Multivariate analysis and prediction of wind turbine response to varying wind field characteristics based on machine learning

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

[1]Dept. of Civil and Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA 94305, USA; email: {jkpark11, smarsly, law}@stanford.edu
[2]Dept. of Civil and Environmental Engineering, Ruhr University Bochum, Universitätsstr. 150, 44780 Bochum, GERMANY; email: hartus@inf.bi.rub.de

ABSTRACT

Site-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 wind turbine structures to varying wind field characteristics. Machine learning algorithms are applied (i) to better understand changes of wind field characteristics due to atmospheric conditions, and (ii) to gain new insights into the wind turbine loads being affected by fluctuating wind. Using Gaussian Mixture Models, the variations in wind field characteristics are investigated by comparing the joint probability distribution functions of several wind field features, which are constructed from long-term monitoring data taken from a 500 kW wind turbine in Germany. Furthermore, based on Gaussian Discriminative Analysis, representative daytime and nocturnal wind turbine loads are predicted, compared, and analyzed.

INTRODUCTION

Variations in wind field characteristics can affect the structural and operational response of wind turbines. Wind field characteristics are often described by time-dependent statistical parameters such as mean wind speed, turbulence intensity, mean wind direction and vertical mean wind profile, which depends on the surface roughness (e.g. land or water) and on the atmospheric stability (e.g. day or night). For example, the daytime and the nocturnal wind field characteristics can differ significantly. Knowledge on the fluctuating wind field and its effects on wind turbines are essential not only for designing, but also for cost-efficiently managing wind turbines. However, these effects have rarely been studied.

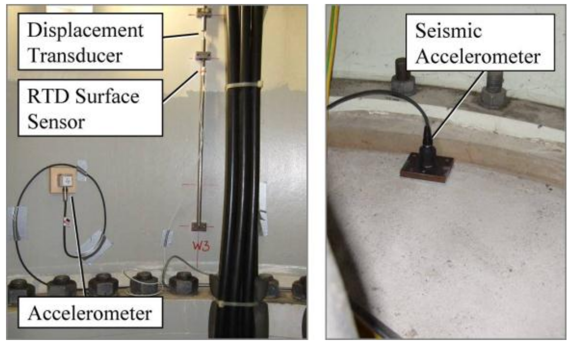
This paper presents a machine learning approach for analyzing and predicting the response of wind turbine structures to varying wind field characteristics. First, an integrated life-cycle management (LCM) framework, which provides long-term monitoring data recorded from a wind turbine, is briefly introduced. The proposed machine learning approach to relate wind field effects on wind turbines is discussed in details. Using the monitoring data provided by the LCM framework, a case study investigating the variation between the daytime and nocturnal wind field characteristics is presented.

WIND TURBINE LIFE-CYCLE MANAGEMENT FRAMEWORK

The wind turbine LCM framework, installed on the aforementioned 500 kW wind turbine, consists of (i) a structural health monitoring (SHM) system and (ii) several software modules installed at spatially distributed locations. The software modules include, for example, a monitoring database for persistent storage of the recorded data sets, a central server for automated data processing, a management module supporting life-cycle analyses, and Internet-enabled user interfaces providing online access to authorized users and to external application programs. Details on the LCM framework and on the integrated software modules have been described in Smarsly *et al.* (2011a,b, 2012a,b,c) and Hartmann *et al.* (2011). In the following subsections, the SHM system is introduced, followed by a description of the data sets used in this study.

Structural Health Monitoring System. The structural health monitoring system comprises of a network of sensors, data acquisition units and an on-site server located in the wind turbine. The wind turbine has a hub height of 65 m and a rotor diameter of 40.3 m. The sensors (accelerometers, displacement transducers, and temperature sensors) are placed at different levels on the inside and the outside of the steel tower and on the foundation of the wind turbine (Figure 1). In addition, two anemometers are deployed to continuously measure the wind speed, the wind direction, and the air temperature. The first anemometer, a cup anemometer, is installed on top of the wind turbine nacelle (h = 67 m). Additionally, the second anemometer, a three-dimensional ultrasonic anemometer, is mounted on a telescopic mast adjacent to the wind turbine (h = 13 m).

