

# A Performance Monitoring Approach for the Novel Lillgrund Offshore Wind Farm

Evangelos Papatheou, Nikolaos Dervilis, Andrew Eoghan Maguire, Ifigeneia Antoniadou, and Keith Worden

**Abstract**—The use of offshore wind farms has been growing in recent years. Europe is presenting a geometrically growing interest in exploring and investing in such offshore power plants as the continent's water sites offer impressive wind conditions. Moreover, as human activities tend to complicate the construction of land wind farms, offshore locations, which can be found more easily near densely populated areas, can be seen as an attractive choice. However, the cost of an offshore wind farm is relatively high, and therefore, their reliability is crucial if they ever need to be fully integrated into the energy arena. This paper presents an analysis of supervisory control and data acquisition (SCADA) extracts from the Lillgrund offshore wind farm for the purposes of monitoring. An advanced and robust machine-learning approach is applied, in order to produce individual and population-based power curves and then predict measurements of the power produced from each wind turbine (WT) from the measurements of the other WTs in the farm. Control charts with robust thresholds calculated from extreme value statistics are successfully applied for the monitoring of the turbines.

**Index Terms**—Machine learning, offshore wind farm, pattern recognition, supervisory control and data acquisition (SCADA), wind turbine (WT) monitoring.

## I. INTRODUCTION

THE idea of using supervisory control and data acquisition (SCADA) measurements for structural health monitoring (SHM) and condition monitoring has received attention from both the wind energy and structural engineering communities, particularly for the monitoring of critical infrastructures [1]. In order to maintain a qualitative profit with large offshore wind farms, a major challenge is to keep operational and maintenance costs to the lowest level by ensuring reliable and robust monitoring systems. For this reason, data mining and machine learning are promising approaches for modeling wind energy aspects

such as power prediction or wind load forecasting. It is perhaps well known that there has been a recent expansion in the use of wind energy, which is likely to continue with an accelerated pace over the coming years. Among the various forms of wind energy, offshore wind farms have become more popular, mainly due to the broader choice regarding their location and also the generally steadier and higher wind speeds that can be found over open water, when compared to land. It is also understood that, although offshore locations may be preferable to land sites, they can radically increase their maintenance costs; thus, the monitoring of offshore wind farms becomes crucial if the expansion of their use continues to grow.

There have been several different approaches proposed for the monitoring of wind turbines (WTs), from traditional non-destructive evaluation (NDE) [2] methods, or vibration approaches on the blades [3], [4], to advanced signal processing in gearboxes [5]–[8]. General reviews can be found in [9] and [10]. Several researchers have tried in recent years to apply damage detection technologies, and these studies were mainly in a laboratory environment [11]–[18]. Briefly, both passive and active sensing technologies have been applied in the context of WTs [13]. In passive sensing techniques, there is no external/artificial excitation as in active sensing techniques, which can have an effect on the sensitivity, the robustness, and the practical application of the approach. Most of the SHM techniques and sensor systems that are discussed in the literature and available to industry have been considered for application to WT blades. For comprehensive reviews and explanation on SHM, the reader is referred to [19] and [20]. In general, NDE approaches work in accessible parts of the structures, require a high degree of expertise, and can have substantial inspection costs, but they can be highly sensitive. SHM incorporates the effort to build a general online monitoring approach for structures, in order to reduce or even replace lengthy inspection costs. Among the methods, which have been applied to WT SHM and fall into the NDE category, are ultrasonic waves (popular with composite structures and mainly use piezoelectric transducers), smart paint (piezoelectric or fluorescent particles), acoustic emissions (usually barrel sensors), impedance tomography (carbon nanotube), thermography (infrared cameras), laser ultrasound (laser devices), nanosensors (electronic nanoparticles), and buckling health monitoring (piezoelectric transducer). Not all NDE methods can be used for online monitoring, i.e., some may require a halt of the operations of the systems for their inspection. Vibration-based monitoring methods generally use accelerometers, piezo- or microelectromechanical

Manuscript received October 3, 2014; revised January 22, 2015 and April 2, 2015; accepted April 30, 2015. Date of publication June 5, 2015; date of current version September 9, 2015. This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/J016942/1 and Grant EP/K003836/2.

E. Papatheou, N. Dervilis, I. Antoniadou, and K. Worden are with the Dynamics Research Group, Department of Mechanical Engineering, The University of Sheffield, Sheffield, S1 3JD, U.K. (e-mail: e.papatheou@sheffield.ac.uk).

A. E. Maguire is with Vattenfall Research and Development, New Renewables, Edinburgh, EH8 8AE, U.K.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIE.2015.2442212

systems (MEMS), laser vibrometry, and also strain- or fiber-optic sensors. These methods tend to be less sensitive but offer global online monitoring capabilities as well as the possibility of monitoring in nonaccessible areas of the structures. They also tend to be cheaper. The aforementioned methods can also be roughly separated into physics or data based, depending on whether they are defined from physical principles or just data driven. Examples of data-based approaches for monitoring WT are given in [21]–[23].

Most modern wind farms will contain some form of SCADA system installed, which can provide for the measurement and the recording of several different variables, such as wind speed, bearing and oil temperatures, voltage, and the power produced, among others. Since the SCADA system records constantly and is primarily used to monitor and control plants, it forms an ideal basis for a complete online SHM approach. In addition, SCADA extracts are perhaps the most direct and potentially useful data obtained from WTs, except of course any direct measurements acquired on the turbines themselves (through accelerometers, laser vibrometry, or any other sensor).

The use of SCADA data for monitoring has been shown in several studies, such as in [24]–[31], and in most cases, it aims at the development of a complete and automatic strategy for the monitoring of the whole turbine or wind farm, although sub-components (e.g., bearings, generator, etc.) may be individually assessed as well. Among the various approaches, power curve monitoring has been popular and successful. The WTs have been designed by manufacturers to have a direct relationship between wind speed and the power produced, and since they require a minimum speed to produce the nominal power, but limit the power generated from higher wind speeds, the power curve usually resembles a sigmoidal function. A critical analysis of the methods for modeling the power curve can be found in [32] and recent work in [33], but in general, researchers have exploited the deviation from a reference curve to perform SHM on turbines. The use of machine-learning approaches for the estimation of power generation can be seen as far back as in [34] and in [35], with more recent works appearing as well [36], [37]. In [38], a steady-state model of a whole wind farm with neural networks was shown to have fair results if the data used were preprocessed, while in [39], three operational curves, i.e., power, rotor, and pitch, were used for reference, in order to produce control charts for the monitoring of WTs. This paper explores the potential of using actual SCADA data for the monitoring of individual turbines, and of the whole farm, by constructing power curves for each turbine and then comparing how well they predict for other turbines. The modeling is done with neural networks and Gaussian processes (GPs) for comparison. Control charts are produced for the individual monitoring of the turbines using standard  $x$  chart plots and extreme value statistics (EVS) for comparison.

