

RESEARCH ARTICLE

A mapping method for anomaly detection in a localized population of structures

Weijiang Lin^{1,*} , Keith Worden¹, Andrew E. Maguire² and Elizabeth J. Cross¹

¹Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom

²Vattenfall Research and Development, New Renewables, The Tun Building, Holyrood Road, Edinburgh EH8 8AE, United Kingdom

*Corresponding author. E-mail: wlin17@sheffield.ac.uk

Received: 02 September 2021; **Revised:** 17 May 2022; **Accepted:** 23 June 2022

Keywords: Population-based structural health monitoring; anomaly detection; environmental and operational variations; Gaussian process regression; wind farm power modeling

Abstract

Population-based structural health monitoring (PBSHM) provides a means of accounting for inter-turbine correlations when solving the problem of wind farm anomaly detection. Across a wind farm, where a group of structures (turbines) is placed in close vicinity to each other, the environmental conditions and, thus, structural behavior vary in a spatiotemporal manner. Spatiotemporal trends are often overlooked in the existing data-based wind farm anomaly detection methods, because most current methods are designed for individual structures, that is, detecting anomalous behavior of a turbine based on the past behavior of the same turbine. In contrast, the idea of PBSHM involves sharing data across a population of structures and capturing the interactions between structures. This paper proposes a population-based anomaly detection method, specifically for a localized population of structures, which accounts for the spatiotemporal correlations in structural behavior. A case study from an offshore wind farm is given to demonstrate the potential of the proposed method as a wind farm performance indicator. It is concluded that the method has the potential to indicate operational anomalies caused by a range of factors across a wind farm. The method may also be useful for other tasks such as wind power and turbine load modeling.

Impact Statement

In light of the rapid growth of the offshore wind sector, monitoring strategies for wind farms becomes more crucial than ever. This paper introduces a mapping method for anomaly detection across a wind farm, which is premised on the idea of population-based structural health monitoring. The wake effect across a wind farm results in spatiotemporal correlations between turbine behaviours, which are captured by the mapping method. A case study is included to demonstrate the method's ability to detect performance anomalies arising from various causes. The method may also be adapted in future to tasks such as power maximisation and load reduction for wind farms.

1. Introduction

Population-based structural health monitoring (PBSHM) allows the health monitoring of a structure to be improved or enabled by using data from other structures within the same population (Worden et al., 2015; Bull et al., 2020; Gardner and Worden, 2020; Gosliga et al., 2020a,b; Lin et al., 2020b; Worden, 2020; Wickramarachchi et al., 2021). A wind farm demonstrates a *localized homogeneous population*,

where a population of nominally identical structures is located in close vicinity to each other, allowing them to share the same environment. According to the *wake effect*, spatiotemporal correlations exist in the environmental and operational conditions across a wind farm, which in turn affect turbine behavior. Such spatiotemporal variations may cause some turbines to be more prone to damage than others in the same farm; for example, a turbine subject to more turbulence fluctuations may be more severely affected by fatigue loading. However, the majority of existing methods for wind farm anomaly detection adopt a single-structure approach, which detects future anomalies based on the past behavior of the same turbine, without considering the interactions between turbines. This paper addresses this limitation by proposing a population-based anomaly detection method that accounts for the spatiotemporal variations across a wind farm.

Anomaly detection using wind turbine supervisory control and data acquisition (SCADA) data has been investigated over the last decade for the purposes of both wind integration studies and predictive maintenance. The former studies the effect of integrating wind power into existing electrical systems, where anomaly detection techniques are used to remove abnormal wind power data in order to obtain an accurate overview of the statistical properties of real turbine production (Dowds et al., 2015). The latter is a proactive, cost-effective maintenance strategy that predicts when maintenance should be performed, based on a data-driven assessment of turbine conditions. In this context, anomaly detection techniques are used to detect turbine performance anomalies, especially as early signs of failure (Tautz-Weinert and Watson, 2016). This concept of anomaly detection in predictive maintenance coincides with that of damage detection in structural health monitoring (SHM), where, in SHM, the type of anomalies to be detected refers to structural damage that potentially leads to failure (Farrar and Worden, 2013). Nonetheless, the methodologies and approaches are the same across various application domains (Chandola et al., 2009). This paper, therefore, aims to develop a novel method of anomaly detection in the field of SHM, specifically in the subdivision of PBSHM, which is applicable to detecting performance anomalies across a wind farm.

In the context of SHM, the spatial and temporal variations caused by the wake effect are referred to as *environmental and operational variations* (EOVs), which are confounding factors that influence both the damage-sensitive features and structural responses in a causal relationship (Farrar and Worden, 2013). The existing data-based approaches to detecting anomalies disguised in EOV are classified into two categories. The first category removes the confounding influence from the data using methods such as projection and cointegration, whereas the second category models the structure in the data, for example, capturing the relationship between EOVs and structural response (Cross, 2012; Farrar and Worden, 2013). This paper focuses on the second approach—EOV modeling—in order to exploit the EOV data that are available in the SCADA system. However, the existing methods are not directly applicable to the current problem, as they adopt a single-structure approach. Hence, the authors propose a method with an additional spatial dimension (Lin et al., 2020a,b) to the existing EOV modeling methods, which focus mainly on temporal changes in EOV and structural response. This proposed method addresses EOV modeling in PBSHM, especially for anomaly (damage) detection in a spatiotemporal setting.

Power-curve monitoring is a method of wind farm performance indication that, like EOV modeling, models the correlation between the environment (wind speed) and performance (power). Previous studies have developed power-curve monitoring as a technique in both PBSHM (Papatheou et al., 2015; Bull et al., 2020) and SHM (Rogers et al., 2020). A power curve describes the dependency of power generated by a wind turbine upon ambient wind speed. Theoretically, turbines with the same design and configuration should show a similar shape of the power curve. It is assumed that a power curve significantly distorted from the normal shape indicates anomalous turbine performance. The power-curve methods model the temporal variations of power in relation to wind speed, while accounting for the spatial variations as part of the *population variance*, which also consists of manufacturing tolerances and varying operational settings (Papatheou et al., 2015; Bull et al., 2020). To exploit the availability of EOV measurements, a new approach is to model the spatial and temporal correlations within a population, which is the focus of this paper.

There are studies of SCADA-based anomaly detection that address the farmwise spatiotemporal correlations, but only in a limited number of cases—using multilevel statistical analysis (Gonzalez et al., 2018) or a pattern recognition algorithm called symbolic dynamics filtering (Yang et al., 2018). However, none of these approaches model the functional relationship between environmental variables and turbine performance. Therefore, this paper brings the idea of EOV modeling into SCADA-based anomaly detection, with a particular focus on the spatiotemporal correlations in data.

As previously mentioned, the proposed method is not limited to being a performance indicator for wind farms. Since most anomaly detection methods can be used across various application domains (Chandola et al., 2009), the proposed method, too, has the potential to be used for damage detection if a damage-sensitive feature, such as natural frequencies, is used. The only prerequisite for the proposed method is to assume spatial and temporal correlations across a population of structures.

The layout of this paper is as follows. Section 1.1 provides an overview of the existing physics-based methods that model the spatiotemporal variations inside a wind farm and outlines the difference between the proposed method and the physics-based models. Section 2 defines the mapping method proposed in this paper, with detailed descriptions of the algorithm used, the functional relationships captured, the discrepancy measures, and the detection thresholds. A case study is given in Section 3, demonstrating the model's capability as a wind farm performance indicator. Section 4 discusses the limitations of the method and potential improvements, and Section 5 provides concluding remarks for the paper.

1.1. Related work

Physics-based models, which simulate the spatiotemporal variations across turbine arrays caused by the wake effect, have been developed for decades. Models of this type are referred to as *wake models*. The existing wake models can be subdivided into two main categories: analytical and numerical. Analytical wake models generally aim to compute only the relative drop in wind speed or power of a downstream turbine with regard to its upstream neighbor, whereas wake models employing computational fluid dynamics (CFD) provide an estimate of the entire vector field of flow velocity across a wind farm, together with estimates of other quantities, such as pressure or temperature, by numerically solving the Navier–Stokes equations (Archer et al., 2018).