The data acquisition units, installed in the maintenance room of the wind turbine, continuously collect, sample, and digitize the structural and environmental data. Referred to as “primary monitoring data”, the data sets are continuously forwarded from the data acquisition units to the on-site server for temporary storage and periodic local backups. Through a permanently installed DSL connection, the on-site server transfers the primary monitoring data to the monitoring database. The database, which is an integral part of the LCM framework, is installed at the Institute for Computational Engineering (ICE) at the Ruhr University Bochum. In addition to the structural and environmental data, operational data is also transferred to the monitoring database. The operational data, such as power production of the wind turbine as well as revolutions and pitch angles of the rotor, is recorded by the wind turbine machine control system. Once being stored in the database, the monitoring data is remotely available online and accessible by authorized personnel and software modules.



(a) (b)

Figure 1. Illustration of sensor instrumentation: (a) Displacement transducer, RTD surface sensor, and accelerometer installed at the 42 m level inside the wind turbine tower, and (b) a seismic accelerometer at the ground level.

Data sets used in this study. The machine learning approach requires different inputs (i.e. wind field statistics) and corresponding outputs (i.e. structural response data) so that structural wind turbine behaviors are "learnt" from the existing input-output patterns. For input, long-term wind field statistics are computed from the measurements recorded by the anemometers described earlier. Specifically, the input feature vector $x = \{U_{67}, U_{67}–U_{13}, \sigma_{U67}\}$ consists of three statistical features, where U_{67} is the mean of a 20-minute wind speed time series at the 67 m height (recorded by the cup anemometer), U_{13} is the mean wind speed at the 13 m height (recorded by the ultrasonic anemometer), and σ_{U67} is the standard deviation of the 20-minute wind speed time series at the 67 m height. For the output, the vertical tower strain data, as calculated from the measurements recorded by the displacement transducer at the 42 m level (Figure 1a), are used. Quartiles of 20-minute strain time series are employed as an indirect measure for the level of fluctuation of the tower bending moment. For a classification of the quartiles, the output quartiles are discretized into 5 levels, i.e. $y = \{1, 2, 3, 4, 5\}$, on the basis of their magnitude. In total, 7,860 pairs of x (input feature vector) and y (discretized quartile levels), provided by the LCM system, serve as the basis for this study.

MACHINE LEARNING APPROACH

Wind Turbine Load Classification using Gaussian Discriminative Analysis. A wind turbine load classification function $\hat{y}(x)$ is defined by mapping the input feature vector x to the predicted wind turbine load class $\hat{y}$. The function $\hat{y}(x)$ is constructed using the Gaussian Discriminative Analysis (GDA), which is a generative learning algorithm that classifies the input features based on learned input feature distributions modeled by the Gaussian distribution. Given the trained input feature distribution $P(x|y)$ conditional on the load class y and the class prior probability $P(y)$, the posterior distribution on y given x is modeled according to the Bayes' rule as follows:

$$P(y \mid x) = \frac{P(x \mid y)P(y)}{P(x)} = \frac{P(x \mid y)P(y)}{\sum_y P(x \mid y)P(y)} \tag{1}$$

Then, for the new input feature vector x_{new}, representing the newly observed wind field statistics, the corresponding class $\hat{y}$ can be predicted according to the maximum a- posteriori detection (MAP) principle expressed as follows:

$$\hat{y} = \arg\max_y P(y \mid x_{new}) = \arg\max_y \frac{P(x_{new} \mid y)P(y)}{P(x_{new})} = \arg\max_y P(x_{new} \mid y)P(y) \tag{2}$$

In particular, GDA models $P(x|y = j)$ as a multivariate Gaussian distribution $N(\mu_j, \Sigma_j)$ and models $P(y)$ as a multinomial distribution, i.e. $P(y = j;\ \phi) = \phi_j$. To construct $P(x|y)$ and $P(y)$, we need to estimate the parameters μ_j, Σ_j, and ϕ_j for each load class. Using the training data sets $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$, the parameters μ, Σ, and ϕ are calculated as the ones maximizing the following log-likelihood function (Li *et al*., 2006):

$$l(\phi, \mu, \Sigma) = \log \prod_{i=1}^{m} P(x^{(i)}, y^{(i)}; \phi, \mu, \Sigma) = \log \prod_{i=1}^{m} P(x^{(i)} \mid y^{(i)}; \mu, \Sigma) P(y^{(i)}; \phi) \tag{3}$$