The layout of this paper is organized as follows: Section II describes the wind farm and the SCADA data, which were available. Section III presents the modeling of the power curves of the WTs, while Section IV displays the monitoring of the turbines with control charts. Finally, the paper is rounded off with some overall conclusions, a discussion of the potential of the approach, and the future work, which is currently planned.



Fig. 1. Location of the 48 turbines in Lillgrund wind farm [40].

## II. DESCRIPTION OF THE LILLGRUND WIND FARM AND THE NOVEL ELEMENT

The Lillgrund wind farm is situated in the sea area between Denmark and Sweden and consists of 48 identical WTs of 2.3-MW rated power [40]; their distribution is shown in Fig. 1. The wind farm is owned and operated by Vattenfall. The original labeling of the turbines made use of a combination of letters and numbers (rows A to D, see again Fig. 1), but for convenience here, the turbines were simply numbered from 1 to 48. In this paper, only the pure number labeling is going to be used. The separation between the turbines in the row is  $3.3 \times D$ , where  $D$  is the diameter of the turbine, and the rows are separated by  $4.3 \times D$  (see Fig. 1). The WTs are Siemens SWT-2.3-93 (see Fig. 2), which are characterized by a rotor diameter of 92.6 m and a hub height of 65 m, giving a rated power of 2.3 MW. The maximum rated power is reached when wind speeds take values of 12 m/s (rated wind speed).

It is important to note that the spacing between the turbines in the specific wind farm is significantly closer than most conventional farms [40], and this unique element is generally expected to affect their performance. This wind farm architecture was created deliberately for analyzing the effects and the interactions of each WT within such closer spacing. The available data used in this study correspond to a full year of operation. All the SCADA extracts consist of 10-min averages, with the maximum, mean, minimum, and standard deviation of the 10-min intervals being recorded and available. The actual sampling frequency is less than 10 min, but it is not disclosed here.

## III. POWER CURVE MONITORING OF WTs

The number of sources of information regarding reliability of WT technology lifetimes is very limited due to the highly competitive market. Offshore wind farms are arguably going to be the pioneers in future regarding the renewable energy sources; however, because they operate in remote areas away from land and are expanding into deeper waters, SHM will be an essential part of the success of these structures in the competitive market. Automatic mechanisms for anomalous performance



Fig. 2. Siemens SWT-2.3-93 in Lillgrund [40].

detection of the WTs are still in the embryonic stage. Following these thoughts, a view of sensitive and robust damage detection methodologies of signal processing is investigated in this paper.

Once more, one has to recall from before that the rare spacing (worldwide) between the turbines is one of the most challenging and interesting motivations behind the analysis of this work. The analysis is based on neural network and GP regression and is used to predict the measurement of each WT from the measurements of other WTs in the farm. Neural networks are standard machine-learning tools, and they were the first to be applied, but as will be shown, GPs have certain useful advantages. Regression model error is used as an index of abnormal response. Furthermore, as will be seen later using this regression error (residual error, which is the difference between the algorithm predictions and the actual SCADA data), a strong visualization that indicates when faults occur will be presented.

A key novel technique in computing these warning levels (thresholds), which indicate novelty (faults), is the usage of sophisticated tools in order to obtain these thresholds. As will be explained later, the warning levels are calculated by using EVS via evolutionary optimization algorithms such as the self-adaptive differential evolution (DE).

### A. ANNs

Artificial neural networks (ANNs) are algorithms or, to be more specific, mathematical models, which are loosely based on the way that the human brain and biological neurons work.

They have been extensively used for regression and classification, and they have been very successful in modeling data originating from various different sources. In the current work, the multilayer perceptron (MLP), i.e., the most common neural network, is used [41]. For more details on the MLPs, the reader is referred to [41] and [42]. Since neural networks have been successfully used for nonlinear regression, they seem ideal for learning the power curve of WTs. The wind speed is available, in 10-min averages, from the SCADA extracts for each turbine (there is an anemometer in each tower). In addition, the SCADA data provide a status for the operation of the turbines, usually in the form of an “error code.” For the creation of the healthy power curve (the reference curve), data from the whole year were used, but only when they corresponded to time instances with a status code equal to “0,” which means “no error” in the turbines. The one-year healthy data were separated into training, validation, and testing sets. The training set is primarily used for the training of the networks, whereas the validation is used to identify the best structure for the network. Different numbers of training cycles are applied, and in the end, the finally selected network is tested with fresh data with the help of the testing set. The search for the network structure here went from one up to ten hidden units, and the finally selected number of training cycles used was 300. All the training was done with the help of the Netlab [43] package, and the optimization algorithm for the network output error minimization was the scaled conjugate gradient method [43], [44]. The measure of the goodness of the regression fit was provided by the normalized mean-square error (MSE) shown in

$$\text{MSE}(\hat{y}) = \frac{100}{N\sigma_y^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

where the caret denotes an estimated quantity,  $y_i$  is the actual observation,  $N$  is the total number of observations, and  $\sigma_y$  is the standard deviation.

In total, 48 different networks (same as the number of turbines) were finally selected to create a power reference curve for the turbines. After that, each network was provided with wind speed data from the rest of the turbines and was asked to predict the power produced from them. In Fig. 3, the normalized MSE errors of each trained network, when tested with wind speed data for the turbine for which they were trained, and subsequently the remaining turbines, is shown. Each axis of the confusion matrix shown in Fig. 3 corresponds to 1 up to 48 turbines, where on the  $y$ -axis is the number of the trained turbine and on the  $x$ -axis the number of the tested turbine. In general, an MSE error below 5 is considered a good fit and below 1 excellent.

