Dimensionality Reduction for the Analysis of Time Series Data from Wind Turbines

Jochen Garcke, Rodrigo Iza-Teran, Marvin Marks, Mandar Pathare, Dirk Schollbach, and Martin Stettner

Abstract We are addressing two related applications for the analysis of data from wind turbines. First, we consider time series data arising from virtual sensors in numerical simulations as employed during product development, and, second, we investigate sensor data from condition monitoring systems of installed wind turbines. For each application we propose a data analysis procedure based on dimensionality reduction. In the case of virtual product development we develop tools to assist the engineer in the process of analyzing the time series data from large bundles of numerical simulations in regard to similarities or anomalies. For condition monitoring we develop a procedure which detects damages early in the sensor data stream.

1 Introduction

Wind energy is nowadays a very important component of the energy production, capable of meeting 11.4% of the EU's electricity demand in 2015 [3]. By 2010, more than 71,000 wind turbines were designed, built, commissioned, and are still operating

Jochen Garcke
Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
Institute for Numerical Simulation, University of Bonn, Wegelerstr. 6, 53115 Bonn, Germany
e-mail: jochen.garcke@scai.fraunhofer.de

Rodrigo Iza-Teran, Marvin Marks, and Mandar Pathare
Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53757 Sankt Augustin, Germany

Dirk Schollbach
Weidmüller Monitoring Systems GmbH, Else-Sander-Straße 8, 01099 Dresden, Germany

Martin Stettner
GE Global Research, Freisinger Landstrasse 50, 85748 Garching bei München, Germany

in the EU[1]. We investigate two data analysis tasks arising in this domain. One is the analysis of data from numerical simulations of wind turbines during product design, certification, and commissioning, the other is the analysis of data arising during the condition monitoring of installed wind turbines.

First, the simulation of the non-linear dynamic response of numerical models of wind turbines is an essential element of design, modification, certification, and site-dependent load assessment [4]. In the course of design and optimization, hundreds of wind turbine configurations may be assessed. In order to evaluate a single configuration, thousands of transient simulations of the model's response to turbulent and deterministic wind in normal operation, in a storm, under mechanical stresses, in fault situations, and during the loss of connection to the power grid - and various combinations thereof - may be conducted. A single simulation typically generates outputs for several hundreds software "sensors" extracting information like local loads, deflections, velocities, and wind inflow from the simulation model, in most cases for simulated 10 minute real time intervals, sampled commonly at 10 Hz, resulting in 5 to 50 MB of data per simulation run. Mining and interpreting the gigabytes of data produced in this way represents a serious challenge.

Second, condition monitoring systems (CMS) rely on empiric models derived from physical sensor data. Applied to wind turbines, they have the goal of determining the health of the components, ensuring safety (e. g. by detecting ice build-ups on rotor blades), and reducing the wear of the mechanical parts. Diagnosis and fine tuning have become attainable thanks to the condition information provided by various sensors and collected by the turbine control station. Precise control strategies can be derived from this information, for example in order to reduce wear, hence prolonging a healthy state of the installation. The condition monitoring of rotating machine parts has a long tradition, whereas the monitoring of rotor blades with oscillation analysis has been studied far less deeply up until now. Therefore, we concentrate on the analysis of rotor blade oscillation data enriched with operational data of the wind turbine. The latter includes for example the meteorological conditions and the operational modes of the wind turbine. The sensor data is kept available for examinations at least as long as the wind turbine remains operational, which can be up to 20 years, to provide input for upcoming analysis and machine learning tasks. Therefore large data repositories can and need to be analysed to provide suistanable services to the wind turbine operators.

This article focuses on data analysis procedures which shall assist the engineer and simplify the overall workflow by providing novel, tailored meta data based on similarity and/or anomaly detection in raw output data. For the analysis of the arising data we investigate dimensionality reduction approaches which obtain a low dimensional embedding of the data. The standard dimensionality reduction method is principal component analysis (PCA), which gives a low-dimensional representation of data which are assumed to lie in a subspace. But if the data in fact resides in a nonlinear structure, PCA gives an inefficient representation with too many dimensions. Therefore, and in particular in recent years, algorithms for nonlinear dimensionality

[1] European Wind Energy Association `www.ewea.org/wind-energy-basics/faq`

reduction, or manifold learning, have been introduced, see e.g. [10]. The goal of these approaches consists in obtaining a low-dimensional representation of the data which respects the intrinsic nonlinear geometry.

In case of numerical simulations, a goal is to facilitate interpretation of the system response as obtained from a vast number of simulations. We use nonlinear dimensionality reduction for time series obtained from the numerical simulations and aim for the detection of patterns, anomalies, or similar behavior. For the condition monitoring, one looks for a small number of significant key indicators which describe the health of a wind turbine. Here, using linear dimensionality reduction, we are able to compute a low-dimensional basis which represents a baseline, deviations from it indicate anomalies in the sensor data, this information can then be used for predictive maintenance.

In Sect. 2 we describe the two application domains in more detail. Sect. 3 contains our data analysis approach for time series from numerical simulations, while in Sect. 4 we present our approach for anomaly detection in sensor data.

2 Time Series Characteristics in Wind Energy

We consider two different situations where time series data arise in wind energy research. First, during the product development, process data stems from numerical simulations of wind turbines. Second, data is obtained during the condition monitoring of installed wind turbines. Other time series data in the wind energy domain, for example, concerns wind speed prediction or wind turbine control [9].

2.1 Numerical Simulations of Wind Turbines

Numerical simulation of wind turbines is an essential element in the design of wind turbine systems, the fine-tuning of installations, or the exploration of the effect of upgrades [4]. Input meta data encompass environmental and operational conditions, referred to as Design Load Cases, DLCs, as specified by certification bodies, as well as instantiations of the wind turbine's individual components such as its foundation, tower, nacelle, drive train, generator, hub, blade, and control system. This information is commonly compiled in files identified by or populated with the aid of this meta data. See Fig. 1 for an overview of the components of the analysis workflow.

The simulation tool itself contains commonly a condensed dynamic representation of the structure, with simplified "engineering models" for the effective description of wind, aerodynamics, or actuator dynamics, based on first principle physics, tuned with field data and engineering knowledge, permitting simulation times on desktop machines which are about one order of magnitude smaller than real time. High performance computing solutions employing 3D computational fluid dynamics solutions such as the Navier-Stokes equations and detailed structural models (such as

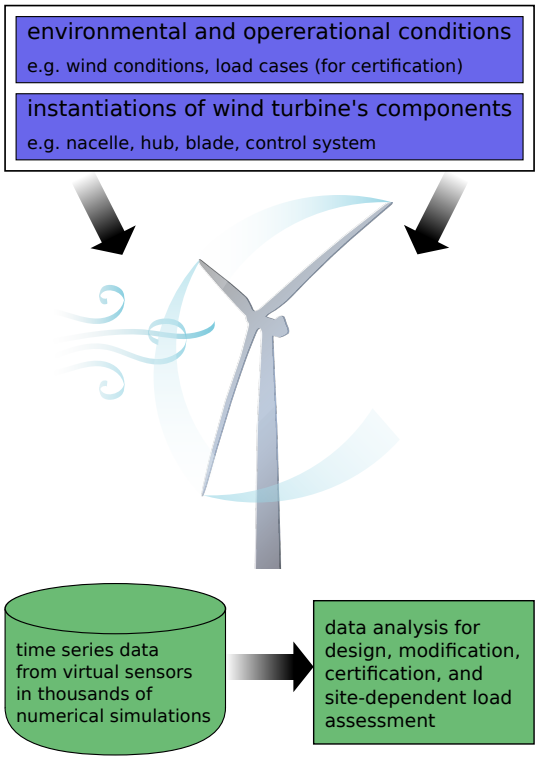


Fig. 1 In the design process of wind turbines, a large number of time series data is generated by numerical simulations, due to the needed investigations of different conditions and components affecting the behavior of the wind turbine in operation.

non-linear FE solvers) might be used during the development of this wind turbine system simulation tool, or in case of root cause analysis, but are very rarely used during the design process of a wind turbine. Such approaches are usually unsuitable because of excessive run times, amplified by the already described need to execute vast amounts of simulations [8].