Wake effect within wind farms, in short, describes the phenomena that involve reduced wind speed and more intensified turbulence after the wind passes through a turbine rotor. The physics behind such phenomena includes how the wind drives the motion in a turbine rotor and, subsequently, the rotation in a gearbox and/or generator. This process is dependent on factors such as atmospheric properties (e.g., air density), wind flow characteristics (e.g., wind speed), turbine design specifications (e.g., blade aerodynamics and drivetrain losses), and control settings (e.g., blade angle of attack). The effect of the wind on the turbine, in turn, results in disruption to the wind flow, forming a *wake* behind the rotor, which is composed of slower and more turbulent wind. The flow field immediately downstream of the turbine is also significantly influenced by the interaction between the rotor and tower/nacelle, with those phenomena yet to be fully understood (OBrien et al., 2017). Further downstream, the wake expands, meanders (having a fluctuating wake trajectory at any time instance), and finally dissipates. Apart from the operating conditions of its parent turbine, the wake progression is also affected by the ambient environment and the existence of other wakes (OBrien et al., 2017; Archer et al., 2018). The level of complexity involved in the creation and progression of a single wake is already very difficult to model physically, not to mention the combined effect of multiple wakes in a wind farm. Hence, existing physics-based models are often designed by confining the scope of the problem to fit specific applications.

One of the most important applications of wake models is in the optimization of wind farm layout. The task is to find a farm layout that yields the best overall power production out of all potential candidates while considering technological, economic, environmental, ecological, esthetic, and social constraints. The standard approach to this task includes the assessment of all candidate layouts, which would only be practical using computationally inexpensive analytical models. However, model accuracy tends to deteriorate once turbine layouts and/or environmental conditions differ from the benchmark scenarios.

The major types of analytical wake models were evaluated for operational wind farms of a range of sizes, terrains, and layouts in a review given by Archer et al. (2018).

Given the trend toward increased turbine sizes to meet the higher demand for wind energy, a second important application of wake models is the improvement of blade design. A larger turbine requires blades that are longer and yet more lightweight. Such new blade designs, however, can be prone to aeroelastic deformation, which inevitably complicates the structure of a turbine wake, especially under the combined effect of individual blade control and rotor–tower interaction (OBrien et al., 2017). CFD-based wake models are the only options with the potential to provide sufficient accuracy. Not surprisingly, the extremely large computational resources required by CFD-based models make them infeasible for most industrial implementations, and even some research cases. Because of cost limitations, such numerical models are only available for the simplest environmental and operational conditions. According to a review by O’Brien et al. (2017), the development of advanced 3D numerical approaches to wake modeling is still “in its infancy,” with enormous potential as well as formidable challenges ahead.

Since the aforementioned wake models are mainly created for design purposes, these models need to be adapted or repurposed for application to SHM, that is, to create a model that can assess current structural conditions based on monitoring data. Assuming that it is possible to repurpose the models, difficulties may arise from high computational burden (if using CFD-based models) or from poor accuracy (if analytical models are chosen), especially in scenarios when turbulence is playing a bigger role. For these reasons, a data-based model, such as the one used in the mapping method, is more likely to produce predictions that supply the accuracy required for monitoring purposes at a relatively lower cost. A data-based framework also allows for more flexible choices of features, so that damage-sensitive features can be used in place of power. As far as the authors are aware, the mapping method described in this paper marks the first venture in the field of SHM to create a data-based model for the spatiotemporal correlations resulting from the wake effect in wind farms.

2. The Mapping Method: A Spatiotemporal Anomaly Detection Method

The current work proposes a method that maps the spatiotemporal variations in EOVs onto an EOV-affected feature; thus, the proposed method is given the name *mapping method*. A key assumption is that the spatiotemporal pattern in the chosen feature will be maintained unless performance anomalies (including damage) occur. The method consists of two steps. First, a data-based model is trained to capture the normal spatiotemporal pattern, which is then used to predict the normal turbine behavior during a test period. Second, the discrepancy between the test predictions and measurements is quantified, and anomalies are detected when the discrepancy value is above a preset threshold.

2.1. A model based on Gaussian process regression

To model the normal spatiotemporal pattern across a wind farm, a flexible regression algorithm is required. In this work, the model of choice is Gaussian process regression (GPR).

Over the last decade, GPR-based methods have gained huge popularity in the SHM community (Cross, 2012; Papatheou et al., 2015; Bull et al., 2020; Lin et al., 2020b; Rogers et al., 2020). As a probabilistic model, a GPR is able to provide a predictive distribution, which indicates both the mean predicted values and their confidence interval (Rasmussen and Williams, 2006). In the context of PBSHM, the predictive distribution also accounts for the population variance, that is, the difference between individual structures within a population that arises from benign sources, such as turbulent boundary conditions and manufacturing tolerances, so that a general *form* can be established across multiple structures of the same “type” (Bull et al., 2020). Another advantage of GPR lies in the fact that it is nonparametric and well suited for learning complex relationships, as opposed to a parametric model whose complexity may be restricted by a predefined functional form.

As an extension to multivariate Gaussian distributions that describe random variables of finite dimensions, a GPR is held over functions, with the function values $f(X)$ specified at a potentially infinite

number of inputs X . The fact that a GPR can provide consistent inferences over any finite number of points makes it a powerful tool to learn functional relationships (Rasmussen and Williams, 2006). A GPR is specified by a mean and a covariance function:

$$f(X) \sim \mathcal{GP}(m(X), k(X, X')). \quad (1)$$

Conventionally, the mean function is assumed to be zero to simplify the formulation, and the covariance function takes the form of *squared exponential* (SE), which is the most commonly used kernel in describing smooth data variations:

$$k_{\text{SE}}(X, X') = \sigma_f^2 \exp\left(-\frac{1}{2l^2}|X - X'|^2\right). \quad (2)$$

The process of model training involves the optimization of hyperparameters, which, according to Eq. (2), include process variance σ_f^2 and input length scale l . Note that characteristic length scales are used in the current model, meaning that each input dimension is assigned a different length scale; for input data with a dimension D , there is a total of D length scales l_1, \dots, l_D . These hyperparameters are optimized here by maximizing the marginal likelihood via gradient-based optimization methods (Rasmussen and Williams, 2006).

With the optimized hyperparameters, a predictive distribution corresponding to the unseen inputs, X^* , can be given by specifying the predictive mean vector, $\bar{\mathbf{f}}^*$, and covariance matrix, $\text{cov}(\mathbf{f}^*)$:

$$\bar{\mathbf{f}}^* = K(X^*, X) [K(X, X) + \sigma_n^2 I]^{-1} \mathbf{y}, \quad (3)$$

$$\text{cov}(\mathbf{f}^*) = K(X^*, X^*) - K(X^*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X^*), \quad (4)$$

where $K(\cdot, \cdot)$ denotes a covariance matrix obtained from Eq. (2), and σ_n^2 is the noise variance associated with the noisy observations \mathbf{y} . Since the goal here is to predict the noisy target \mathbf{y}^* , the covariance for noisy predictions should be $\text{cov}(\hat{\mathbf{y}}^*) = \text{cov}(\mathbf{f}^*) + \sigma_n^2 I$.

For more details on GPR formulation/implementation, readers are referred to Rasmussen and Williams (2006).

2.2. Functional relationships captured by the model

As mentioned earlier, the mapping method aims to capture the spatiotemporal correlation between the EOVs and turbine performance. The choices of features are given as follows: power is chosen as the target feature, since it is a direct result of environmental variations as well as operational control, and, consequently, a direct measurement of turbine performance; features representing the EOVs include the spatial coordinates at all turbine locations and the wind speed at a subset of locations. A GPR model is trained to predict the power time series at all turbine locations in a wind farm, from the corresponding spatial coordinates and wind speed at a subset of locations. In other words, the model intends to capture the wind–power relationship and how it is affected by the wake-induced correlation across space and time.