From the results, it is clear that almost all the trained networks are very robust and the maximum MSE error is around 5, which mainly occurs in turbines 3 and 4, which are located in the outside row of the wind farm. It can also be seen that, on the diagonal of the confusion matrix (which corresponds to the testing set of the trained turbines when tested with data from themselves), the MSE error is very low, with the maximum appearing in turbine 39 (MSE = 1.4708) and the

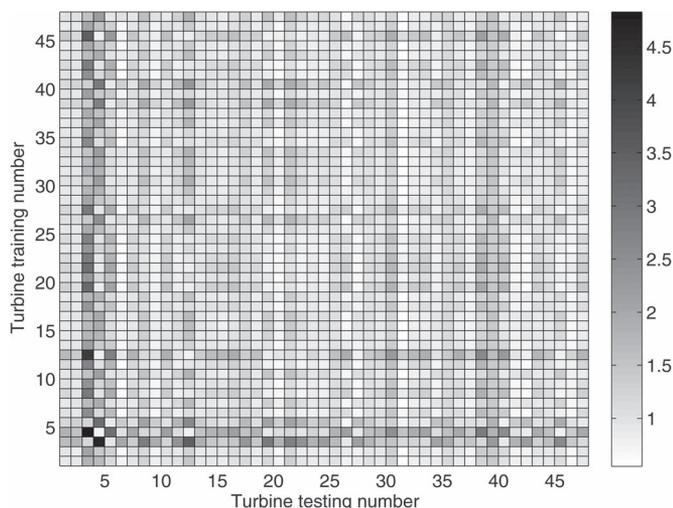


Fig. 3. Confusion matrix with MSE errors created from the neural networks: testing set.

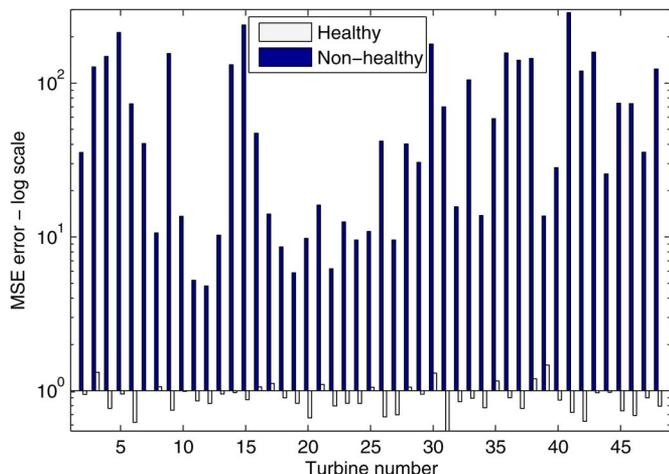


Fig. 4. MSE error of the neural network models when presented with data not corresponding to error code “0.”

minimum in turbine 31 with MSE at 0.5408. When the trained networks for an individual turbine were fed data originating from the same network, but which did not correspond to “no error statuses,” the MSE error was everywhere larger, as shown in Fig. 4; the lowest was 4.7991, which appeared in turbine 12, and it was still larger than the 0.8262 of the healthy data. In turbine 4, for example, the MSE increased from 0.768 to 149.033 and the standard deviation of the regression error from 0.0593 to 0.3685. Subsequent scanning of the data revealed that the majority of the instances where the regression error becomes high (in turbine 4) happened when the turbine was not working, either from emergency stops or manual stops. The emergency stops are associated with faults, but as these results derive from actual working data, the types of faults are limited to what was present during the recording period. Essentially, Fig. 3 shows a map of potential thresholds, which can be used for the monitoring (in a novelty detection scheme) of the turbines individually or as a population.

### B. GPs

Neural networks have proved to be a very powerful tool, particularly for nonlinear regression, but they also present several challenges during their modeling stage. The structure of the network, including hidden layers and hidden nodes along with the training cycles, plays a prominent role in the accurate modeling of data and the overall results of any such analysis. In addition, different initial conditions for the network weights must always be explored, and issues regarding overfitting of data are generally present, making the process nontrivial.

Alternatively, in the area of monitoring a WT via a regression analysis and in the exact same philosophy as the one described earlier, another powerful technique can be adopted, which is much simpler and faster. This technique is the GP for regression. The GP is a research area of increasing interest, not only for regression but also for classification purposes. The GP is a stochastic nonparametric Bayesian approach to regression and classification problems. These GPs are computationally very efficient, and the nonlinear learning is relatively easy. Regression with these algorithms takes into account all possible functions that fit to the training data vector and gives a predictive distribution of a single prediction for a given input vector. As a result, a mean prediction and confidence intervals on this prediction can be calculated from this predictive distribution. For more details, the reader is referred to [42] and [45].

The initial and basic steps in order to apply GP regression is to obtain mean and covariance functions. These functions are specified separately and consist of a specification of a functional form and a set of parameters called hyperparameters. Here, a zero-mean function and a squared-exponential covariance function are applied (see [45]). When the mean and covariance functions are defined, then the inference method specifies the calculation of the exact model and, in simple terms, describes how to compute hyperparameters by minimization of the negative log marginal likelihood. The software used for the implementation of GP regression was provided in [45].

In Fig. 5, a similar confusion matrix to that produced for the MLPs is shown. The results appear to be very good again, with the same level of robustness and similar levels of MSE error, with the worst being again in turbines 3 and 4. In terms of the comparison between neural networks and GPs, it appears that the results are very similar with the networks performing with a slightly lower MSE error. It should be noted that the GPs are trained with about a third of the data that the neural networks were provided, but the testing sets are everywhere the same.