Joint Probability Density Estimation using Gaussian Mixture Models. The site-specific wind field characteristics are described by the wind fields statistics as specified in the input feature vector *x*, in particular the mean wind speed and the standard deviation. In this study, different wind field characteristics (by day and by night) are compared using the joint Probability Density Function (PDF) $f_{X|W}(x|w)$ for input feature *x* conditional on the atmospheric setting w. For a certain condition w, the joint PDF for *x* is constructed based on the Gaussian Mixture Model (GMM), in which PDF is expressed as the linear combination of *K* probability density function as follows (Figueiredo and Jain, 2002):

$$f(x) = \sum_{j=1}^{K} \phi_j f_j(x) = \sum_{j=1}^{K} \phi_j P(x \mid \mu_j, \Sigma_j) \tag{4}$$

where $P(x|\mu_j, \Sigma_j)$ is the *j*th Gaussian Probability Density Function (GPDF), which is given by a multivariate normal distribution $N(\mu_j, \Sigma_j)$, and ϕ_j is the weight of the *j*th GPDF.

To construct the Gaussian Mixture Density (GMD) function *f*(*x*), the parameters μ_j, Σ_j, and ϕ_j are preliminarily estimated for each *j*th distribution function. To this end, the log-likelihood of the data sets $\{x^{(i)}, \ldots, x^{(m)}\}$ is expressed as:

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log P(x^{(i)}; \phi, \mu, \Sigma) = \sum_{i=1}^{m} \log \sum_{z^{(i)}=1}^{k} P(x^{(i)} \mid z^{(i)}; \mu, \Sigma) P(z^{(i)}; \phi) \tag{5}$$

and the parameters μ, Σ, and ϕ can be found as the ones that maximize Eq. (5). The latent random variable $z^{(i)}$ specifies one of the k possible Gaussian distributions from which $x^{(i)}$ is drawn and $z^{(i)}$ follows a multinomial distribution. That is, when $z^{(i)} = j$ with the probability $P(z^{(i)} = j;\, \phi) = \phi_j$, the feature vector x follows the jth Gaussian distribution as $P(x^{(i)}|z^{(i)}; \mu, \Sigma) \sim N(\mu_j, \Sigma_j)$. Because $z^{(i)}$, being a random variable, is not known, the estimation of the parameters based on the maximum likelihood estimation is non-trivial. The Expectation Maximization (EM) algorithm is one efficient method for estimating the parameters, given the existence of the latent random variables z.

The EM algorithm consists of two iterative steps: (1) the "E-step" evaluates the probability of $z^{(i)}$, given the current data $x^{(i)}$ and the parameters estimated in the preceding iteration; (2) the "M-step" updates the parameters ϕ , μ, Σ that maximize the log-likelihood function according to (Eq. 5) using the $z^{(i)}$ estimated in the preceding E-step (Render and Walker 1984). These two steps are repeated until the parameters converge to some acceptable values.

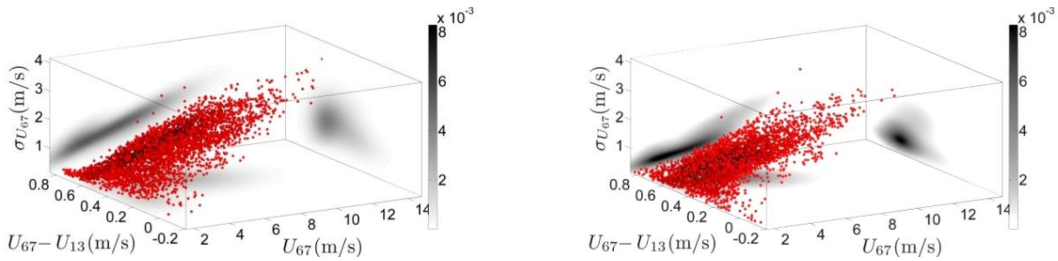
Expected Wind Turbine Load. The expected wind turbine load $E[\hat{y}|w]$, on condition that the atmospheric setting *w* can be assumed, can be expressed as:

$$E[\hat{y} \mid w] = \int_x \hat{y}(x) f_{X|W}(x \mid w) dx \approx \sum_{i=1} \hat{y}(x^{(i)}) P_{X|W}(x^{(i)} \mid w) \tag{6}$$

Here, $\hat{y}(x)$ is the wind turbine load classification function which maps the wind input feature *x* onto the discretized load class $\hat{y}$, and $f_{X|W}(x|w)$ is the joint PDF for wind field input features *x* given a certain atmospheric setting *w*. Note that the wind turbine loads depend only on the wind field input features *x*, and that the PDF for *x* is subject to change depending on the atmospheric condition *w*. The expectation value $E[\hat{y}|w]$ thus provides the insights into how a wind turbine experiences different levels of the true load class *y* given an easily observable atmospheric setting *w*.

