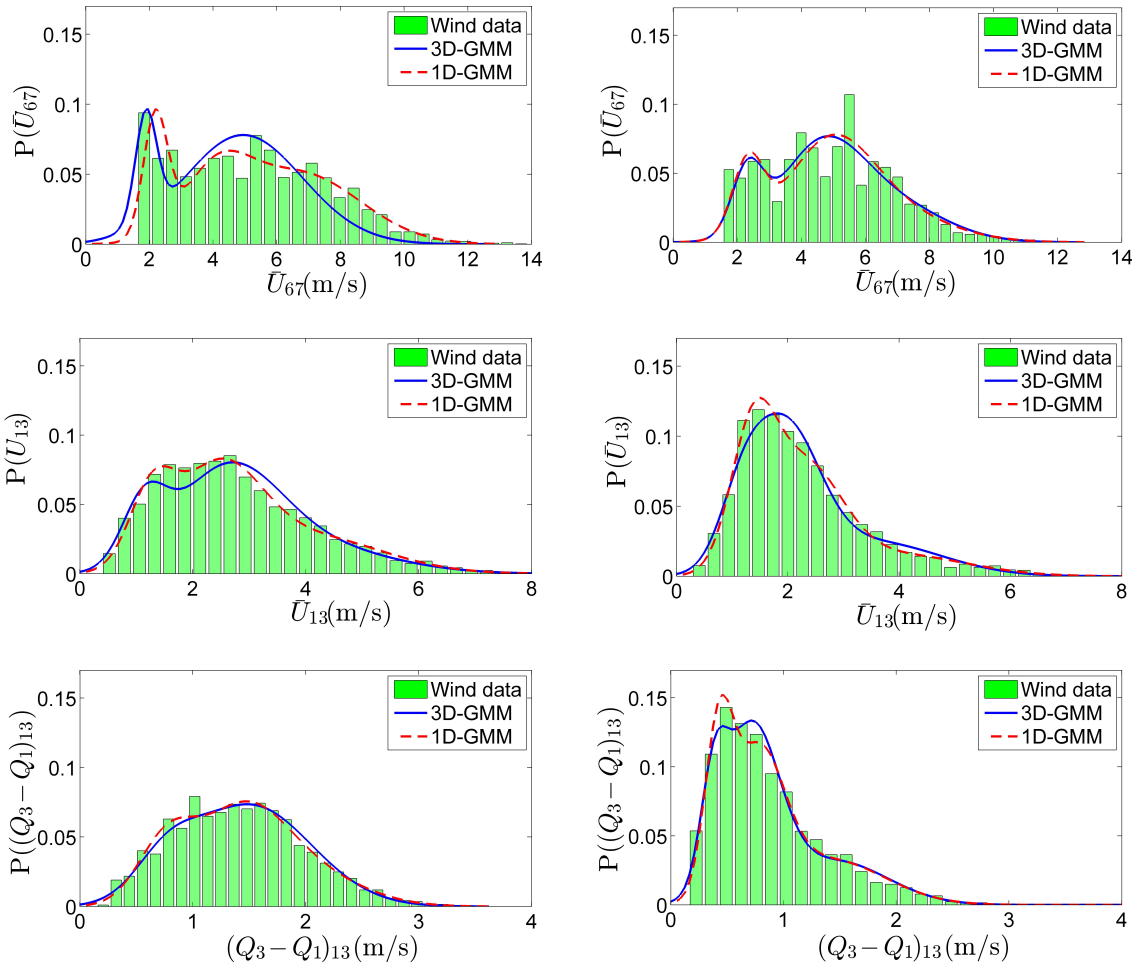
(b) $p(\boldsymbol{x}|t_p = 2)$ (night time)

Figure 5. Multivariate probability density functions for the wind field input features.



(a) Diurnal (b) Nocturnal

Figure 6. Marginal PDFs showing the characteristics for the diurnal and nocturnal wind fields. The contours represent PDFs constructed from GMM applied to the input data, represented as dot.



(a) Diurnal (b) Nocturnal

Figure 7. PMFs showing the characteristics for the diurnal and nocturnal wind fields.

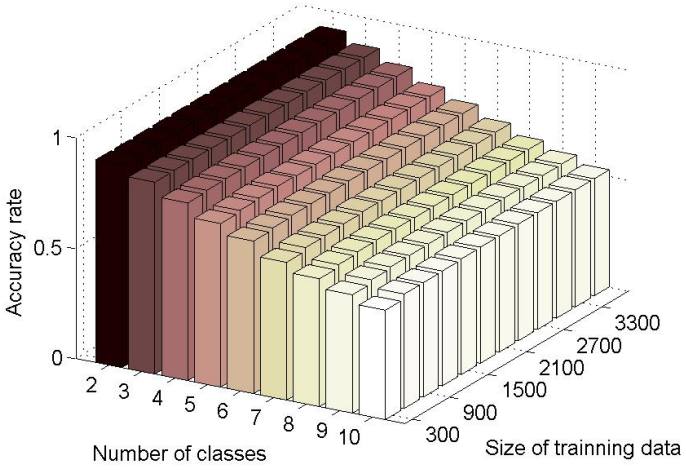


Figure 8. Comparison of the accuracy ratios for the response classification functions using different numbers of classes and different sizes of training data sets.