Fig. 6 simply shows the average MSE errors contained in the confusion matrix shown in Fig. 3. It shows how well each trained (reference) power curve predicts the power produced in the rest of the turbines on average, and also how well the power produced in each turbine is predicted by the rest of the trained curves (corresponding to the rest of the turbines). In Fig. 6, it can be seen again that the worst turbines are 3 and 4, which predict and are also predicted by the rest of the turbines with a greater error than the rest. Similar plots as to those in Fig. 6 can be produced for the GPs, and the results are equivalent; hence,

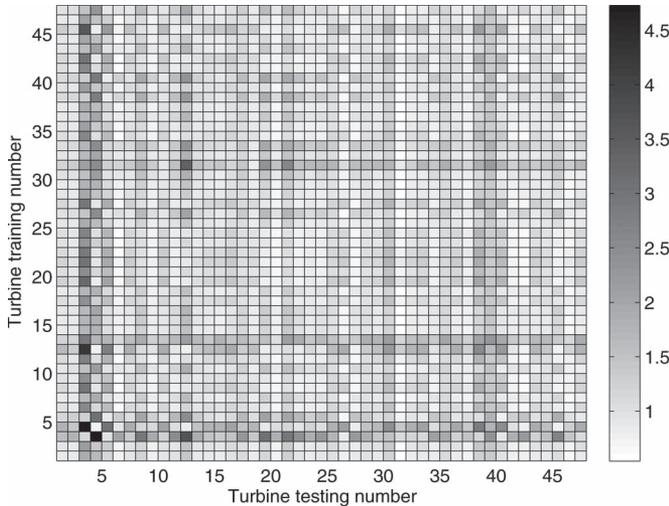


Fig. 5. Confusion matrix with MSE errors created from the GPs: testing set.

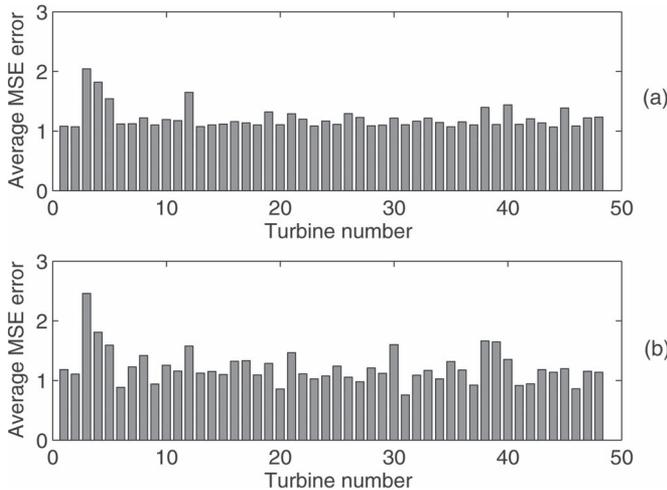


Fig. 6. Average MSE error showing (a) how well the power produced in each turbine is predicted by the networks trained in the rest of the turbines and (b) how well each trained network produces the power produced in the other turbines.

they are not included here. The very low MSE errors show that the power curves have the potential of being used as a feature for the monitoring of the whole farm, as they were shown to be generally robust to the individual differences that the turbines inevitably present (location, different sensors, different generators, etc).

#### IV. VISUALIZING THE DATA USING GP ANALYSIS THROUGH ROBUST EVS THRESHOLDS

As previously mentioned, GP regression considers all possible families of functions that fit to the training set observations and, as a result, provides a predictive distribution, which gives a mean prediction and confidence intervals on this prediction. The advantage of the approach is illustrated in Fig. 7, where the actual and the predicted power for turbine 9 are shown. The intervals plotted in Fig. 7 are plus and minus three standard

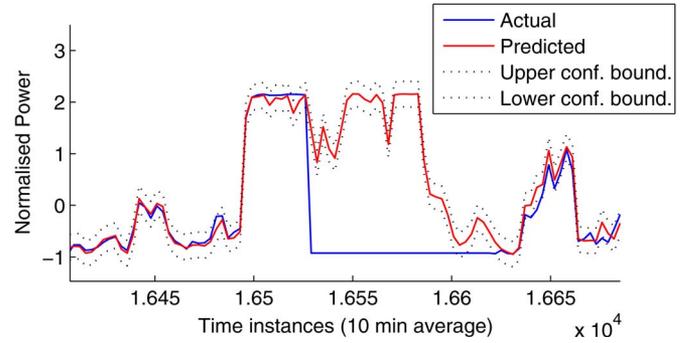


Fig. 7. Predicted and actual power when there was a fault in turbine 9 (smoke in the nacelle).

deviations of the predictive distribution. If one assumes that an accurate model was obtained, when an observation is outside these confidence intervals, then it may be considered as an outlier/novelty and, potentially, a fault. This approach is superior to the neural networks, where no confidence intervals are available, and also offers the advantage that, if highly abnormal conditions occur, which differ from those in the training set, the confidence intervals will increase, meaning that a performance anomaly might not be classified wrongly as an outlier.

As previously mentioned, all data were separated equally into training, validation, and testing sets. The training set was only used for training, and the testing set was used for the judgment of the models. The choice for the turbine for illustration was fairly arbitrary, but turbine 9 had a relatively low MSE of 0.76 in the testing set (see previous section). Fig. 7 shows the predicted and the actual power for turbine 9 in some time instances where a fault had been identified. The fault involved smoke in a part of the nacelle, and it is clearly shown that the actual values lie very far from the confidence boundaries of the GP. It is possible then by monitoring the regression error of the identified model to identify severe faults in the turbine, which affect its performance. This critical analysis is investigated in the next section.

#### A. Control Chart Monitoring With $x$ Chart Plots

The predictive distribution provided by the GP regression may present an easy way of monitoring the performance of a WT; however, an online approach will demand a more robust quantity to be monitored other than just the confidence intervals of the prediction. The simplest such quantity is the residual error between the actual and the predicted power produced. A control chart is a straight forward way of monitoring the residual error in the turbines. A standard  $x$  chart control approach was the first to be applied. The residual error used for all control charts presented here was provided by the models, which were trained on measurement data from the same WTs, which they were intended to monitor, and not from the rest of the turbines. The residual error thus corresponds to the values of the diagonal of the confusion matrix shown in Fig. 5. The control limits were calculated by  $\mu + 2.58\sigma$  [46] for a 99% level confidence. Fig. 8 shows the results of the  $x$  chart plot in the training error, and Fig. 9 shows the results for the validation set. It is reminded