The wind–power correlation can be summarized by a power curve, with an example shown in Figure 1. It is seen that power is strongly correlated with wind speed, forming an approximately sigmoidal (or cubic within a specific range) relationship. The shape of the curve is piecewise as a result of the turbine control strategy; thus, the modeling of a full power curve is a complicated task in itself. The use of a simple GPR model, without incorporating physical knowledge or partitioning the input space, might result in reduced accuracy in regions of control intervention (e.g., above-rated power; Rogers et al., 2020). Therefore, it is unlikely that a simple GPR is able to capture the full wind–power correlation as well as the spatiotemporal correlation induced by wakes. As a result, the data used in the following analysis belong to a restricted set of operating conditions, in which the wind speed variations are kept within the range in-between the cut-in

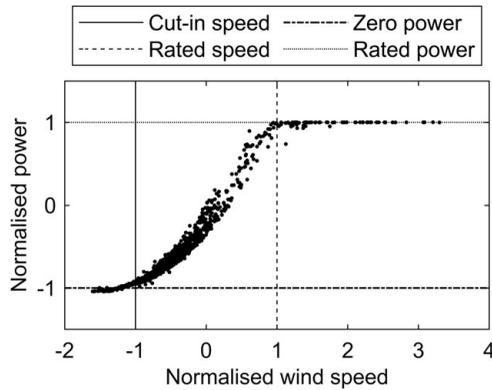


Figure 1. An example of a power curve.

and rated values (Figure 1). An extension to modeling the full wind–power correlation may be useful once the current model is proved valid.

At the risk of repetition, it should be clarified that power-curve monitoring is not part of the aim of the mapping method (and the turbine-by-turbine wind–power correlation is not modeled directly). Instead, wind speed and power are the chosen features that represent EOVs and structural response, respectively. The aim of the mapping method is to achieve EOV modeling in a spatiotemporal setting, and the aim of this paper is to demonstrate the potential of this method as a performance indicator for a wind farm.

The GPR model not only captures how power changes with wind speed temporally, but how this correlation varies spatially. The spatial aspect comes from the wake effect in the wind farm, which affects wind speed and power in a similar manner. As illustrated in Figure 2, the maximum power output in a farm is found at the front row of turbines; the power levels decrease as the wind progresses further downstream across the farm. This spatial correlation is modeled by using as input variables (a) the spatial coordinates of all turbine locations and (b) the wind speed time series at a fixed set of reference locations. The reference locations are selected to ensure that the model learns the spatial and temporal correlations through interpolation (i.e., to avoid extrapolation).

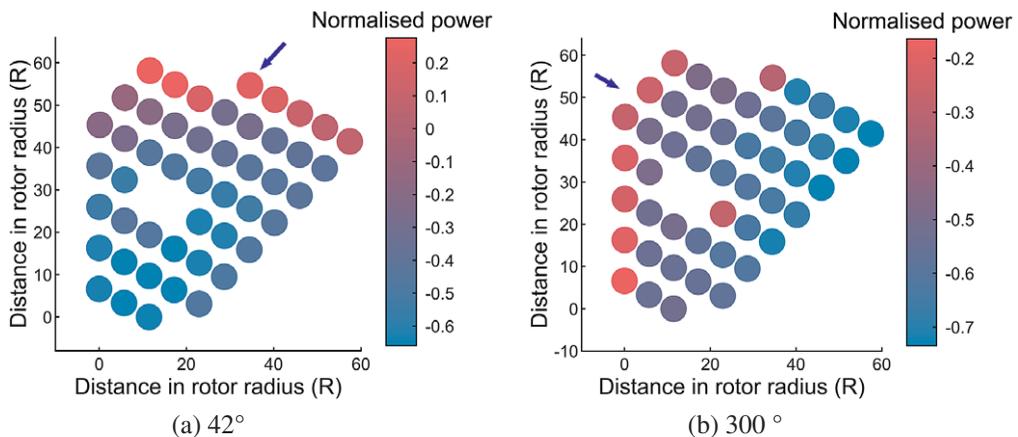


Figure 2. Example power variations across the Lillgrund wind farm corresponding to different wind directions, which are indicated by the blue arrows.

Figure 2 also shows that the spatial pattern is sensitive to wind direction, which changes with time. This time-varying spatial pattern is also captured by the model.

To conclude, the GPR model captures (a) the correlation between wind speed at reference locations and power at each output turbine, (b) how the wind–power correlation varies spatially, and (c) how the spatial pattern changes with wind direction (which is time-varying).

2.3. Discrepancy measures and thresholds

After establishing a model of normal condition, the expected normal turbine performance during a test period can be predicted, and the discrepancy between the prediction and the measurement is computed. As a population-based method, the mapping method detects anomalies via comparisons across the structures in a population. Therefore, for each turbine, the difference between the predicted and measured power time series needs to be summarized into a single discrepancy value to facilitate the comparison between turbines. Two discrepancy measures are used for the purpose of this study, namely *normalized mean-squared error* (NMSE) and *mean standardized log loss* (MSLL), which are error metrics commonly used for model quality evaluation in the literature of GPR.

The NMSE is given by

$$\text{NMSE}_s = \frac{100}{T\sigma_{y^*}^2} \sum_{t=1}^T \left(y_{t,s}^* - \bar{f}_{t,s}^* \right)^2, \quad (5)$$

where y and \bar{f} denote the measured and predicted (mean) target variables, respectively. The superscript * indicates the test dataset, and the subscripts t and s denote the time instance $t = 1, \dots, T$ and the turbine index $s = 1, \dots, S$. The mean-squared error term is normalized by the term $100/\sigma_{y^*}^2$. $\sigma_{y^*}^2$ represents the variance of the measured target variable in the test set across all turbines, in order to remove error sensitivity to the scale of testing data. The scaling factor 100 converts the error values into percentages. A score of 100% means that the predictions are equivalent to the target mean of the test data. The threshold of NMSE is therefore set to be 100%, below which the model provides better predictions than the mean of test data and is assumed to have captured some correlations between input and output. Note that the calculation of NMSE does not include the information about GPR predictive variance.

The second metric, MSLL, includes both the mean and variance of the GPR predictive distribution in its formulation (Rasmussen and Williams, 2006):

$$\text{MSLL}_s = \frac{1}{2T} \sum_{t=1}^T \left(\log \left(2\pi\sigma_{\hat{y}_{t,s}}^2 \right) + \frac{\left(y_{t,s}^* - \bar{f}_{t,s}^* \right)^2}{\sigma_{\hat{y}_{t,s}}^2} \right) - \frac{1}{2T} \sum_{t=1}^T \left(\log \left(2\pi\sigma_y^2 \right) + \frac{\left(y_{t,s}^* - \bar{y} \right)^2}{\sigma_y^2} \right). \quad (6)$$

The first term in Eq. (6) is the negative log probability that indicates how likely the target measurements are to be predicted using the trained model. Here, the GPR-predicted variance for a noisy target is $\sigma_{\hat{y}_{t,s}}^2 = \text{diag}(\text{cov}(\hat{\mathbf{y}}^*)) = \text{diag}(\text{cov}(\mathbf{f}^*) + \sigma_n^2 \mathbf{I})$. The second term in Eq. (6) represents the (negative log) probability obtained by treating the training mean \bar{y} and variance σ_y^2 as the model. It standardizes the first term such that $\text{MSLL} > 0$ if the GPR predictions are worse than the trivial model of training mean and variance. The threshold of MSLL is set to be 0, below which the model prediction is considered acceptable.

In the analysis that follows, it is demonstrated that the abovementioned discrepancy measures, NMSE and MSLL, and their corresponding thresholds can be used for anomaly detection.

3. A Case Study on an Offshore Wind Farm

The objective of this case study is to demonstrate the potential applicability of the mapping method to the problem of wind farm anomaly detection using SCADA data. Section 3.1 summarizes the data used for

model training and testing. It is followed by testing results which demonstrate (a) the model’s capability to predict power production with adequate accuracy (Section 3.2) and (b) the potential to use the model as a performance indicator for wind farms (Section 3.3).

3.1. Datasets

The SCADA data are collected from an offshore wind farm, Lillgrund. In this study, *time series data* refer to a series of mean values summarizing every 10-minute section of the raw data streams. Detailed description and analysis of the wind farm can be found in Jeppsson et al. (2008) and Dahlberg (2009).