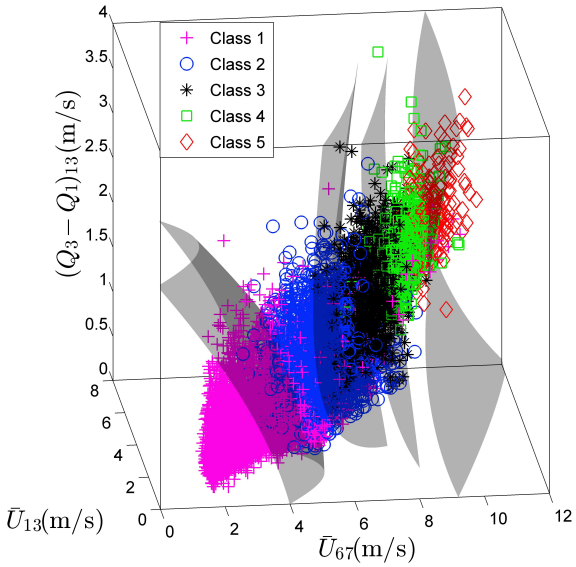
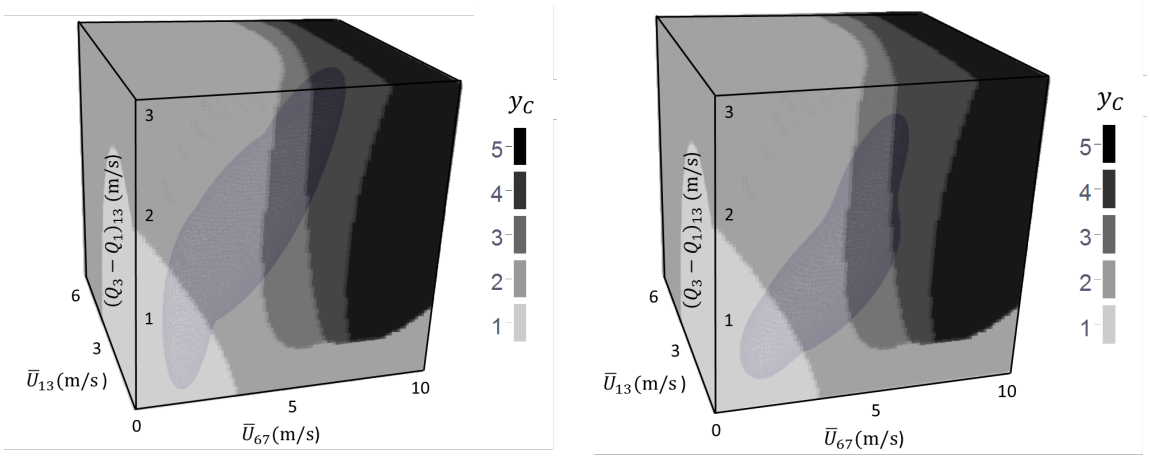


Figure 9. Training data for GDA and the constructed boundaries classifying the response classes.



(a) Day (b) Night

Figure 10. Time-specific PDFs, $p(\boldsymbol{x}|t_p = 1)$and $p(\boldsymbol{x}|t_p = 1)$, and the response classification function $g_C(\boldsymbol{x})$ for evaluating the class probabilities and the expected response classes.

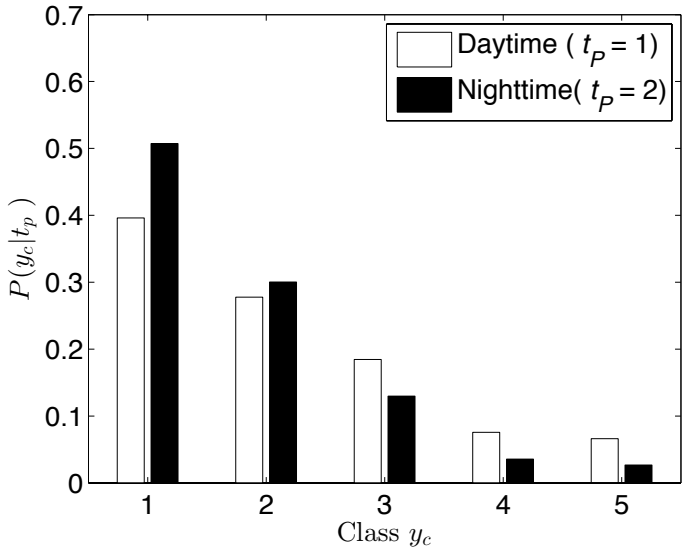


Figure 11. Comparison of the estimated class probabilities for the day and the night times.

Table 1. The ranges of interquartile value for each output response class.

Class y_C	1	2	3	4	5
Interquartile range	0.0 ~ 3.8 (mg)	3.8 ~ 7.6 (mg)	7.6 ~ 11.4 (mg)	11.4 ~ 15.2 (mg)	15.2 ~ 19.0 (mg)

Table 2. Comparison of the class probabilities and the expected loads.

		$y_C = 1$ (class 1)	$y_C = 2$ (class 2)	$y_C = 3$ (class 3)	$y_C = 4$ (class 4)	$y_C = 5$ (class 5)	Expected class
$t_P = 1$ (day)	Measured	0.3835	0.2901	0.1849	0.0799	0.0616	2.1461
	Estimated	0.3961	0.2776	0.1845	0.0756	0.0662	2.1281
$t_P = 2$ (night)	Measured	0.5202	0.2807	0.1468	0.0279	0.0244	1.7557
	Estimated	0.5074	0.3004	0.1299	0.0356	0.0266	1.7737