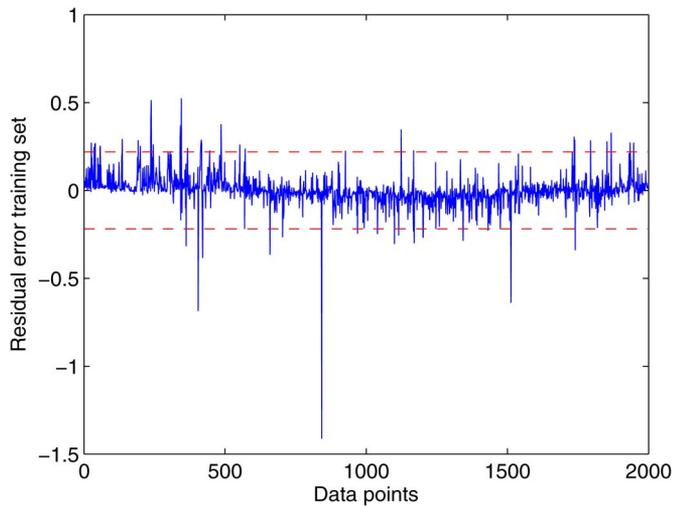


Fig. 8. Standard  $x$  chart plot for the training error in turbine 9.

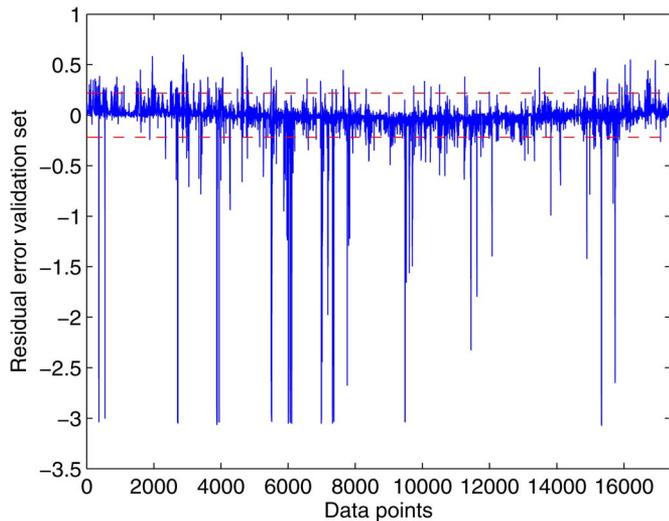


Fig. 9. Standard  $x$  chart plot for the validation set error in turbine 9.

that the validation set was used for the identification of the best neural network structure, but it was not used at all in the GP regression. In addition, the training set does not contain any data, which corresponds to time instances with faults present (error codes equal to “0” are considered healthy). It is shown in Fig. 8 that there are 55 instances, which are considered outliers, out of the 2000 total. If the controlled variable is assumed to follow a Gaussian distribution, then the proposed control limits should incorporate 99% of the data, which would mean approximately 20 outliers out of the 2000, i.e., a number smaller than the 55 shown in Fig. 8.

In order to check for the correct ratio of false positives (FPs) and false negatives (FNs) during the monitored period of time, in this case either the validation or the testing set, the use of the error codes provided is crucial. The “non-healthy” data, which appeared during the year, correspond to various types of error codes, each one representing different parts of the monitored system, including the whole wind farm and the grid. The error codes, therefore, were separated into categories of different

severity, according to the type and the recommended action upon their presence, information which was provided by the company owning the wind farm, i.e., Vattenfall. For example, a fault that indicated a warning with a recommended action of autoreset, was considered of lower severity than the one that indicated an alarm and a call for a technical team to the turbine. Finally, for the purposes of identification of the FP and FN in the specific work, the error codes were separated into those which were considered relevant to the power produced, which would therefore be detectable by monitoring the power curve, and those which were irrelevant (e.g., lamp failures, etc.) and consequently undetectable by the current methodology.

By using the aforementioned classification of the faults for the period corresponding to the validation set, Fig. 9 reveals that there are 486 FPs, meaning wrongly identified outliers, and 235 FNs, meaning instances where faults were present but were not detected. This means that out of the total 576 time instances with faults present, 341 were detected.

### B. Extreme Value Threshold Calculated With DE

It was shown that the standard  $x$  control charts demonstrated a detection rate of approximately 59% for the whole year of the recorded data for turbine 9, but they also indicated a high number of FPs. To alleviate this, EVSs [47] were employed in order to produce more robust thresholds. The assumption of a Gaussian distribution, although very common and very attractive, may not be entirely appropriate in cases where the problem at hand is concerned with tails of a distribution. In those cases, EVS may be more suitable.

In order to study the tails of an arbitrary parent distribution, the maximum of the samples should be used for its right tail and the minimum for its left. According to Fisher and Tippett [48], when the number of vector samples originating from an arbitrary parent distribution tends to infinity, the induced distribution on the maxima of the samples can only take one of three forms: Gumbel, Weibull, or Frechet. The distribution used here is the Gumbel, for reasons that will be explained further on, and it can be seen in (2) for the maxima and (3) for the minima, i.e.,

$$H(x) = e^{-e^{-\frac{x-\lambda}{\delta}}} \quad (2)$$

$$L(x) = 1 - e^{-e^{-\frac{x-\lambda}{\delta}}} \quad (3)$$

where  $\lambda$  and  $\delta$  are model parameters to be estimated from the data. The key point in such a situation would be to fit models to parts of the parent distribution’s tails, as they should follow one of the three possible distributions and thus identify a parametric model. Once the parametric model is obtained, a threshold can be calculated based on the fitted distribution. The chosen approach for the modeling here was DE [49].

DE belongs to the family of evolution-based algorithms, where an initial random population of solutions is propagated through a repeated cycle of mutation and crossover operations until an optimal (or near optimal according to desired criteria) solution is obtained. The process is explained in detail in [50]

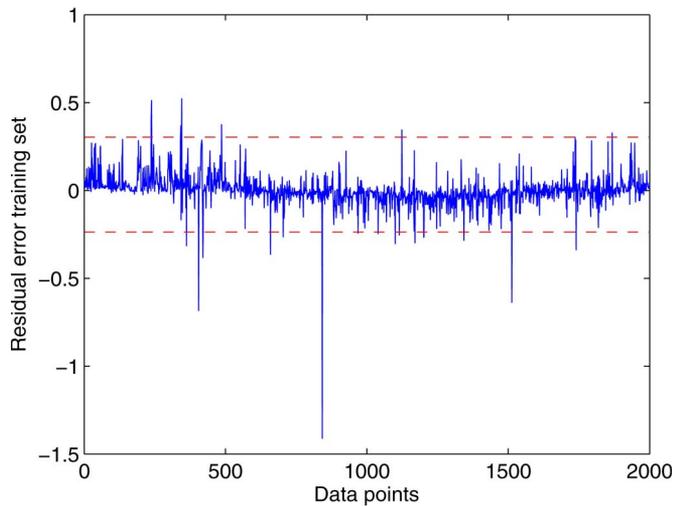


Fig. 10. Control chart for the training error in turbine 9 with EVS thresholds.