A GPR-based model of normal condition is trained to predict power from spatial coordinates and wind speed, as mentioned in Section 2.2. One of the inputs to the model is the spatial locations, which are given as the Cartesian coordinates of each turbine position. Another input feature is the wind speed at a number of reference turbines. A set of 10 reference turbines is randomly selected across the wind farm, by manually choosing the minimum distance among them such that they are (roughly) evenly distributed across the entire space. For the purpose of this preliminary study, the only criterion behind the selection is to (visually) ensure that the reference locations are spread across the wind farm in order to avoid extrapolation in space, since a standard/zero-mean GPR model tends to predict less accurately during extrapolation (Cross and Rogers, 2021). More advanced methods for reference selection, such as Latin hypercube sampling, may be considered for future studies. The chosen reference locations are indicated by the rectangles in Figure 3.

The training and testing datasets are selected based on the following criteria. First, the data on wind speed and power used in the case study are confined to the range between cut-in and rated values (Figure 1), for the reasons mentioned in Section 2.2. Second, the GPR model includes the effect of multiple wind directions as part of the definition of normal condition; thus, the training data cover a range of wind directions. It is noted that the training data are selected so as to avoid rapid/extreme fluctuations in wind directions. If wind direction changes too rapidly, the turbines may not be given enough time to respond to the change, and the resulting spatial pattern may not be reflective of the wind direction. Given that the mapping method aims to capture how spatial patterns change with wind directions, data describing steady changes in wind directions are necessary for model training. A summary of the training

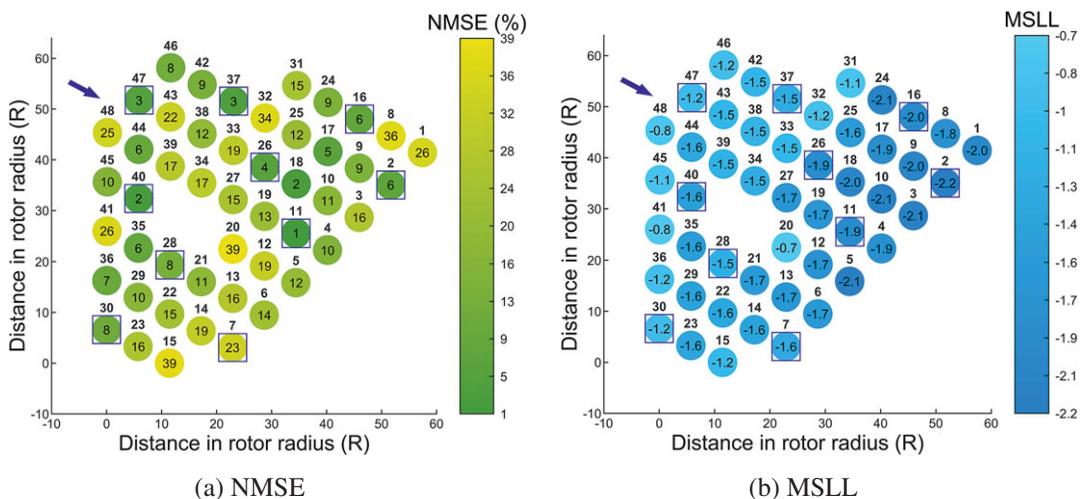


Figure 3. Maps of (a) NMSE and (b) MSLL averaged across a testing data period (of 2 hours 40 minutes) that describes normal operational conditions. The incoming wind direction is indicated by the blue arrow, and the reference turbine locations are marked with the blue rectangles. The turbine numbers (1–48) are also indicated, outside the circles.

Table 1. A summary of training and testing datasets.

	Normalized wind direction range	Normalized wind speed range	Normalized power range	Number of data points	Duration (hr)
Training sets	0.09–0.22	0.18–0.81	0.15–0.91	336	2.3
	0.10–0.22	0.24–0.91	0.21–0.95	312	2.2
	0.21–0.34	0.06–0.90	0.09–0.93	408	2.8
	0.23–0.36	0.28–0.98	0.18–0.95	360	2.5
	0.51–0.64	0.05–0.94	0.08–0.94	1,080	7.5
	0.51–0.64	0.09–0.88	0.10–0.84	408	2.8
	0.71–0.83	0.02–0.88	0.07–0.81	456	3.2
	0.72–0.86	0.21–0.89	0.19–0.95	312	2.2
Test Set 1	0.72–0.86	0.00–0.68	0.08–0.59	816	2.8
Test Set 2	0.02–0.94	0.09–0.72	0.00–0.64	624	2.2

and testing datasets used in the case study is given in Table 1. Note that the numbers of data points for the training period in Table 1 correspond to the numbers obtained after subsampling the original SCADA data by a factor of 2 (in order to reduce computation time), whereas the testing data are not subsampled. Thus, to make test predictions, the model interpolates not only in space but in time. The training data correspond to about 25.5 hours' worth of data for all 48 turbines in the Lillgrund wind farm, and the test datasets 1 and 2 correspond to 2.8 and 2.2 hours' worth of data, respectively. Note that the individual data sections are the longest continuous periods during which the mentioned criteria (of wind speed, power, and wind direction) are satisfied.

3.2. Capability to predict normal power production

The trained GPR model is used to predict the power output during a testing period, when all turbines are operating under normal conditions. This example refers to Test Set 1 in Table 1. The differences between measurements and predictions are summarized by the maps of NMSE and MSLL shown in Figure 3. It is seen that no specific spatial pattern can be found in the NMSE map (Figure 3a), whereas the MSLL results tend to get better toward the downstream side of the farm (Figure 3b). Such a spatial trend in the MSLL values is arguably caused by the calculation of the MSLL requiring an indirect comparison of the mean and variance between training data and GPR predictions; that is, the two terms in Eq. (6) correspond to the log loss obtained by comparing the testing data with GPR predictions and training statistics, respectively. Since the predicted mean power levels tend to decrease in the direction of the wind, the MSLL values change in the wind direction as well, but the way the MSLL changes depends on how far away the predictions are from the training mean. Note that the confidence levels of model predictions remain roughly constant throughout the farm. Overall, the error values are well below the corresponding thresholds (100% for NMSE; 0 for MSLL). At reference turbine locations, that is, where wind speed measurements are provided as the model inputs, there is a tendency to obtain relatively better NMSE results, while whether reference inputs are provided does not seem to affect the MSLL values. The difference between the two error metrics mainly stems from the fact that only the mean predictions are used in the NMSE, whereas both the predictive mean and variance are included in the calculation of the MSLL.

Examples of time series predictions are visualized in Figure 4, where the GPR prediction constitutes a mean and a confidence interval (given as three times the GPR predictive standard deviation $\sigma_{\hat{y}^*}$). Figure 4a illustrates one of the best predictions, excluding those at reference turbines. It is seen that the mean prediction follows the measured trend almost exactly, even at regions of relatively wider confidence intervals (from Data Point 11 onward). Figure 4b demonstrates the model prediction with a medium level of errors, where there are periods of mismatch between the mean prediction and the measurement, but the