Raw data output of these simulations are time series of “sensors”, for instance representing the response of the first bending mode in a direction normal to the blade’s chord-wise extension, labeled as “first flap mode”, or the bending moment in this direction as experienced at the root of a rotor blade. Sensor signals will exhibit broad-banded “noise” resulting from the structure’s response to wind turbulence. For sensors on the rotating system (rotor and drive train), these signals will be superimposed by pronounced periodicity at multiple, varying frequencies.

Due to the complexity and volume of this raw data, automated post-processing is mandatory to assist the engineer in their analysis. The standard case requires extraction of extrema on several levels of simulation bundles, or accumulation of so-called Damage Equivalent Loads, DELs. This meta data is then used either to size

components or to check margins to allowable values ensuring safe operation of the system.

In some cases, however, this type of meta data is insufficient, for instance as information on physical behavior leading to the identified extreme values is lost. A similar need occurs when trends with regard to the impact of input meta data must be identified. Resolving to raw output data is possible, but cumbersome and time consuming.

2.2 Condition Monitoring of Wind Turbines

A problem in the day-to-day operation of wind turbines is their downtime due to structural damage and other faults. In order to increase their output, one would want to minimize these outages. This goal can be achieved by implementing *condition monitoring* (CM) techniques. In contrast to conventional maintenance strategies (like run to breakdown), CM allows early detection and even prediction of damage. This is done by monitoring the current state of the turbines with the help of sensors [9].

With the use of CM it is possible to perform condition based maintenance, which has financial advantages to traditional strategies since, in particular, downtime and maintenance can be better planned. With the increasing number of installed wind turbines, the importance of CM will only grow in this field. Furthermore, with the higher efficiency of modern turbines, outages result in higher economical losses. Further advantages of CM are the prevention of secondary defects [11] caused by the failure of certain components, and total loss of the engine can be prevented by a timely shut-down.

Damage can occur in all components and control systems of the turbine. The annual frequency of these damage types varies and they have different impacts on the downtime of the turbine. While just 7% of all outages are caused by rotor blade failures, the average caused downtime of this damage type amounts to four days [5]. This motivates our investigation to improve CM methods in this field.

Vibration analysis is one of the main methods used in CM [14]. The idea behind this method is to analyze the change of vibrational properties which occur with the emergence of structural damage [1]. Therefore, we concentrate on the analysis of rotor blade oscillation data enriched with operational data of the wind turbine. Here, Fourier analysis is widely used to obtain the frequency spectrum of the measured vibrations. Often spectra of the undamaged case are taken as a point of reference for further measurements. Additionally, frequency bands which are associated with damage features can be easily isolated [1].

The analysis of the rotor blade oscillation data has to be based on robust behavioral models of the rotor blade. This means that these models ideally should be generally applicable for different wind turbine types, and should also be valid for a broad range of operational modes and meteorological conditions. Since in condition monitoring the models themselves are deduced from historical data, it is necessary to investigate the data from a vast amount of wind turbines over a broad range of operational

conditions in order to validate these behavioral models. Various kinds of analysis methods are suitable for this task, especially signal analysis and statistical methods for regression, clustering, and optimization. Since the approach is empirical, the results are fraught with statistical uncertainty.

In particular, to detect changes in vibration behavior, machine learning methods [6] are used. In real applications there is often not enough data available to train supervised learning methods, so unsupervised learning algorithms are employed. Especially anomaly detection, also called novelty detection, is a method from this domain. Here, one builds a model from the existing data of the undamaged turbine. Additional data points, i.e. measured during the normal operation, are checked for conformity with this model. When such current data deviates from the model, the occurrence of damage is assumed and further investigations are triggered.

3 Exploration of Time Series Data from Numerical Simulations

For the analysis of time series data obtained from numerical simulations, we here introduce an approach based on nonlinear dimensionality reduction. We propose an interactive exploration tool assisting the engineer in the detection of similarities or anomalies in large bundles of numerical simulations of wind turbines. The specific situation we consider is the so-called storm load case, in which the engineer searches for abnormal system behavior. In this scenario strong winds have caused the wind turbine to shut down and to assume a survival configuration in which blades are rotated to cause minimum drag when the hub points directly into the wind. Furthermore, after shutdown, connection with the power grid is lost so that no controls are operational. This means in particular that the machine cannot react to changes in wind direction anymore. Numerical simulations now provide time series data of the behavior of the wind turbine in such a "parked" position under storm conditions.

3.1 Virtual Sensor Data from Wind Turbine Simulations

As a result from the numerical simulation, we obtain a time series $\mathcal{T}$ of given length D, where D is the length of the simulation time in seconds divided by the time step size for saving the sensor data, which can but does need to be the same as the time step of the numerical simulation. The impact of wind direction is studied by using wind directions $\alpha, \alpha = 1, \ldots, M$, where M typically is 36, i.e. the wind directions are changed in steps of 10°. To address the turbulent nature of the wind, multiple (random) wind seeds are used, which we denote by $k, k = 1, \ldots, K$, where a typical value of K is 10. The average wind speed is given by regulations. For one wind turbine configuration we already have $M \times K$ simulations $\mathcal{T}_\alpha^k$, e.g. 360 for typical setups. In Fig. 2 we show some exemplary time series of a virtual sensor indicating the contribution of the so-called flapwise bending mode of one blade, which in this

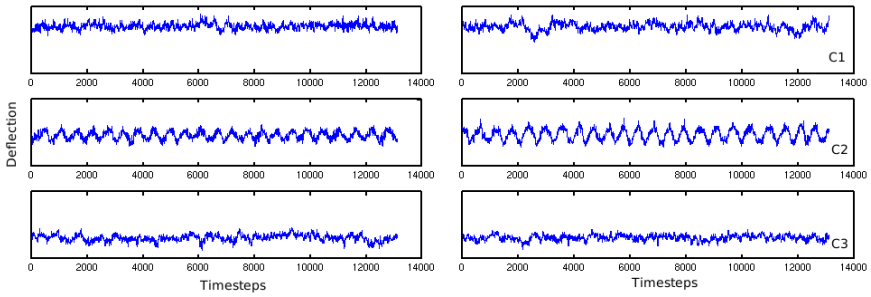


Fig. 2 Exemplary time series of a (virtual) flapwise mode sensor on a blade, for different wind directions. The first row shows time series data from two simulations from cluster C1, second row from cluster C2, and third row from cluster C3, refer Fig. 4 for details.

storm case represents bending within the plane of the rotor. The different quality of the response of the parked (standing still or idling) wind turbine as a function of the (now changed) wind direction is obvious to the human observer.