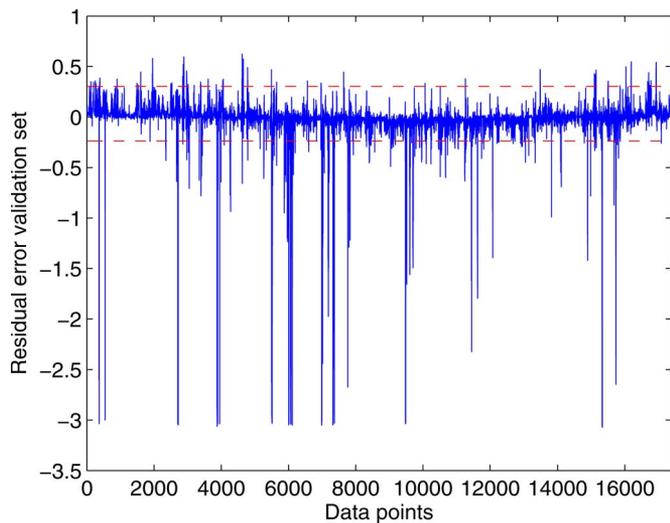


Fig. 11. Control chart for the validation set error in turbine 9 with EVS thresholds.

and [51]. Inherent in an evolution process is the calculation of a fitness function, which in the particular problem here is to fit a parametric model to a given cumulative distribution function (CDF). A normalized MSE similar to (1) was also applied here.

In the case of turbine 9, all three distributions were tried with the Gumbel providing the best results. After the same distribution was applied to a few more turbines with similar results, it was decided to be used for all of them. The parameters of the DE were a crossover ratio of 0.5, a scaling factor in the mutation vectors of 0.9, and a population size of 30. A fixed number of maximum generations was set at 100, after it was seen that, in most cases, the rate of change of MSE was stabilizing long before actually reaching this number, and its increase did not improve the results. Figs. 10 and 11 show the training and validation sets when the EVS was used in order to calculate the thresholds. The level of confidence used was 99%, meaning that 1% of outliers was expected to be indicated.

A quick comparison with Fig. 8 reveals that the number of outliers has been reduced to 24 from 55, i.e., a number which is very close to the theoretical 1% (out of 2000 training data points). Subsequent analysis of the validation set reveals that the number of FPs has now dropped to 252 (from 486), i.e., a significant improvement, although the number of FNs has slightly increased to 250 from 235.

The same approach was followed for all 48 turbines, thresholds were calculated with the standard  $x$  chart, and EVS approaches and FPs and FNs were calculated for the training and the validation sets. Table I shows those results for ten WTs, along with the mean values from all of them. It is reminded here that this table contains results from the residual error supplied from the models, which were trained on the measurement data from the same turbines they were intended to monitor and not from any nearby turbines. It is clear that, in almost all turbines, the EVS thresholds significantly reduce the FPs (approximately 53% in the training set and 52% in the validation, on average) and slightly increase the FNs (approximately 3%, on average). In the validation set, there is an average of 1.17% of FPs with the extreme value thresholds, which is very close to the theoretical 1% of outliers for the 99% confidence level, while the  $x$  chart approach has 2.44% of FPs. It is reminded that, in order to identify FNs, the error codes, which were not considered relevant to the power curves, were removed. The results shown in Table I consider the same error codes as in the results presented previously for turbine 9. The procedure of removing error codes is subjective because there were in total 249 different error codes present during the whole year of the data and it is not always trivial to understand which are relevant to the power curves.

The whole analysis was repeated without removing any error codes, and the detection rate was shown to be 23.4% (average from all turbines) for the  $x$  charts and 20.8% for the extreme value thresholds. When six error codes are removed, then the detection rate goes to 47.1% (average from all turbines) and 45% for  $x$  charts and EVS, respectively. Some turbines have a low detection rate, which lowers the overall average value. Those six error codes, which represent types of faults, are shown in Table II. Those results are a bit higher than those shown in Table I mainly because two specific error codes were not previously removed (4 and 6). With more understanding of the error codes and of the data, it is expected that a maximum detection rate can be found, which should indicate the actual sensitivity of the power curves. The faults that were not identified may correspond to very small changes in the system, which are essentially undetectable or correspond to faults for which the power curve feature is insensitive. Finally, the identification of faults relies heavily on the reliability of the potential faults, which were present and of the “healthy” data. It is possible that, while data appeared normal (meaning “0” error code), they were not in reality and ended up being used in the reference power curve. The same can be said about the faults, as some error codes may mask the presence of other faults.

The analysis presented is based on the real SCADA data from the whole year, which also contain significant influence from environmental effects, such as temperature, which may complicate the detection of faults. Although the regression

TABLE I

FPS AND FNS FOR TEN WIND TURBINES AND THE MEAN VALUES FROM ALL 48. DR STANDS FOR DETECTION RATE. THE MODELS USED HERE WERE TRAINED ON DATA FROM THE SAME WTS AS THOSE THEY WERE INTENDED TO MONITOR