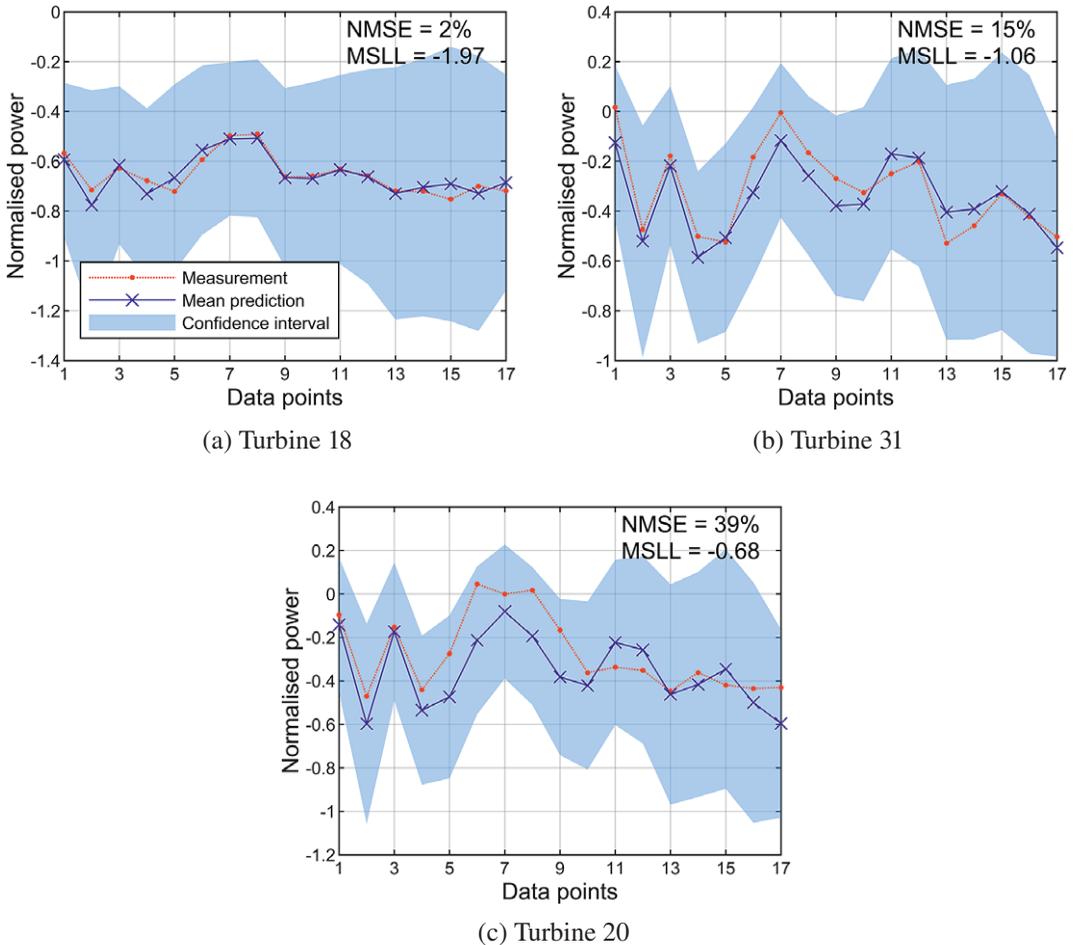


Figure 4. Examples of the predicted and measured power time histories for the normal testing set, in cases of (a) small, (b) medium, and (c) large errors.

amount of deviation is considered small compared with the confidence interval. The largest-error prediction is illustrated in Figure 4c; it can be seen that the mean prediction deviates from the measurement throughout the time window while still loosely following the trend. In regions of large deviation (Data Points 5–9), the measured data are close to the boundary of the confidence interval, which leads to the relatively large MSLL as well as NMSE. In summary, the model is able to provide mean predictions that follow (at least roughly) the measured trends, as well as confidence intervals that capture all the benign power variations in the given testing dataset.

3.3. Performance indicator for wind farms

Having established an appropriate model of normality, the model is used to predict the power production in a test set that contains potential performance anomalies. Figure 5 shows the discrepancy between predictions and measurements in terms of maps of NMSE and MSLL. The color schemes of the error maps are designed such that the turbines with an error value above the threshold are highlighted (in warm colors). The thresholds for NMSE and MSLL are 100% and 0, respectively, with reasons given in Section 2.3. It is seen that five turbines are highlighted as candidate anomalies in the NMSE map, that is, Turbines 22, 29, 33, 38, and 48, and there is one additional candidate anomaly, Turbine 25, highlighted in

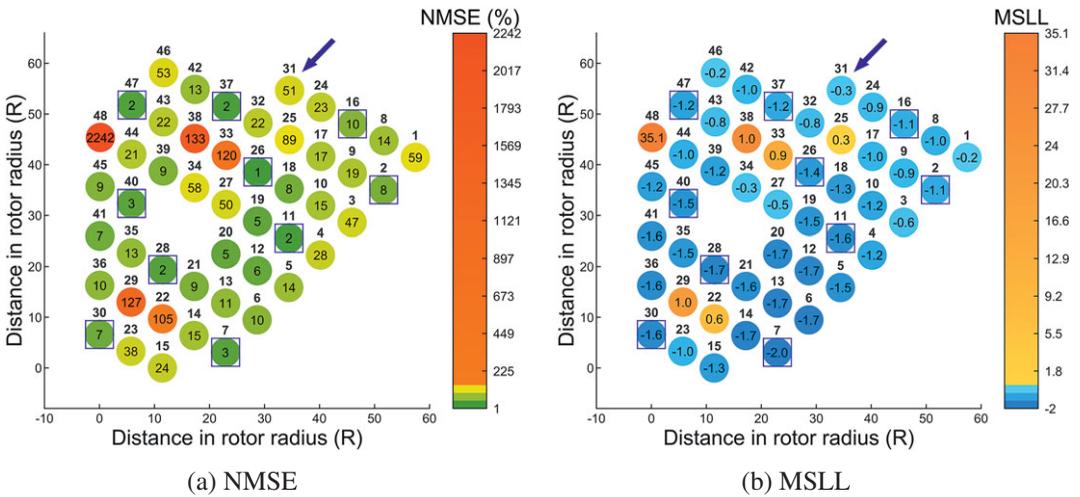
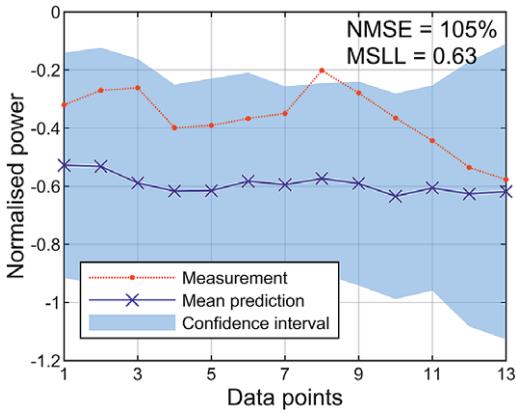


Figure 5. Maps of (a) NMSE and (b) MSLL averaged across a 2-hour testing period with potential anomalies. The incoming wind direction is indicated by the blue arrow, and the reference turbine locations are marked with the blue rectangles. The turbine numbers (1–48) are also indicated, outside the circles.

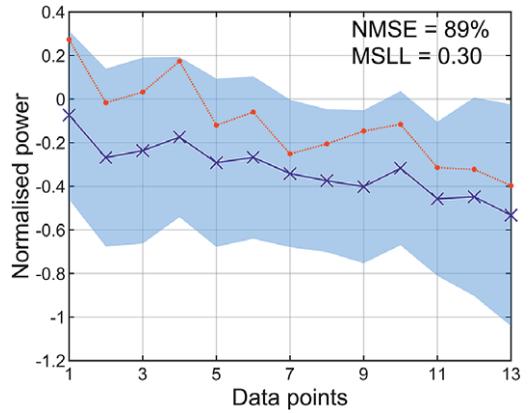
the MSLL map. Among these highlighted turbines, Turbine 48 is disqualified from being a potential anomaly, since it was manually switched off during the time period being considered. This powers down results in extremely high values of NMSE and MSLL. Although this is a known control action, it provides very basic validation that the model is able to detect anomalous behavior. The other highlighted turbines will now be investigated. Note that the dataset available is currently unlabeled; that is, the true reason for suspected anomalies is unknown—here some possible reasons for the observed behavior will be explored.

The time histories of the predicted and measured power for the candidate anomalies (Turbines 22, 25, 29, 33, and 38) are shown in Figure 6. For all candidates, there are obvious discrepancies between mean predictions and measurements throughout the time window. However, the trends of measured power are still captured by the confidence intervals most of the time. Interestingly, in all five cases, the model underpredicts the power produced, implying that there may be a factor that causes many turbines to produce more power than usual during this period of time. With respect to the error values, the further away the mean predictions are from the measurements, the higher the NMSE values, whereas higher MSLL values tend to result from scenarios when the measured data trends are consistently close to or beyond the uncertainty boundaries. In general, none of the candidate anomalies exhibits large fluctuations in power, and the relatively high errors seem to result from predictions with shifted means.