CASE STUDY: APPLICATION OF THE MACHINE LEARNING APPROACH USING MONITORING DATA

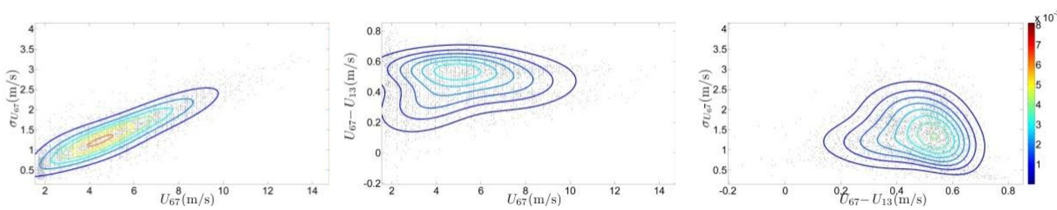
Construction of Joint PDFs for Wind Field Characteristics. For comparing the wind field characteristics during the daytime and the night time, 7,860 data sets are categorized into daytime and nocturnal data sets. Using four Gaussian mixtures, the joint PDF on relevant input features *x* is constructed for each data set using the EM algorithm. Figure 2 shows the joint PDFs of daytime and nocturnal wind field input features. The axes in the Cartesian plot denote the mean of a 20-minute wind speed time series at 67 m (U_{67}), the differences between the mean wind speed at 67 m and 13 m ($U_{67} - U_{13}$) and the standard deviation of 20-minute wind speed time series at 67 m (σ_{U67}). Therefore, the location of each dot in Figure 2 corresponds to a 20-minute wind field whose features are specified by the values in graph axes. Furthermore, the probability of each input feature vector *x* is expressed by dot's color. The two PDFs as plotted in Figures 2(a) and 2(b) clearly show the different variation in the characteristics between the day and night wind field. Wind fields in the daytime have larger dispersions in each input feature space compared to nocturnal wind fields. This is because the non-steady state boundary layer at daytime causes a more active air flow mixing.



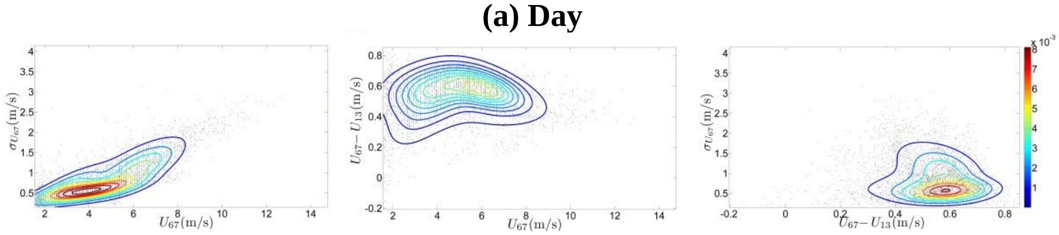
(a) Day **(b) Night**

Figure 2. Joint probability density functions for the wind field input features: (a) joint PDF for the daytime, and (b) joint PDF for the night time.

Based on the modeled PDFs, the probability for any arbitrary wind input feature *x* can be calculated. The probabilites of wind input features is shown in terms of three marginal PDFs as shown in Figure 3. Figures 3(a) and 3(b) show the correlation among the input features for the day time and for the night time, respectively. It can be seen that the mean and the standard deviation of the wind speed at 67 m are strongly correlated. In addition, the level of turbulence, which is depicted by the standard deviation, is lower at night than during the day, possibly because of the more stable nocturnal atmosphere.



(a) Day



(b) Night

Figure 3. Marginal joint probability density functions:
(U_{67}, σ_{U67}), (U_{67}, $U_{67} - U_{13}$), and ($U_{67} - U_{13}$, σ_{U67}).