Turbine No	Training set		Validation set							
	$x$ chart	EVS	$x$ chart				EVS			
	FP	FP	FP	FP %	FN	DR %	FP	FP %	FN	DR %
1	69	24	530	3.06	433	37.88	177	1.02	475	31.85
3	41	20	309	1.78	394	44.58	130	0.75	416	41.45
6	55	24	471	2.70	313	29.19	196	1.12	321	27.34
9	55	26	486	2.80	235	59.20	252	1.45	250	56.60
12	61	38	241	1.38	212	34.16	148	0.85	216	32.92
15	42	24	387	2.23	345	41.72	205	1.18	352	40.54
18	57	29	384	2.21	390	30.85	198	1.14	405	28.19
21	51	31	306	1.76	237	44.63	160	0.92	239	44.16
24	67	29	489	2.80	232	30.12	221	1.27	236	28.92
27	14	25	134	0.77	240	42.03	237	1.36	215	48.07
average	51.69	24.17	421.92	2.44	305.98	38.71	202.73	1.17	314.83	36.95

TABLE II

EXAMPLES OF ERROR CODES THAT WERE REMOVED FROM THE ANALYSIS (THE NUMBERS DO NOT CORRESPOND TO THE ACTUAL CODES USED IN THE SCADA)

Error code no	Error code
1	AviLight, Lamp 1 failure
2	AviLight, Lamp 2 failure
3	GrdInv: 143 Ethernet ch.1 loss
4	Secondary WS unit out of range
5	Inv.(tow) coolwater pres. war
6	Many Anemometer activations

error was shown to be small for the whole year, it is possible that modeling the power curves on a smaller time scale, e.g., monthly or weekly, will probably increase the detection of faults, but this may require the frequent update of the reference power curve. In addition, the SCADA data currently used are 10-min averages, which may also be a reason for lower detection of faults; it is expected that 1-min data will improve the results. The approach does not require a complex physics-based model of the WTs and does not preprocess the data or filter out potential outliers; it is generally straightforward to apply since it only requires wind measurements and power produced for the development of a reference curve.

The results shown in this paper were presented for  $x$  chart plots under an assumption of Gaussian statistics and using EVS distributions. It is shown that the EVSs provide the appropriate thresholds for a given confidence level. One could argue that better thresholds could be obtained by using an  $\bar{x}$  chart. This is certainly true; however, it requires grouping the data. For example, assuming a group size of 12 here for an  $\bar{x}$  chart would mean that diagnostic results would only be checked every 2 h instead of 10 min.

## V. CONCLUSION

This paper has presented an exploration of the suitability of SCADA extracts from the Lillgrund wind farm for the purposes of monitoring the farm via sophisticated machine-learning architectures. ANNs and GPs were used to build a reference power curve (wind speed versus power produced) for each of the 48 turbines existing in the farm and as well as EVS via an optimization algorithm in order to define alarm thresholds. Then, each reference model was used to predict the power produced in the rest of the turbines available, thus creating a confusion matrix of the MSE errors for all combinations. The results showed that nearly all models were very robust with the highest MSE error to be 4.8291, and this was happening when the model trained in turbine 4 was predicting power from turbine 3. Both turbines 3 and 4 are located in the outside row of the wind farm. It was shown that, when wind speed data, which did not come from time instances where the error status was "0" (meaning healthy data), were used as an input to the trained neural networks, the MSE error was significantly larger for neural networks and GPs. Comparison between neural networks and GPs showed that there is no significant difference in their performance, but the inherent ability of the GPs to produce confidence intervals is advantageous. The residuals of the models, which used data from the same turbines, which they were intended to monitor, were next used in a novelty detection scheme. It was shown that it is possible to monitor significant events that will affect the performance of the turbines by simple control charts, although certain faults remained undetected. An approach for the calculation of robust thresholds in the residuals made use of EVS and showed a significant reduction in the number of FPs. Future work will also focus on extra features other than the power curve for the improvement of the approach. In addition, the full analysis of the error statuses that were presented during the recorded time can lead to a more intelligent identification of faults, including their classification.