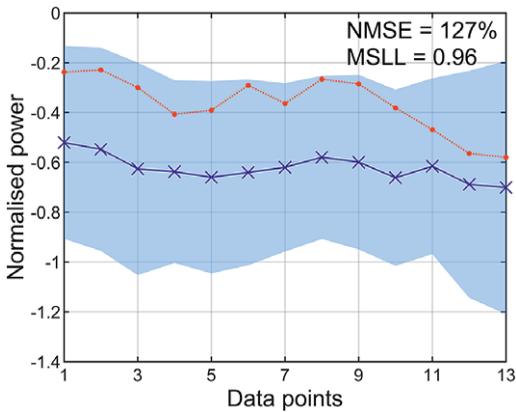
One reason why the actual power production deviates from model prediction is that the testing dataset describes a spatial correlation different from that depicted in the training data. This difference in spatial correlation is the main cause of the candidate anomalies at Turbines 25 and 38. In Figure 7, the training and testing wind–power correlations for the first three turbine rows, that is, the neighborhood of Turbines 25 and 38, are illustrated, with comparisons made available between training and testing data. A comparison between Figure 7a,b illustrates how the spatial correlations between Turbine 25 and its neighbors change from training to testing data. It is seen that, although Turbine 25 tends to generate relatively high power, given the same wind speed compared to its neighbors during training (Figure 7a), the amount of extra power it generates in the testing period becomes more distinct from the neighborhood (Figure 7b). However, what is shown in Figure 7b can (arguably) be merely a demonstration of the population variance; perhaps that is why Turbine 25 is flagged as a candidate anomaly by only one of the error metrics. For Turbine 38, the difference between training and testing data lies mainly in its relationship with the upstream Turbine 37. During training, the effect of wake shielding is apparent as the wind and power at Turbine 38 are significantly less than those at Turbine 37 (Figure 7c), which is not



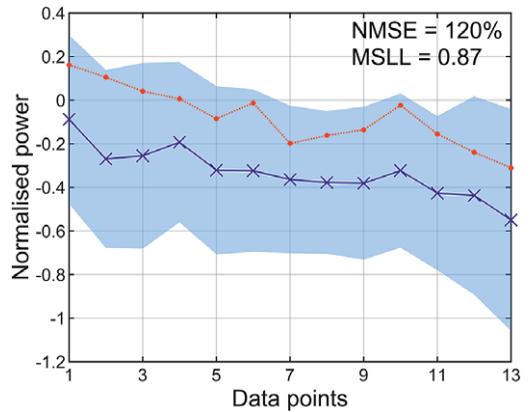
(a) Turbine 22



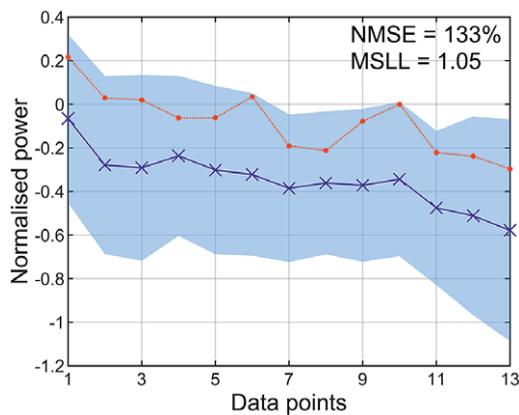
(b) Turbine 25



(c) Turbine 29

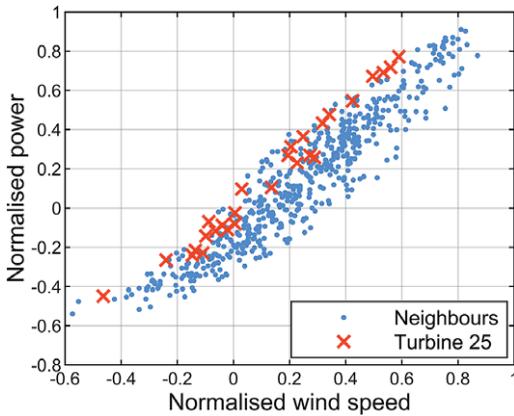


(d) Turbine 33

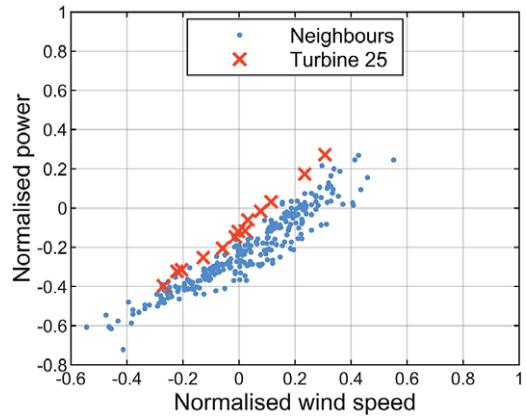


(e) Turbine 38

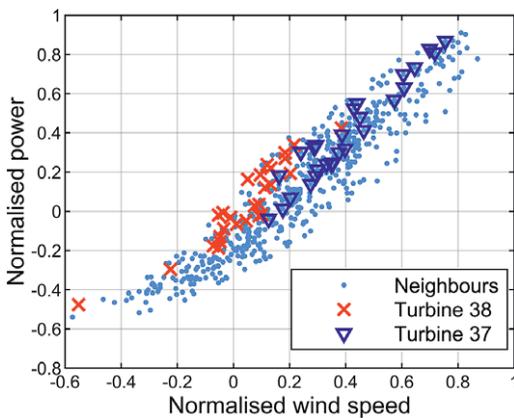
Figure 6. Time histories of the predicted and measured power for five highlighted candidate anomalies.



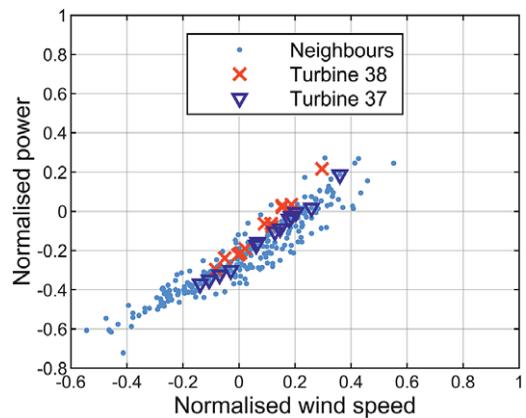
(a) Training wind-power correlations (with Turbines 25 highlighted).



(b) Testing wind-power correlations (with Turbines 25 highlighted).



(c) Training wind-power correlations (with Turbines 37 & 38 highlighted).



(d) Testing wind-power correlations (with Turbines 37 & 38 highlighted).

Figure 7. Wind-power correlations associated with (a)-(b) Turbine 25 and (c)-(d) Turbine 38. The “Neighbors” are referred to turbines in the neighborhood of Turbines 25 and 38, in this case the first three rows of turbines in the wind direction (excluding other candidate anomalies).

seen in the testing period (Figure 7d). Wake deviations due to yaw misalignment may be a reason for such difference, which remains hypothetical since the relevant data are unavailable to the authors. The difference in spatial pattern around Turbine 38 is captured by both error metrics.

Another cause of deviation from normal condition is an unexpected local nacelle direction. In particular, if a turbine turns away from the wake(s) of the upstream turbine(s), the power produced by the turbine may be underestimated by the model. Such an unexpected nacelle direction is found at Turbine 33. Normally, Turbine 33 should be affected by the combined wakes of the two upstream turbines (Turbines 31 and 32). However, during the testing period, this turbine is constantly facing Turbine 37, a region likely to be affected by fewer wakes. Therefore, Turbine 33 ends up with more power output than predicted, as seen in Figure 6d, and this overprediction has been successfully highlighted by both maps in MSL and NMSE.