In order to quantify similarities between all bundles of simulation results, the first step in the analysis workflow is to extract the required data from all the simulations for the chosen sensor or sensors. If required, pre-processing steps are applied depending on the sensor/signal characteristics, e.g. scaling for data normalization, transformation to polar coordinates for blade position and deflection values, transformation to frequency domain for periodic signals, peak detection, etc. Generally, the pre-processing steps depend on the analysis goal and the specific sensor. Here, we are interested in an initial anomaly detection and proceed with the raw time series data.

3.2 Nonlinear Dimensionality Reduction for Time Series Analysis

We aim for approaches which are capable of comparing several time series in a semi-automatic fashion, and allow the detection of similarities or anomalies in the results. The basis for such a comparison is a suitable concept of distances between the high-dimensional data points, which we then aim to preserve in a so-called nonlinear dimensionality reduction, where points which are nearby in the high-dimensional space are still nearby in the lower dimensional space [10]. For visualization purposes one often embeds the simulation data into three dimensions, i.e. each numerical simulation is represented as a point in this space. Datasets can now be efficiently explored to find similarities and differences, and the embedded space can be navigated to investigate the properties of simulated time series which are close to each other, i.e. similar [10].

In applications one needs a suitable notion of distance between different time series to allow such a dimensionality reduction. The Euclidean distance is commonly the first choice for such a measure. However, it is not suited for all conceivable

applications, which may require completely different distance functions. For instance, the Euclidean distance will classify signals which are shifted in time, but are otherwise very similar in nature and magnitude, as dissimilar. But this is a situation typical for wind turbine simulation data; the time shift itself is due to variation in initial conditions and wind speed and seed, and is irrelevant when the dynamic characteristics of the wind turbine are investigated. One frequently used approach for signals which may vary in time or speed is called dynamic time warping (DTW) which consists of an optimal mapping between two signals implemented by a minimization of the pairwise distances under certain restrictions [7]. Lower distances indicate parts of the signal with high similarity. Evaluation has a computational complexity of $O(D^2)$, but under some assumptions it can be optimized by restricting the computation to only a suitable chosen part of the evaluation domain [7]. A variant of this algorithm is FastDTW, which uses a multilevel approach that recursively projects a solution from a coarser resolution and refines the projected solution, reducing the effort to $O(D)$ under suitable assumptions [15].

For applications in signal processing it is also necessary to be able to capture signal similarity at different time scales, specially in the presence of noise or locally influencing factors. Therefore, a way to decompose a signal into several components with varying degrees of detail is very useful. For example, one method that has been employed with good success for such a task is empirical mode decomposition [13]. The method does not use a set of fixed basis functions like in the case of Fourier or wavelet transforms, but it actually uses a spline interpolation between maximum and minimum in order to successively generate so called "intrinsic mode functions" (IMF). Once the IMFs are obtained, a nonlinear dimensionality reduction method can be applied based on these, albeit with the assumption that all signals generate compatible signal modes. Other such mathematically inspired pre-processing approaches can also be used before the eventual nonlinear dimensionality reduction.

The overall workflow of the studied data analysis procedure now consists of four steps.

1. In the first step of *data extraction* the specific data to be analyzed is extracted from the archived raw time series data. The selection of the data can depend on the wind turbine specifications, the load case, the wind directions, the chosen sensors, the analysis goal, etc.
2. During the *pre-processing* step the raw time series data is modified. For example transformations to the frequency domain can be applied, large variations in individual signals scales may be removed, or signals could be combined.
3. For the *dimensionality reduction*, computations are performed with the pre-processed data. For example, suitable distances are computed between all pairs of time series data, followed by a decomposition of the obtained similarity matrix.
4. The obtained low dimensional embedding is then used for an interactive *exploration* of the data. One can study the properties of nearby points in the embedding, which represent similar time series. The interactive study can be supported by computing the correlation, the mutual information, or similar quantities measuring relations between attributes, between the embedded dimensions and design parameters, or between quantities of interest in the output.

3.3 Diffusion Maps

In this work, we consider diffusion maps [2] as a nonlinear dimensionality reduction approach. Here, first a weighted graph $G = (\Omega, W)$ is constructed, where the vertices of the graph are the R observed data points $\mathcal{T}_\alpha^k \in \mathbb{R}^D$. The weight which is assigned to the edge between two data points $\mathbf{u} := \mathcal{T}_\alpha^k \in \Omega$ and $\mathbf{v} := \mathcal{T}_{\tilde{\alpha}}^{\tilde{k}} \in \Omega$ is given by $w(\mathbf{u}, \mathbf{v}) = \mathrm{e}^{-\Delta^2(\mathbf{u},\mathbf{v})/\epsilon}$, where the value of ϵ controls the neighborhood size. The term $\Delta(\mathbf{u}, \mathbf{v})$ can in general be an application-specific, locally defined distance measure that defines the way signals are to be compared. In our application of time series data, we could use the DTW distance measure if the time shift between signals is irrelevant, or the Euclidean distance if capturing time shift is important. The employed Gaussian kernel $e^{-\Delta^2}$ is a common choice in many applications, and we apply it in our experiments. This stems from the relationship between heat diffusion and random walk Markov chains, the basic observation is that if one takes a random walk on the data, walking to a nearby data-point is more likely than walking to another one that is far away.

With this construction, one defines a Markov process on the graph where the transition probability for going from $\mathbf{u}$ to $\mathbf{v}$ is given by $p(\mathbf{u}, \mathbf{v}) = w(\mathbf{u}, \mathbf{v}) / \sum_{z \in \Omega} w(\mathbf{u}, \mathbf{z})$. Hence, if the points are similar, then we have a high transition probability, and if they are not, then the transition probability is low. The Markov chain can be iterated for q time steps and a so called "diffusion distance" D_q can be defined in a natural way [2].

Diffusion maps now embed high-dimensional data into the low-dimensional space such that the diffusion distances D_q are (approx.) preserved. It has been shown in [2] that this is accomplished by the map $\Psi_q : v \to [\lambda_1^q \psi_1(v), \lambda_2^q \psi_2(v), ..., \lambda_s^q \psi_s(v)]$, where λ_j and ψ_j are the eigenvalues and right eigenvectors of the similarity matrix $P = p(\mathbf{u}, \mathbf{v})_{\mathbf{u},\mathbf{v}}$. Furthermore, one can approximate the distance between transition probabilities using the Euclidean distance in the reduced space. The number of terms used for the embedding, i.e. the dimension s of the low-dimensional space, depends on the weight w and the value of ϵ, besides the properties of the data set itself. In case of a spectral gap, i.e. when only the first few eigenvalues are significant and the remaining ones are small, one can obtain a good approximation of the diffusion distance with only a few terms. For details about the theoretical background of the approach we refer to [2].