## REFERENCES

- [1] A. Grilo, J. Chen, M. Diaz, D. Garrido, and A. Casaca, "An integrated WSA and SCADA system for monitoring a critical infrastructure," *IEEE Trans. Inf. Informat.*, vol. 10, no. 3, pp. 1755–1764, Aug. 2014.
- [2] J. E. Doherty, "Nondestructive evaluation," in *Handbook on Experimental Mechanics*. Bethel, CT, USA: Soc. Exp. Mech., Inc., 1987, ch. 12.
- [3] N. Dervilis *et al.*, "On damage diagnosis for a wind turbine blade using pattern recognition," *J. Sound Vib.*, vol. 333, no. 6, pp. 1833–1850, Mar. 2014.
- [4] W. Yang, Z. Lang, and S. Tian, "Condition monitoring and damage location of wind turbine blades by frequency response transmissibility analysis," *IEEE Trans. Ind. Electron.*, to be published.
- [5] I. Antoniadou, "Accounting for nonstationarity in the condition monitoring of wind turbine gearboxes," Ph.D. dissertation, Faculty Eng., Univ. Sheffield, Sheffield, U.K., 2013.
- [6] W. J. Staszewski and K. Worden, "Classification of faults in gearboxes—Pre-processing algorithms and neural networks," *Neural Comput. Appl.*, vol. 5, no. 3, pp. 160–183, 1997.
- [7] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, no. 1/2, pp. 108–126, Feb. 2013.
- [8] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, "Monitoring wind turbine gearboxes," *Wind Energy*, vol. 16, no. 5, pp. 728–740, Jul. 2013.
- [9] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 1, pp. 1–39, Jan. 2009.
- [10] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renew. Energy*, vol. 46, pp. 169–178, Oct. 2012.
- [11] D. Adams, J. White, M. Rumsey, and C. Farrar, "Structural health monitoring of wind turbines: Method and application to a HAWT," *Wind Energy*, vol. 14, no. 4, pp. 603–623, May 2011.
- [12] E. R. Jørgensen *et al.*, "Full scale testing of wind turbine blade to failure—Flapwise loading," Risø Nat. Lab., Roskilde, Denmark, Tech. Rep., Jun. 2004.
- [13] G. R. Kiriker *et al.*, *Encyclopedia of Structural Health Monitoring*. Hoboken, NJ, USA: Wiley, 2009, ch. 147.
- [14] O. J. Kristensen *et al.*, *Fundamentals for Remote Structural Health Monitoring of Wind Turbine Blades—A Preproject, Annex E: Full-Scale Test of Wind Turbine Blade, Using Sensors and NDT*. Roskilde, Denmark: Risø Nat. Lab., May 2002.
- [15] L. Overgaard and E. Lund, "Structural collapse of a wind turbine blade. Part B: Progressive interlaminar failure models," *Composites A, Appl. Sci. Manuf.*, vol. 41, no. 2, pp. 271–283, Feb. 2010.
- [16] L. C. Overgaard, E. Lund, and O. T. Thomsen, "Structural collapse of a wind turbine blade. Part A: Static test and equivalent single layered models," *Composites A, Appl. Sci. Manuf.*, vol. 41, no. 2, pp. 257–270, Feb. 2010.
- [17] B. F. Sørensen *et al.*, "Improved design of large wind turbine blade of fibre composites based on studies of scale effects," Risø National Lab. Bæredygtig Energi, Danmarks Tekniske Universitet, Roskilde, Denmark, Tech. Rep., Sep. 2004.
- [18] J. R. White, D. E. Adams, and M. A. Rumsey, "Modal analysis of CX-100 rotor blade and Micon 65/13 wind turbine," *Struct. Dyn. Renew. Energy*, vol. 1, pp. 15–27, Jun. 2011.
- [19] H. Sohn *et al.*, "A Review of Structural Health Monitoring Literature: 1996–2001," Los Alamos Nat. Lab., New Mexico, USA, Tech. Rep. LA-13976-MS, 2004.
- [20] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. Hoboken, NJ, USA: Wiley, 2013.
- [21] S. Yin, G. Wang, and H. R. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, vol. 24, no. 4, pp. 298–306, Jun. 2014.
- [22] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.
- [23] S. Yin, S. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.
- [24] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, Sep. 2009.
- [25] A. Kusiak and Z. Zhang, "Adaptive control of a wind turbine with data mining and swarm intelligence," *IEEE Trans. Sustain. Energy*, vol. 2, no. 1, pp. 28–36, Jan. 2011.
- [26] A. Kusiak and W. Li, "The prediction and diagnosis of wind turbine faults," *Renew. Energy*, vol. 36, no. 1, pp. 16–23, Jan. 2011.
- [27] A. Kusiak, Z. Zhang, and A. Verma, "Prediction, operations, and condition monitoring in wind energy," *Energy*, vol. 60, pp. 1–12, Oct. 2013.
- [28] A. Kusiak and Z. Zhang, "Optimization of power and its variability with an artificial immune network algorithm," in *Proc. IEEE/PES PSCE*, Mar. 2011, pp. 1–8.
- [29] S. Li, D. Wunsch, E. O'Hair, and M. Giesselmann, "Comparative analysis of regression and artificial neural network models for wind turbine power curve estimation," *J. Solar Energy Eng.*, vol. 123, no. 4, pp. 327–332, Jul. 2001.
- [30] I. Colak, S. Sagiroglu, and M. Yesilbudak, "Data mining and wind power prediction: A literature review," *Renew. Energy*, vol. 46, pp. 241–247, Oct. 2012.
- [31] N. Dervilis *et al.*, "An SHM view of a CFD model of Lillgrund wind farm," in *Applied Mechanics and Materials*, vol. 564. Durnten, Switzerland: Trans Tech Publications, Jun. 2014, pp. 164–169.
- [32] V. Thapar, G. Agnihotri, and V. Sethi, "Critical analysis of methods for mathematical modeling of wind turbines," *Renew. Energy*, vol. 36, no. 11, pp. 3166–3177, Nov. 2011.
- [33] S. Shokrzadeh, M. Jafari Jozani, and E. Bibeau, "Wind turbine power curve modeling using advanced parametric and nonparametric methods," *IEEE Trans. Sustain. Energy*, vol. 5, no. 4, pp. 1262–1269, Oct. 2014.
- [34] G. Kariniotakis, G. Stavrakakis, and E. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE Trans. Energy Convers.*, vol. 11, no. 4, pp. 762–767, Dec. 1996.
- [35] S. Li, D. Wunsch, E. O'Hair, and M. Giesselmann, "Using neural networks to estimate wind turbine power generation," *IEEE Trans. Energy Convers.*, vol. 16, no. 3, pp. 276–282, Sep. 2001.
- [36] S. Gill, B. Stephen, and S. Galloway, "Wind turbine condition assessment through power curve copula modeling," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 94–101, Jan. 2012.
- [37] A. Kusiak, H. Zheng, and Z. Song, "On-line monitoring of power curves," *Renew. Energy*, vol. 34, no. 6, pp. 1487–1493, Jun. 2009.
- [38] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Appl. Energy*, vol. 98, pp. 574–583, Oct. 2012.
- [39] A. Kusiak and A. Verma, "Monitoring wind farms with performance curves," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 192–199, Jan. 2013.
- [40] J. A. Dahlberg, "Assessment of the Lillgrund Windfarm: Power Performance, Wake Effects," Vattenfall AB, Solna, Sweden, Tech. Rep., Sep. 2009.
- [41] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, Nov. 1995.
- [42] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, vol. 4, no. 4, 2006.
- [43] I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*. London, U.K.: Springer-Verlag, 2004.
- [44] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, 1993.
- [45] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [46] D. C. Montgomery, *Introduction to Statistical Quality Control*, 4th ed. New York, NY USA: Wiley, 2001.
- [47] E. Castillo, *Extreme Value Theory in Engineering*. New York, NY, USA: Academic, 1988.
- [48] R. Fisher and L. Tippett, "Limiting forms of the frequency distributions of the largest or smallest members of a sample," *Proc. Cambridge Philos. Soc.*, vol. 24, pp. 180–190, 1928.
- [49] R. Storn and R. Price, "Differential evolution—A simple and efficient heuristic for global optimisation over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, Dec. 1997.
- [50] K. Worden, G. Manson, H. Sohn, and C. Farrar, "Extreme value statistics from differential evolution for damage detection," in *Proc. 23rd Int. Modal Anal. Conf.*, 2005, pp. 1–10.
- [51] K. Worden, D. Allen, H. Sohn, D. Stinemat, and C. Farrar, "Extreme value statistics for damage detection in mechanical structures," Los Alamos Nat. Lab. Los Alamos, NM, USA, Tech. Rep. LA-13903-MS, Aug. 2002.

Authors' photographs and biographies not available at the time of publication.