The remaining candidate anomalies, Turbines 22 and 29, stand out from their neighbors, as they experience unexpectedly high wind speed and, thus, produce more power. Figure 8 shows the wind-power correlations of Turbines 22 and 29 in relation to those of the neighboring turbines. As part

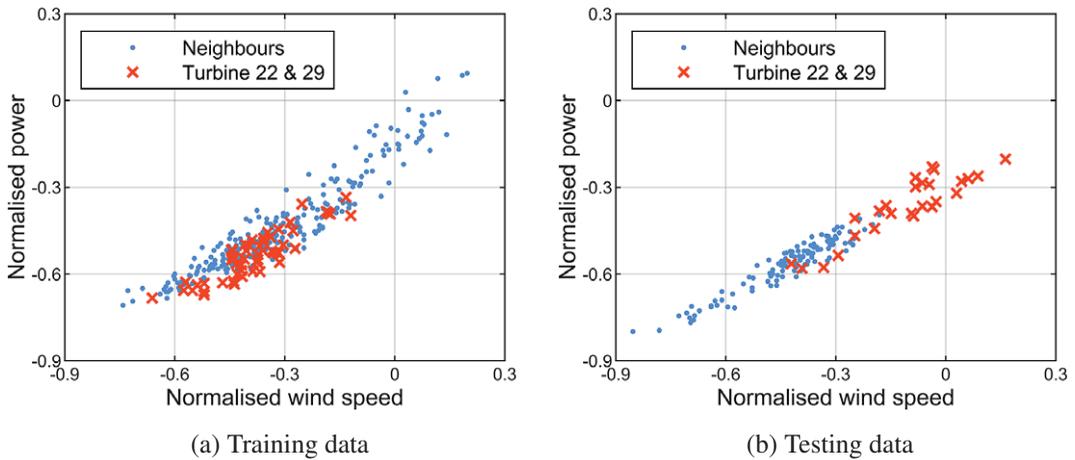


Figure 8. Wind-power correlation for a subset of turbines toward the back of the farm, that is, Turbines 13–15, 21–23, 28–30, 35, and 36, obtained from (a) training and (b) testing data.

of the normal spatial pattern acquired by the model, Turbines 22 and 29 generate relatively low power in comparison to their surrounding turbines (Figure 8a). In the testing period, the wind and power at the two turbines extend far beyond the ranges covered by their neighbors while maintaining the gradient of the wind–power correlations (Figure 8b). Therefore, the difference between training and testing data is likely to stem from changes in the environment rather than changes in the turbine systems. Upon further examination of the data under similar environmental and operational conditions, the unexpected high speeds at Turbines 22 and 29 are unlikely to be caused by unexpected nacelle direction or ambient temperature, nor are there seasonal or diurnal patterns in the occurrence of such phenomena. Given the existence of a central gap in the wind farm of interest (see Figure 5), more wind eddies may be generated around the gap region, bringing about more short-term turbulence that causes the occasional high wind around Turbines 22 and 29. The fact that these infrequent environmental variations are not described by the training data has led to the seemingly sensitive detection results. However, the occasional exposure to wind eddies that give rise to counterintuitive wake patterns may subject Turbines 22 and 29 to higher fatigue loads and, thus, a higher risk of anomalous performance. The model results, therefore, provide insight into this potential risk.

3.4. Effect of reference wind speed inputs

As previously stated, the mapping model is designed to predict the power across a wind farm, as a function of turbine locations and wind speed at a fixed set of reference turbines. In the previous case study, 10 reference turbines are chosen, with their locations indicated in Figures 3 and 5. This section focuses on investigating how the numbers of reference turbines may affect model accuracy. For a given number of reference turbines, the locations were randomly selected five times, and a GPR was trained each time. The NMSE results corresponding to different numbers of reference turbines are illustrated in Figure 9. It is seen that, as the number of reference turbines increases, there is an approximately exponential decrease in both the NMSE values and the error variances. If the reference locations are to be optimized, the NMSE improvements will be larger at fewer numbers of references. Optimized reference locations are considered to bring significant NMSE reduction when the number of references is smaller than or equal to 8. The average NMSE level for 10 references is around 30%. For the set of reference turbines used in the case study, the overall NMSE is around 20% across all normal testing datasets. Hence, in the current study, an optimized set of reference locations may not bring significant improvements in predictive accuracy.

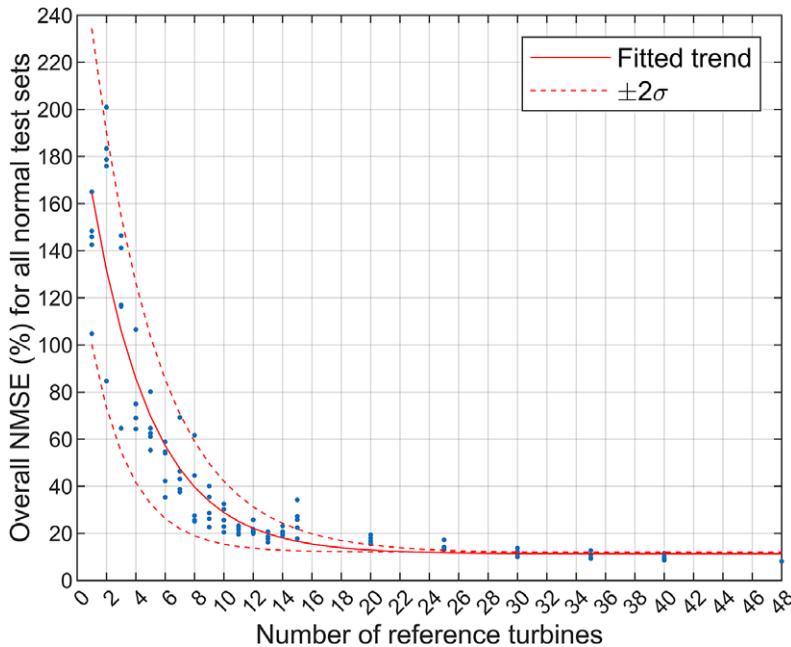


Figure 9. The NMSE for various numbers of reference turbines.

Nonetheless, an optimization algorithm, such as a genetic algorithm (Worden et al., 2008), will be necessary if fewer references are to be selected in future studies.

4. Discussion

In this paper, the mapping method is presented as a two-step process that includes modeling and detection. In the modeling step, a GPR model is trained, which captures the spatiotemporal variations of power across a wind farm under normal conditions. In the detection step, the discrepancy between predictions and measurements quantified, and the discrepancy values above thresholds are considered (potentially) anomalous.

The GPR model is associated with several assumptions. First, the functional relationships captured by the model are assumed stationary. It indicates that the covariance (within or between datasets) is a function of the difference between input points rather than their absolutes, which greatly simplifies the complexity of the model (Rasmussen and Williams, 2006). Another key assumption arises from the chosen SE covariance function, which intrinsically assumes *smoothness*. Although some argue that this function might be too smooth to reflect any realistic process, the SE covariance function is still commonly employed across disciplines (Rasmussen and Williams, 2006). The smoothness assumption is acceptable here, as no apparent discontinuity is found in the data. First, the training and testing data are confined to a specific range of wind speed and power that represents the continuous region in the wind–power correlation (Figure 1). Second, the spatial variations in power are smooth when all turbines are operating normally, which is supported by the illustration given in Figure 2. Third, the training data are selected to avoid extreme fluctuations in wind directions, in an attempt to avoid discontinuities. As seen in Table 1, although the training data do not cover the entire range of possible wind directions, they cover the wind direction range of the normal test set (Test Set 1). The unexpectedly wide wind direction range in Test Set 2 is a result of anomalous turbine angles as explained in Section 3.3. Therefore, the smoothness assumption holds for the given case study. Both mentioned assumptions are the most commonly used and give rise to one of the simplest forms of the GPR model. Since the aim of the current work is to prove

the concept of the proposed mapping method, it is reasonable to start with a relatively simple and easily accessible model.

Two types of discrepancy measures are used in the analysis, namely NMSE and MSLL. The NMSE focuses on how the mean predictions deviate from the measurements, whereas the MSLL also accounts for the position of the data trends in relation to the boundaries of the confidence intervals. In the case when two turbines return similar NMSE values, the MSLL results might help distinguish which turbine has the more confident prediction (and thus less likely to be erroneous). In summary, the two types of error metrics contain information that complements one another, and should be used in conjunction.

The detection thresholds used in the case study are chosen based on the (mathematical) definitions of NMSE and MSLL (Section 2.3). As a preliminary study, the current paper focuses on establishing the framework of the mapping method while keeping the individual elements, such as the form of the GPR model and detection thresholds, as simple as possible. Thus, this paper demonstrates the potential of the mapping method under the simplest possible setup, which justifies further investigation into this method. In future studies, more sophisticated methods of threshold selection, such as extreme value statistics (Farrar and Worden, 2013; Papatheou et al., 2017), may be considered.