3.4 Numerical Results

To evaluate the proposed data analysis procedure, we study a generic utility sized 1.5MW power wind turbine model with three blades of roughly 40 meters length. We consider numerical simulations of a storm load case with turbulent wind with average speed of 39 m/sec. All (external) controls are off due to loss of connection to the grid. The wind direction is varied in 10° steps from 0° to 360°, with 10 wind seeds per wind directions, resulting in 360 individual simulations. The time window

for the wind turbine simulation is roughly 11 minutes, where the sensor data is sampled at 0.05 second intervals for 1360 sensors. The dimension of each time series is equal to the number of data samples, i.e. it is 13107 in this case. We concentrate in the following on the analysis of sensor data for rotor speed, rotor position, blade deflection (flapwise and edgewise), tower deflection, and tower torsion only.

We first consider data from the rotor speed sensor with no pre-processing, and use the Euclidean distance. The dimensionality is reduced to three for visual purposes, while capturing the main behavior of the data. Figure 3 (left) shows the low dimensional representation. There are 360 points in the displayed embedding plot, each point in three dimensions represents one simulation. Simulations in the same cluster in such an embedding have similar behavior in regard to the rotor speed sensor.

Since each simulation corresponds to a particular wind direction and wind seed, we can also visualize the results with a polar plot (see Fig. 3, right) by using the first three coordinates from the low-dimensional embedding as RGB values to derive a color value for each simulation. Each cell in the plot represents the time series from one simulation, while the position of each cell is based on the wind direction (azimuthal) and the wind seed (radial). For example, the 36 simulations with seed A form the innermost circle, where the wind direction in degrees is marked on the boundary of the outer most cell. Simulations with similar behavior are close to each other in the low dimensional coordinates, and are therefore shown in similar color in the polar plot. For example, one easily observes that the wind directions marked by C1, C2, and C3 in the right plot form clusters in the adjoining embedding plot on the left of Fig. 3. Further investigation of the raw data confirms that the rotor is rotating while in the idle configuration for the simulations in these three clusters, but for the other wind directions it is not. Additionally, the rotor speed in cluster C1 is higher than in C2 and C3, and the rotor speed in simulations from the C2 cluster is again higher than in C3. This is also apparent from the location of these clusters along one dimension of the low-dimensional space, which appears to be dominated by the rotor's rotational speed sensor. No information can be obtained on the quality of rotor rotation, however.

As a second example we consider the flapmode sensor for blade one, again no pre-processing was performed for the sensor. The embedding plot in Fig. 4 also shows three different clusters (note that C1-C3 are different from the cluster shown previously in Fig. 3). Further investigations of the raw data time series confirm that the rotor is in a still position for the simulations in cluster C1 and C3, and the position of blade 1 for simulations in cluster C1 is different from that in cluster C3. Simulations in cluster C2 have a rotating rotor. In simulations with wind direction of -40° the rotor did rotate for half of the simulation time, and is inconsistent with the behavior with simulations having adjacent wind directions of -30° and -50°. This anomaly is very relevant from an engineering perspective and can be easily identified with our data analysis procedure.

Using DTW for this data provides slightly more details in cluster C2. We considered only the range of wind directions from this cluster, where the rotor is actually turning. We performed an embedding using both the Euclidean distance and the distance obtained from DTW, as seen in Fig. 5. While the Euclidean distance incorporates

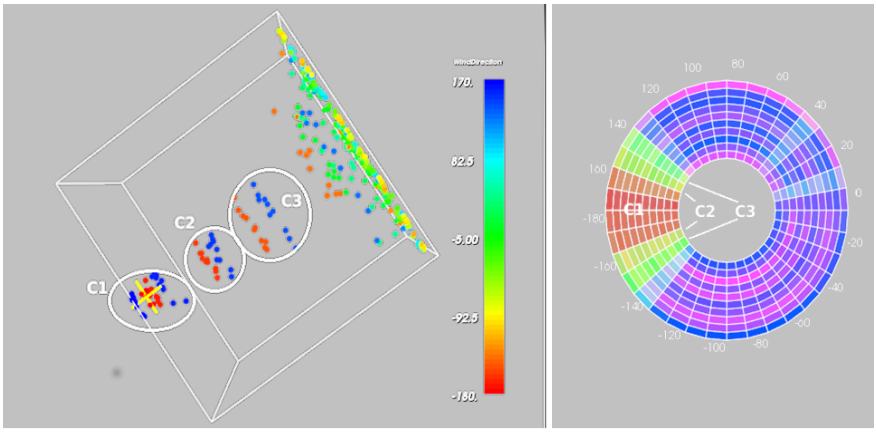


Fig. 3 Embedding results for rotor speed sensor data, using diffusion maps with the Euclidean distance. For simulations in clusters C3 to C1, wind turbine is observed to be rotating with increasing speed, i.e. simulations in cluster C1 have highest rotor speed.

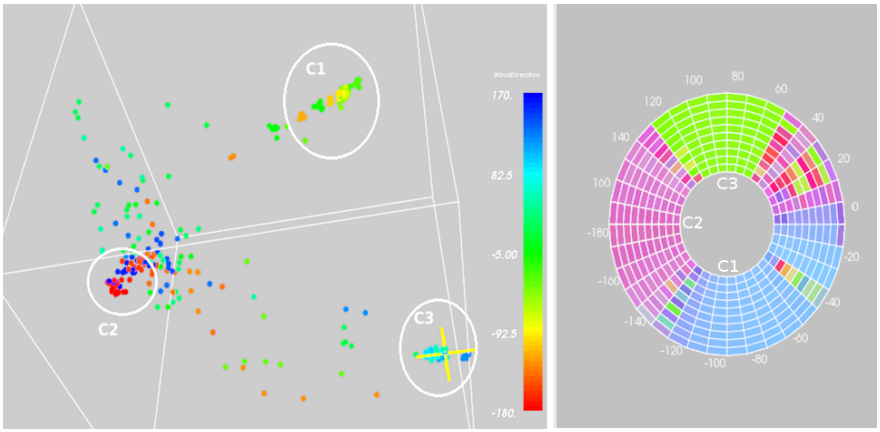


Fig. 4 Embedding results for flapwise mode sensor data from blade 1, using diffusion maps with the Euclidean distance. Similar behavior is observed among simulations in each cluster and an anomaly is detected for simulations with wind direction -40°.

differences from different wind seeds, initial conditions, and thus time shifts in signals, DTW effectively removes this dependency permitting a clear distinction for the different wind directions.

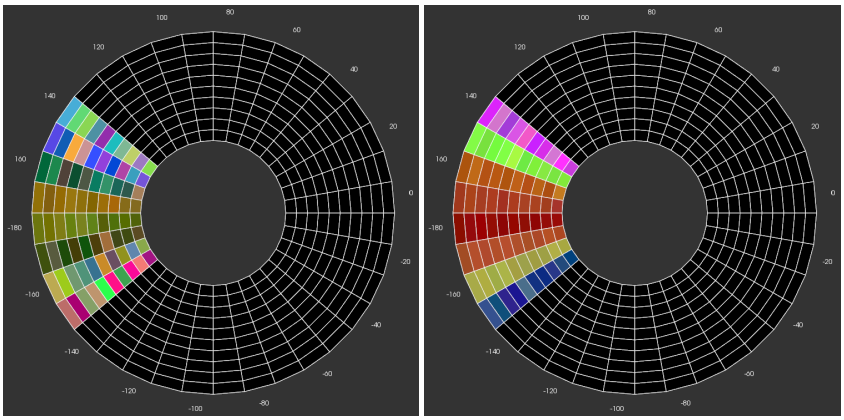
These examples show the possibilities of the nonlinear dimensionality reduction to allow a quick identification of similarly behaving simulations and the detection of anomalies in certain simulations.