Classification of the Wind Turbine Response. Using the GDA approach, the wind turbine load class $\hat{y}$ is predicted. For this purpose, the quartiles of the 20-minute tower strain time series – corresponding to 3,980 sets of input features vectors – are used for training for the GDA algorithm, whereas another 3,980 sets of input features are used for testing the proposed machine learning approach. The histograms for the measured and predicted classes are shown in Figure 4. In addition, the classification errors, defined as $y - \hat{y}$, are plotted for each input-output pair. As a result, the percentage of the exact classification ($y - \hat{y} = 0$) is about 78%. It should be emphasized that, if the error criterion is relaxed ($|y - \hat{y}| \leq 1$), the error percentages is reduced to 3.6%. Although the exact load classification is not directly needed, it is required for determining the distribution of the measured and predicted load classes, which vary due to the condition of the adjacent atmosphere and of the service of the wind turbine.

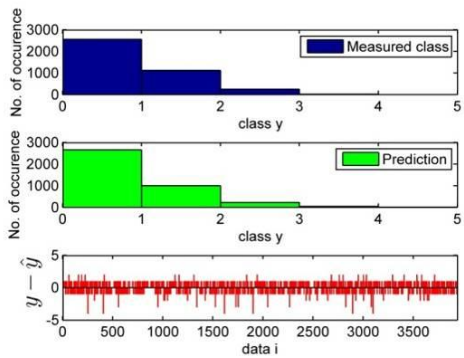


Figure 4. Wind turbine response class prediction (using the quartile range of 20-minute strain time series as wind turbine load output).

Wind Turbine Load Prediction. The expected (time-variant) load class $E[\hat{y}|w]$ is calculated based on the input feature joint PDF and the class-mapping function. The expected classes representing the daytime and nocturnal conditions are compared in Table 1. The effectiveness of the class-mapping model is evaluated by comparing the expected class based on the measured $y^{(i)}$ and the predicted class $\hat{y}(x^{(i)})$. The comparison results which show good agreement. In addition, the trend in load statistic variations can be studied by comparing the expected classees for the daytime and the nocturnal conditions. The expected load class is higher during the day than during the night, and this trend is well captured by the statistical models employed in this study.

Table 1. Comparison of the expected classes.

	Measured class $\sum_{i=1}^{N} y^{(i)} P_{X\|W}(x^{(i)}\|w)$	Predicted class $\sum_{i=1}^{N} \hat{y}(x^{(i)}) P_{X\|W}(x^{(i)}\|w)$
w = day	1.9608	2.1505 (9.7% overestimate)
w = night	1.3794	1.3021 (5.6% underestimate)

SUMMARY

A machine learning approach towards analyzing and predicting the response data of a wind turbine structure subjected to transient wind fields is investigated. Long-term monitoring data, provided by an integrated life-cycle management framework, is used to evaluate the proposed machine learning approach. The results obtained indicate that Gaussian Discriminative Analysis (GDM) and Gaussian

Mixture Models (GMMs) can be coupled to estimate wind turbine loads in various atmospheric conditions. In the study presented in this paper, it could be observed that – according to the monitoring data used – greater wind speeds and larger standard deviations are generally observed during the day.

ACKNOWLEDGEMENTS

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

REFERENCES

Figueiredo, A. M. and Jain, K. A. (2002). "Unsupervised learning of finite mixture models." *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 24(3), pp. 381-396.

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.

Li, T., Zhu, S. and Ogihara, M. (2006). "Using discriminant analysis for multi-class classification: an experimental investigation." *Knowledge and Information Systems*, 10(4), pp. 453-472.

Render, A. R. and Walker, F. H. (1984). "Mixture Densities, Maximum Likelihood and the EM Algorithm." *SIAM Review*, 26(2), pp. 195-239.

Smarsly, K., Law, K. H. and Hartmann, D. (2011a). "Implementation of a multiagent-based paradigm for decentralized real-time structural health monitoring." In: *Proceedings of the 2011 ASCE Structures Congress*. Las Vegas, NV, USA, April 14, 2011.

Smarsly, K., Law, K. H. and Hartmann, D. (2011b). "Implementing a Multiagent-Based Self-Managing Structural Health Monitoring System on a Wind Turbine." In: *Proceedings of the 2011 NSF Engineering Research and Innovation Conference*. Atlanta, GA, USA, January 4, 2011.

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.