The key assumption of the mapping method is that the spatiotemporal pattern of the target variable changes when anomalies occur. In Section 3.3, many of the candidate anomalies (Turbines 22, 25, 29, and 38) are justified by comparing their wind–power correlations with those of the surrounding turbines. Such anomalies would be difficult to detect without considering the interrelations between turbines. Hence, the mapping method, as a population-based approach, possesses a unique advantage in detecting anomalies.

On the other hand, the fact that some anomalies result from the difference between training and testing data raises the question of whether the pattern described by the testing data is necessarily an “anomalous” one. To reduce false positives (and negatives), training data should be representative of the conditions expected—as is the case with any data-driven model. In practice, this may be difficult given the range of conditions/benign variations that may occur. This issue suggests that a gray-box approach may be suitable (Cross and Rogers, 2021) and will be the topic of future work.

5. Conclusion

The aim of this paper is twofold. First, this paper proposes and develops a population-based anomaly detection method that accounts for the spatiotemporal correlations in environmental and operational conditions, by investigating the problem of anomaly detection across a wind farm. Second, this paper demonstrates the proposed method as a wind farm performance indicator. The proposed mapping method is a two-step process, in which, firstly, a GPR-based model of normal condition is trained and, secondly, anomalies are detected based on the discrepancy between model predictions and measurements.

The GPR model has demonstrated its ability to learn the functional relationships including:

1. the correlation between wind speed at reference locations and power at all locations,
2. how the wind–power correlation varies spatially, and
3. how the spatial pattern changes with time-varying wind directions.

As shown in the given case study, the candidate anomalies detected by the model can be a result of:

- powering off (unexpected wind–power correlation),
- unexpected turbine angles (unexpected spatial pattern), or
- spatially counterintuitive environmental conditions (unexpected spatial pattern).

In conclusion, this case study has successfully demonstrated the capability of the mapping model to identify performance anomalies across a wind farm, using wind turbine SCADA data.

Data Availability Statement. The data used in the case study were accessed under a confidentiality agreement with Vattenfall.

Author Contributions. Conceptualization: W.L., K.W., and E.J.C.; Data curation: A.E.M.; Formal analysis: W.L.; Funding acquisition: K.W. and E.J.C.; Methodology: W.L., K.W., and E.J.C.; Supervision: K.W. and E.J.C.; Writing—original draft: W.L.; Writing—review and editing: W.L., K.W., and E.J.C. All authors approved the final submitted draft.

Funding Statement. The authors would like to acknowledge the support of the EPSRC, particularly via grant reference numbers EP/R004900/1, EP/S001565/1, and EP/R003645/1. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Competing Interests. The authors declare no competing interests exist.

References

- Archer C, Vassel-Be-Hagh A, Yan C, Wu S, Pan Y, Brodie J and Maguire A** (2018) Review and evaluation of wake loss models for wind energy applications. *Applied Energy* 226, 1187–1207.
- Bull L, Gardner P, Gosliga J, Maguire A, Campos C, Rogers T, Haywood-Alexander M, Dervilis N, Cross E and Worden K** (2020) Towards population-based structural health monitoring, Part I: Homogeneous populations and forms. In: Mao, Z. (eds) *Model Validation and Uncertainty Quantification*, Volume 3. Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer.
- Chandola V, Banerjee A and Kumar V** (2009) Anomaly detection: A survey. *ACM Computing Surveys* 41, 15:1–15:57.
- Cross E** (2012) *On Structural Health Monitoring in Changing Environmental and Operational Conditions*. Ph.D. thesis, Department of Mechanical Engineering, University of Sheffield.
- Cross E and Rogers T** (2021) Physics-derived covariance functions for machine learning in structural dynamics. *IFAC-PapersOn-Line* 54(7), 168–173.
- Dahlberg J-A** (2009) *Assessment of the Lillgrund Windfarm: Power Performance and Wake Effects*. Technical report, Vattenfall Vindkraft AB.
- Dowds J, Hines P, Ryan T, Buchanan W, Kirby E, Apt J and Jaramillo P** (2015) A review of large-scale wind integration studies. *Renewable and Sustainable Energy Reviews* 49, 768–794.
- Farrar C and Worden K** (2013) *Structural Health Monitoring: A Machine Learning Perspective*. Hoboken, NJ: John Wiley and Sons.
- Gardner, P., Bull, L.A., Gosliga, J., Dervilis, N., Worden, K.** (2020). Towards Population-Based Structural Health Monitoring, Part IV: Heterogeneous Populations, Transfer and Mapping. In: Mao, Z. (eds) *Model Validation and Uncertainty Quantification*, Vol. 3. Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer.
- Gonzalez E, Tautz-Weinert J, Melero J and Watson S** (2018) Statistical evaluation of SCADA data for wind turbine condition monitoring and farm assessment. *Journal of Physics: Conference Series* 1037, 032038.
- Gosliga J, Gardner P, Bull L, Dervilis N and Worden K** (2021a) Towards population-based structural health monitoring, Part II: Heterogeneous populations and structures as graphs. In: Dilworth, B., Mains, M. (eds) *Topics in Modal Analysis & Testing*, Vol. 8. Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer.
- Gosliga J, Gardner P, Bull L, Dervilis N and Worden K** (2021b) Towards population-based structural health monitoring, Part III: Graphs, networks and communities. In: Dilworth, B., Mains, M. (eds) *Topics in Modal Analysis & Testing*, Vol. 8. Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer.
- Jepsson J, Larsen P and Larsson Å** (2008) *Technical Description Lillgrund Wind Power Plant*. Technical report, Vattenfall Vindkraft AB.
- Lin W, Worden K and Maguire AE** (2021) Towards population-based structural health monitoring, Part VII: EOv fields: Environmental mapping. In: Dilworth, B., Mains, M. (eds) *Topics in Modal Analysis & Testing*, Vol. 8. Conference Proceedings of the Society for Experimental Mechanics Series. The University of Sheffield: Springer.
- Lin W, Worden K, Maguire A and Cross E** (2020) Power mapping: A wind turbine performance indicator in population-based structural health monitoring. In *Proceedings of the 29th International Conference on Noise and Vibration Engineering*, ISMA.
- O'Brien J, Young T, OMahoney D and Griffin P** (2017) Horizontal axis wind turbine research: A review of commercial CFD, FE codes and experimental practices. *Progress in Aerospace Sciences* 92, 1–24.
- Papathéou E, Dervilis N, Maguire A, Antoniadou I and Worden K** (2015) A performance monitoring for the novel Lillgrund offshore wind farm. *IEEE Transactions on Industrial Electronics* 62, 6636–6644.
- Papathéou E, Dervilis N, Maguire A, Campos C, Antoniadou I and Worden K** (2017) Performance monitoring of a wind turbine using extreme function theory. *Renewable Energy* 113, 1490–1502.
- Rasmussen C and Williams C** (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.
- Rogers T, Gardner P, Dervilis N, Worden K, Maguire A, Papathéou E and Cross E** (2020) Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression. *Renewable Energy* 148, 1124–1136.
- Tautz-Weinert J and Watson S** (2016) Using SCADA data for wind turbine condition monitoring—A review. *IET Renewable Power Generation* 11, 382–394.

- Wickramarachchi C, Brennan D, Lin W, Maguire A, Harvey D, Cross E and Worden K** (2022) Towards population-based structural health monitoring, Part V: Networks and databases. In: Madarshahian, R., Hemez, F. (eds) *Data Science in Engineering*, Vol. 9. Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer.
- Worden, K.** (2021). Towards population-based structural health monitoring, Part VI: Structures as geometry. In: Pakzad, S. (eds) *Dynamics of Civil Structures*, Vol. 2. Conference Proceedings of the Society for Experimental Mechanics Series. Cham: Springer.
- Worden K, Cross E, Dervilis N, Papatheou E and Antoniadou I** (2015) Structural health monitoring: From structures to systems-of-systems. *IFAC-PapersOnLine* 48, 1–17.
- Worden K, Manson G, Hilson G and Pierce S** (2008) Genetic optimisation of a neural damage locator. *Journal of Sound and Vibration* 309, 529–544.
- Yang W, Liu C and Jiang D** (2018) An unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring. *Renewable Energy* 127, 230–241.