4 Anomaly Detection Based on Linear Dimensionality Reduction for Condition Monitoring Sensor Data from Wind Turbines

For the detection of anomalies in the sensor data of wind turbines we now propose an approach based on dimensionality reduction. The method utilizes principal component analysis (PCA) on vibrational sensor measurements of a wind turbine, given in the form of frequency spectra (computed from the time series data), to obtain a baseline by way of the low-dimensional data representation. Further measurement samples are projected into this basis to detect deviation of the coefficients from those of the baseline. Our approach is evaluated with real life data which is provided by Weidmüller Monitoring Systems (WMS). Note that initial investigations of nonlinear dimensionality reduction approaches for this real life data did not show advantages for anomaly detection against the employed linear approach.

4.1 Sensor Data from Rotor Blades

The vibrational data from the rotor blades are collected by two acceleration sensors per blade, mounted at different angles. The *flap*-sensors align with the flat side of the blade, and the *edge*-sensors are mounted towards the edge of the blade, see Fig. 6. Instead as raw time series data, the information is stored in the form of frequency spectra and also contains environmental and operational parameters. The different orientation of the edge and flap sensors cause differences in the kind of vibrations



(a) Embedding with Euclidean distance (b) Embedding with DTW

Fig. 5 Embedding results for flapwise mode sensor data from blade 1, using different distance measures for Δ in diffusion maps. In difference to the Euclidean distance, DTW allows a clear distinction for the different wind directions.

they are monitoring. In particular, flap sensors undergo stronger vibrations because they monitor vibrations in the wind direction.

The data points stem from hourly measurements. In each measurement cycle the vibration from the sensors is captured over the course of 100 seconds and afterwards transformed into the frequency domain. Additionally, the following operational parameters are saved: Time stamp of measurement, rotational speed, power output of the turbine, wind velocity, pitch angle of the rotor blades, and environmental temperature, where the instationary values are averaged over the time window.

A typical frequency spectrum is shown in Fig. 7, it can be interpreted as a vector of length D containing real numbers. To visualize different spectra at the same time, we arrange them into a spectrogram as seen in Fig. 8. This allows an overview of all spectra from one sensor, and events like shifts of peaks can be easily detected. For example, Fig. 8 shows clearly some differences between certain measurement points, i.e. the blue lines are from measurements during the downtime of the turbine, therefore having lower amplitudes than the other spectra.

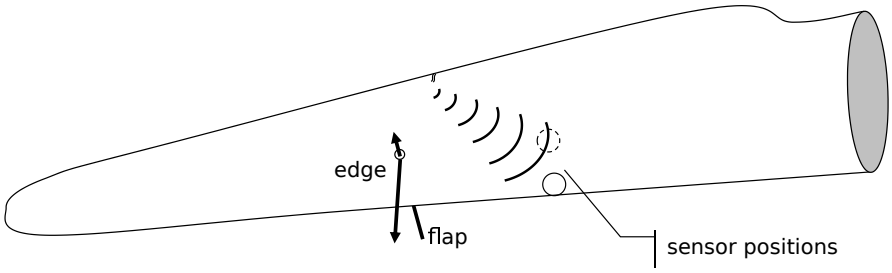


Fig. 6 Blade with edge/flap sensor positions.

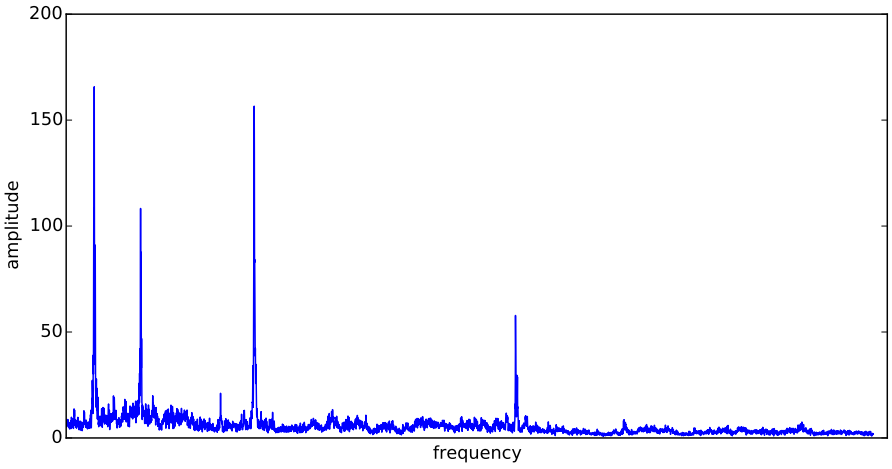


Fig. 7 A typical frequency spectrum, taken from a single sensor during one measurement cycle.

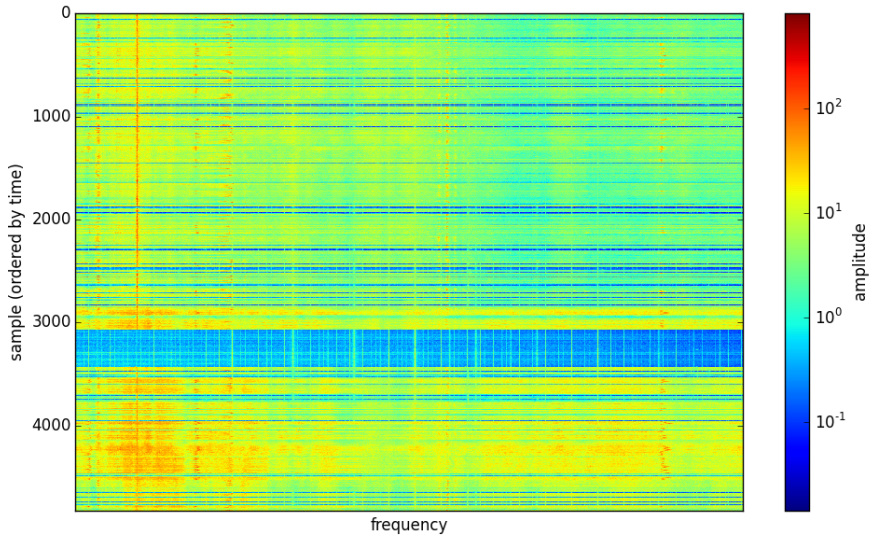


Fig. 8 Spectrogram of frequency data of a single sensor, spectra are sorted by time parameter from top to bottom.

4.1.1 Pre-processing

Before one analyzes the data, one needs to perform some pre-processing steps. For example, the data points are divided into different classes according to their operational parameters. These classes represent different operational modes, e.g. grouped by rotational speed or similar wind velocities. In the next step the variables to be examined are selected; these are certain frequency bands which are associated with damage features.

Experiments have shown that rotational speed is one of the main contributors to the change in the observed frequency spectra. Especially since the data sets originate from turbines which are essentially operated with constant rotational speeds (see Fig. 9), it is advised to separate data classes by rotational speed.

Figure 10 shows the spectrogram of measurements taken under the same rotational speed. Compared to Fig. 8, it is clearly seen that these frequency spectra are more similar to each other.

Instead of processing the frequency spectra as a whole, smaller frequency bands can be selected for analysis. The underlying idea behind analyzing smaller portions of the spectrum separately is that each band can contribute to the damage signal on its own. This is favorable when damage manifests locally in some frequency bands which are dominated in variance by other non-relevant frequencies. On the downside, this approach can result in a higher false positive rate. Furthermore, one might know from engineering principles in which parts of the spectrum certain damage features manifest themselves, and therefore one can select these bands for detection of these

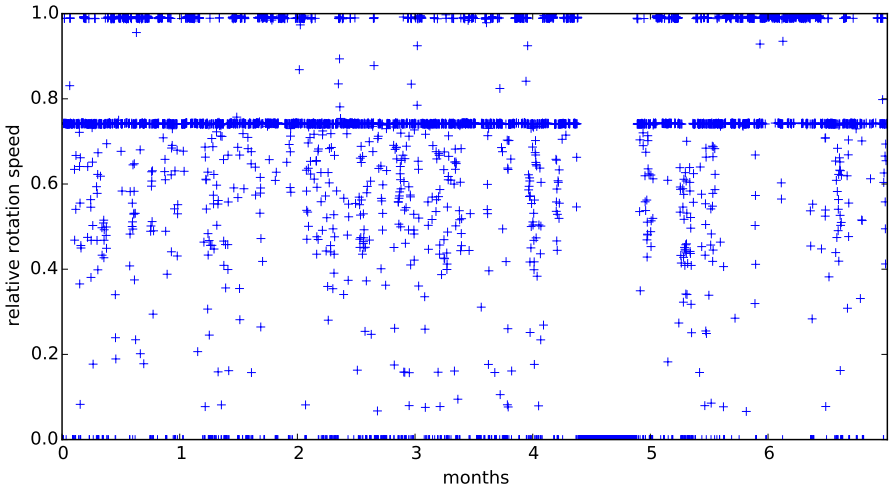


Fig. 9 Rotational speed over time of one turbine. The two horizontal lines on which most of the points accumulate represent the two operational speeds at which the turbine is operated.

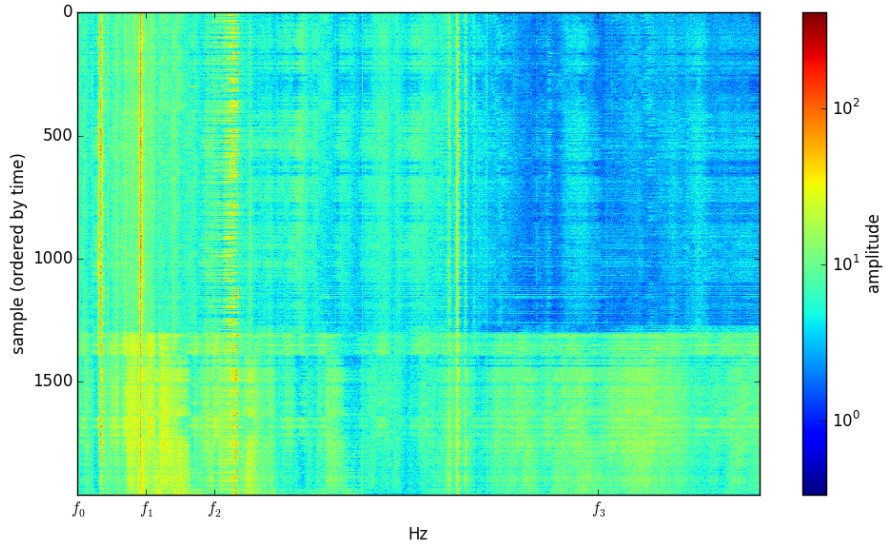


Fig. 10 Spectrogram of samples taken under same rotational speed. Frequencies f_i are marked to illustrate ways to select variables described in section 4.1.1.

specific features. Additionally, it saves computational time when several narrow frequency bands are analyzed since fewer variables are needed.

In the following, we assume that $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\} \subset \mathbb{R}^D$ is the set of m samples of dimension D which remain after the described pre-processing steps are taken, in particular this means that the vectors $\mathbf{y}_i$ do not contain the whole spectrum, but

the selected frequency band. In our case the vectors $\mathbf{y}_i \in \mathbb{R}^D$ can be interpreted as discrete frequency spectra as seen in Fig. 7.

4.2 Anomaly Detection in Sensor Data

4.2.1 Model of Undamaged State

One common approach for anomaly detection is to create a baseline model of the undamaged condition of the wind turbine. To achieve this, we select a point in time t until which the turbine is assumed to be free of damage. Let

$$\mathcal{Y}_t := \{\mathbf{y}_1, \ldots, \mathbf{y}_t\} \subset \mathcal{Y}$$

be the set of data points gathered until this point in time.

From $\mathcal{Y}_t$, we calculate the baseline model using principal component analysis (PCA), which is a widely used method for linear dimensionality reduction. The core principle of PCA is to find a linear projection of the original data set into a new orthogonal basis by computing an eigenvalue decomposition. In this basis the individual components are statistically independent, allowing a distinction of relevant and redundant information in the data set. By discarding all but the first few principal components corresponding to latent variables with the highest variance, a low dimensional representation of the data is found which keeps most of the relevant information. Note that the use of PCA implies a further pre-processing step, namely the centering and scaling of the data $\mathcal{Y}_t$.

We collect the data samples $\mathcal{Y}_t$ into a matrix $\mathbf{Y}_t := [\mathbf{y}_1, \ldots, \mathbf{y}_t] \in \mathbb{R}^{D\times t}$. Let $\mathbf{C} := \frac{1}{t}\mathbf{Y}_t\mathbf{Y}_t^T \in \mathbb{R}^{D\times D}$ be the sample covariance matrix and $\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ its eigenvalue decomposition, where $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_D)$ consists of eigenvalues ordered by magnitude and $\mathbf{V}$ contains the according eigenvectors. The projection matrix $\mathbf{P} \in \mathbb{R}^{S\times D}$ is given by

$$\mathbf{P} := \mathbf{I}_{S\times D}\mathbf{V}^T. \tag{1}$$

The samples $\mathbf{y}_i \in \mathbb{R}^D$ are mapped to the first S principal components by $\mathbf{P}\mathbf{y}_i = \mathbf{x}_i \in \mathbb{R}^S$, with $S < D$. Deviations of the remaining points $\mathcal{Y} \setminus \mathcal{Y}_t$ from this baseline are measured in a suitable fashion. Note that a similar approach is described in [12].

4.2.2 Deviation from the Undamaged State

The PCA can be used to detect anomalous behavior in a process or system by measuring the variation of the coefficients of the samples in the PCA model. For this purpose the Q-statistic and T^2-statistic are frequently used tools. In order to apply them, we have to assume that the underlying process is normally distributed. The Hotelling's T^2-statistic is a generalization of the Student's t-statistic and is based on the scores of the projection of a sample.

As stated above, $\mathbf{P}$ is the transformation matrix from the PCA-model of $\mathcal{Y}_t$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_S)$ is the covariance matrix of the latent variables, obtained by the PCA algorithm. The T^2-statistic of sample $\mathbf{y}_i$ is calculated by

$$T_i{}^2 := \mathbf{P}\mathbf{y}_i\mathbf{\Lambda}^{-1}\mathbf{y}_i^T\mathbf{P}^T = \sum_{j=1}^{S} \frac{x_{ij}^2}{\lambda_j}\,. \tag{2}$$

The T^2-statistic detects variation in the subspace of the first S (not discarded) principal components, measuring the variation of samples within the PCA model. We then define a tolerance threshold m_T by:

$$m_T := \max_{\mathbf{y}\in\mathcal{Y}_t} |\mathbf{P}\mathbf{y}\mathbf{\Lambda}^{-1}\mathbf{y}^T\mathbf{P}^T|. \tag{3}$$

If the values from the T^2-statistic of the remaining data points exceed m_T, the presence of an anomaly, potentially indicating a damage or fault, is assumed.

An alternative would be the Q-statistic, which is a measure for the (residual) difference between a sample and its projection onto the model. Note that in our experiments we observed essentially the same results for the T^2-statistic and the Q-statistic.

4.3 Methodology for Damage Detection

The pre-processing steps and the calculation of the baseline parameters are now collected into a damage detection method. This method analyzes data $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\} \subset \mathbb{R}^D$ from a single sensor of the turbine and detects deviations from the undamaged state and consists of the following steps:

1. pre-processing
   - divide data points into classes determined by operational modes
   - select frequency ranges to be analyzed
2. select data points $\mathcal{Y}_t$ describing the undamaged state of the turbine
3. compute baseline model from $\mathcal{Y}_t$ using PCA
4. set thresholds for T^2- or Q-statistics
5. use the statistics to measure deviations from the baseline model

Let $\mathbf{Y} \subset \mathcal{Y}$ be the class of data points which satisfy the parameters of a selected operational mode, e.g. with the same rotational speed. Furthermore let $j, k \in \{1, 2, \ldots, D\}$ be the lower and upper bound of a selected frequency band. Let t be the point in time until which the installation is assumed to be without damage. Algorithm 1 returns the baseline computed from the frequency spectra describing the undamaged state. To cover the biggest possible portion of the initial data $\mathcal{Y}$, several classes of operational modes have to be chosen accordingly. Additionally, when l many frequency bands are selected for analysis, we iterate the above algorithm over these selections yielding l different baseline models.

Algorithm 1: Calculate Baseline Model for a Sensor

Data: $\mathbf{Y} \subset \mathcal{Y}$ frequency spectra describing undamaged state, j, k frequency range
Result: baseline model $\mathbf{P}$, $\mathbf{\Lambda}$, m_T
$n := |\mathbf{Y}|$, $i := 1$
for $\mathbf{y} = (y_1, \ldots, y_D) \in \mathbf{Y}$ **do**
  $\mathbf{Y}_t[i] \leftarrow (y_j, \ldots, y_k)^T$ ▷ selection of frequency variables
  $i \leftarrow i + 1$
calculate matrices $\mathbf{P}$, $\mathbf{\Lambda}$ for baseline model after (1) using $\mathbf{Y_t}$
calculate threshold m_T from $\mathbf{Y_t}$ according to (3)
return $\mathbf{P}$, $\mathbf{\Lambda}$, m_T

Algorithm 2: Calculate Damage Signal

Data: $\mathbf{Y} \subset \mathcal{Y}$ frequency spectra, j, k frequency range, baseline model $\mathbf{P}$, $\mathbf{\Lambda}$, m_T
Result: damage signal $\mathbf{s} \in \mathbb{R}^n$
$n := |\mathbf{Y}|$, $i := 1$
for $\mathbf{y} = (y_1, \ldots, y_D) \in \mathbf{Y}$ **do**
  $\mathbf{y} \leftarrow (y_j, \ldots, y_k)$ ▷ selection of frequency variables
  compute T_i^2 after (2) ▷ compute T^2-statistic for each $\mathbf{y}$
  $\mathbf{s}[i] \leftarrow \begin{cases} 0 & \text{if } T_i^2 \le m_T \\ T_i^2/m_T & \text{else} \end{cases}$ ▷ compute damage signal for each $\mathbf{y}$
  $i \leftarrow i + 1$
return $\mathbf{s}$

For a given collection of frequency spectra and a baseline model, Algorithm 2 returns the corresponding damage signal $\mathbf{s}$. This output can be interpreted as a one dimensional time series. When l many frequency bands are selected, we obtain i different damage signal time series in the form of a vector $[\mathbf{s}_1, \ldots, \mathbf{s}_l]$. To obtain one signal value, we sum them up entrywise in our analysis $\mathbf{s}[i] := \sum_{k=1}^{l} \mathbf{s}_k[i]$. Alternatively one can take the maximum over the signal entries $\mathbf{s}_k[i]$.

The data points occurring after time stamp t can be interpreted as a continuous data feed in real time. Therefore, after the computation of a baseline model with Algorithm 1, this data feed can be directly evaluated by Algorithm 2, and a continuous signal feed is generated in real time.

4.4 Numerical Results

In this section we present the application of the proposed method on real wind turbine data as described in Sect. 4.1. Each data point consists of a frequency spectrum from each sensor and operational parameters taken during the measurement cycle. The data set contains around 4000 data points taken over the course of seven months. Data from the first month was used to compute the baseline model. The occurrence of damage in

the given time frame is known, and its date was determined by a proprietary method of WMS.

During pre-processing, the selection of data points by environmental parameters takes place. Because the turbine at hand is operated under constant rotational speed, the data points originating from the first rotational mode are selected for analysis. This leaves us with around 2000 data points which cover the time domain of the data quite well. Additionally, we perform selection of frequency bands. Here, we distinguish between two different approaches:

- (**F1**) analysis of the whole frequency band between the frequencies f_2 and f_3 in Fig. 10
- (**F2**) the whole spectrum is divided into several smaller frequency bands of the same width, i.e. intervals of length $f_1 - f_0$ in Fig. 10, which are analyzed separately, and their results are summarized afterwards

Based on the decay of the eigenvalues for this data, we keep the first three principal components for the computation of the baseline model. The data points are centered and scaled by the mean and standard deviation of the baseline data points.

The following figures illustrate the results of the method, which was applied to the data from the three *flap*-sensors. The three plots in one figure show the values of the damage signal (computed by the described algorithm) over time. The damage signal is an indicator for deviation from the baseline model, and big values can be associated with the occurrence of damage. The left vertical line indicates the point in time for which the baseline was selected. The dashed line represents the date t_d at which the damage is first detected by the proprietary method of WMS evaluated for the stored sensor data. Note that the method of WMS was not available for live monitoring of the wind turbine at that time.

In Fig. 11 we show results for which frequency band F1 was evaluated. The first rotor blade does not show significant deviations, the second blade on the other hand shows deviations before the assumed date of damage occurrence t_d. If this has to be attributed to anomalies which were not included into the baseline model or if it is an indicator for damage is unknown. One can see clearly that the damage signal from rotor blade 3 increases after date t_d, so the purely data-driven procedure obtains results very similar to the original approach.

After the fault, the turbine was shut down during month 4, visible by the gap between month 4 and 5. After re-activation of the turbine in month 5, the deviations persist and even increase in blade 1 and 2, which indicates a change of operational or environmental conditions due to the repair of the rotor blades and other maintenance efforts. For the purpose of condition monitoring, a new baseline would need to be derived in such a situation.

Figure 12 shows the sum of damage signals from frequency bands **F2**. The results are essentially the same, showing that the anomaly detection procedure is, at least somewhat, independent from the frequency bands which are chosen for the evaluation.

We also applied our approach on sensor data from a second wind turbine with a similar fault and observed results consistent to those presented here, i.e. the fault can be detected at more or less the same time as the proprietary method.

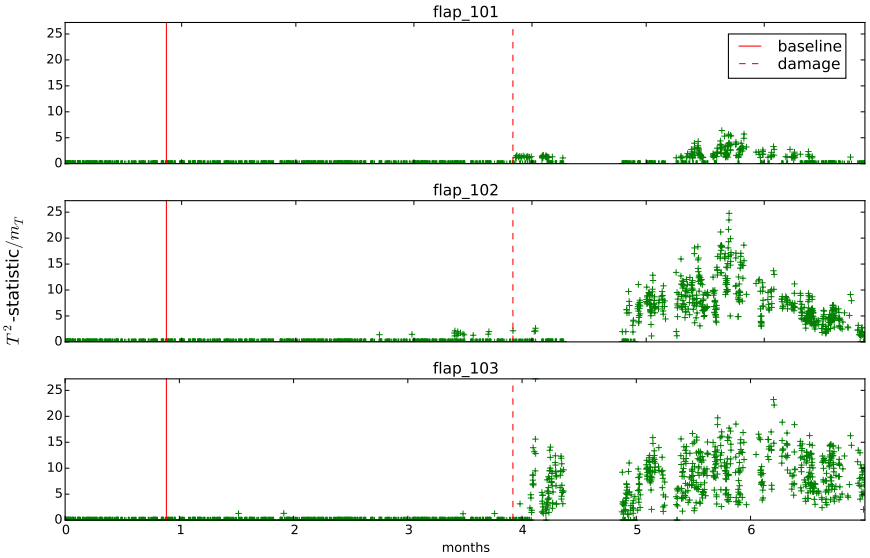


Fig. 11 Damage signal from evaluation of frequency band **F1**. The signal from rotor blade 3 increases after date t_d, the assumed date of the damage occurrence.

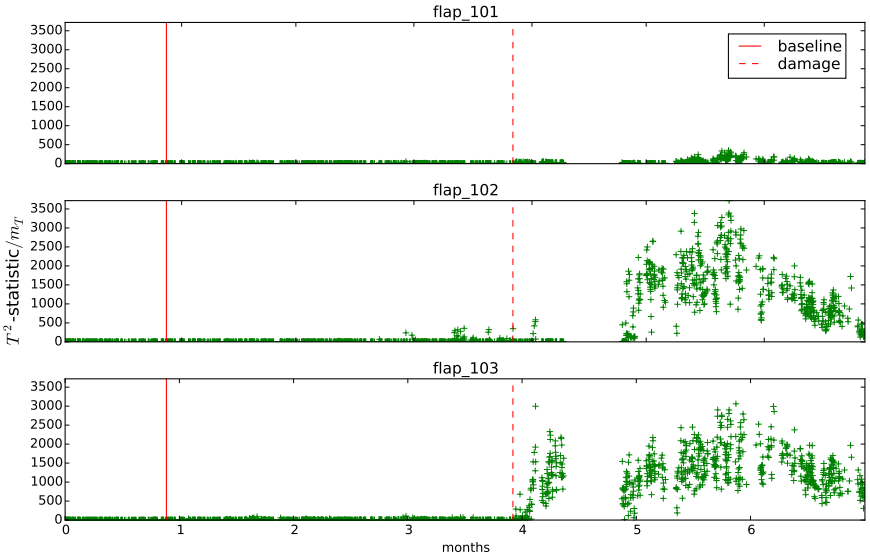


Fig. 12 Sum of damage signals from evaluation of frequency bands **F2**. The signal from rotor blade 3 increases after date t_d, the assumed date of the damage occurrence.

5 Conclusions

We described how linear and nonlinear dimensionality reduction approaches can be used for analyzing time series data in the wind energy domain. For the analysis of time series data stemming from numerical simulations in the virtual product development process, we introduced an approach which allows the quick identification of similarly behaving simulations and the detection of anomalies in simulation results. Furthermore, we proposed an approach based on linear dimensionality reduction which is able to detect changes in vibrational properties of rotor blades and which can be associated with the occurrence of faults.

A natural extension of this work would be to evaluate the data of more than one sensor at a time. For analyzing multiple sensors concurrently, suitable concepts of distances have to be introduced in this setting, and one has also to investigate, e.g., correlations in time between different sensors. Generally, other distance measures suitable for time series have to be investigated, in particular those which can be interpreted to observe certain physical principles and are invariant to certain transformations.

Acknowledgements The authors would like to thank the German Federal Ministry of Education and Research (BMBF) for the opportunity to do research in the VAVID project under grant 01IS14005. We cordially thank Henning Lang and Tobias Tesch for their assistance with the numerical experiments.

References

1. I. Antoniadou, *Accounting for nonstationarity in the condition monitoring of wind turbine gearboxes*, PhD thesis, University of Sheffield, 2013.
2. R. Coifman and F. Lafon, *Diffusion maps*, Applied and computational harmonic analysis, 21 (2006), pp. 5–30.
3. E. W. E. A. (EWEA), *Wind in power - 2015 european statistics*, tech. rep., European Wind Energy Association, February 2016.
4. R. Gasch and J. Twele, eds., *Wind Power Plants*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
5. B. Hahn, M. Durstewitz, and K. Rohrig, *Reliability of wind turbines*, in Wind Energy, J. Peinke, P. Schaumann, and S. Barth, eds., Springer Berlin Heidelberg, 2007, pp. 329–332.
6. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Second Edition*, Springer, 2009.
7. F. Itakura, *Minimum prediction residual principle applied to speech recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 23 (1975), pp. 67 – 72.
8. J. M. Jonkman, *The new modularization framework for the fast wind turbine cae tool*, in Proceedings of the 51st AIAA Aerospace Sciences Meeting, 2013. also Tech. RepNREL/CP-5000-57228, National Renewable Energy Laboratory, Golden, CO.
9. A. Kusiak, Z. Zhang, and A. Verma, *Prediction, operations, and condition monitoring in wind energy*, Energy, 60 (2013), pp. 1–12.
10. J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction*, Springer, 2007.
11. W. Lu and F. Chu, *Condition monitoring and fault diagnostics of wind turbines*, in Prognostics and Health Management Conference, 2010. PHM '10., Jan. 2010, pp. 1–11.

12. L. Mujica, J. Rodellar, A. Fernández, and A. Güemes, *Q-statistic and T^2-statistic PCA-based measures for damage assessment in structures*, Structural Health Monitoring, 10 (2011), pp. 539–553.
13. N. E. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, and H. Liu, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454 (1998), pp. 903–995.
14. R. B. Randall, *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*, Wiley, 2011.
15. S. Salvador and P. Chan, *FastDTW : Toward Accurate Dynamic Time Warping in Linear Time and Space*, Intelligent Data Analysis, 11 (2007), pp. 561–580.