Data Driven Modeling for System-Level Condition Monitoring on Wind Power Plants

Jens Eickmeyer[1], **Peng Li**[2], **Omid Givehchi**[2], **Florian Pethig**[1] and **Oliver Niggemann**[1,2]
[1]Fraunhofer Application Center Industrial Automation IOSB-INA
e-mail: {jens.eickmeyer, florian.pethig, oliver.niggemann}@iosb-ina.fraunhofer.de
[2] inIT - Institute Industrial IT
e-mail: {peng.li, omid.givehchi, oliver.niggemann}@hs-owl.de

Abstract

The wind energy sector grew continuously in the last 17 years, which illustrates the potential of wind energy as an alternative to fossil fuel. In parallel to physical architecture evolution, the scheduling of maintenance optimizes the yield of wind power plants. This paper presents an innovative approach to condition monitoring of wind power plants, that provides a system-level anomaly detection for preventive maintenance. At first a data-driven modeling algorithm is presented which utilizes generic machine learning methods. This approach allows to automatically model a system in order to monitor the behaviors of a wind power plant. Additionally, this automatically learned model is used as a basis for the second algorithm presented in this work, which detects anomalous system behavior and can alarm its operator. Both presented algorithms are used in an overall solution that neither rely on specialized wind power plant architectures nor requires specific types of sensors. To evaluate the developed algorithms, two well-known clustering methods are used as a reference.

1 Introduction

According to a wind market statistic by the GWEC (Global Wind Energy Council) [1], the global wind power capacity grew continuously for the last 17 years. In 2014, the global wind industry had a 44 % rise of annual installations and the worldwide total installed capacity accumulated to 369553 megawatt at the end of 2014. In Europe, renewable energy from wind power plants (WPP) covers up to 11% of the energy demand [2]. With this rapid continuous growth, the wind power is considered as one of the most competitive alternative to fossil fuels.

In a case study, Nilsson [3] denotes an unscheduled downtime with 1000 € per man-hour, with costs of up to 300000 € for replacements. This does not take into account the reduced yield through production loss. Therefore, the objective of maintenance is to reduce WPPs downtimes and provide high availability and reliability.

High availability is currently achieved by two different strategies. On the one hand, maintenance is planned as regular time-interval based on the manufacturer's data of specific WPP parts. This is performed in order to prevent wearout failure. On the other hand, there is the strategy of corrective maintenance, which reacts to occurred failures. Both strategies need time for actual maintenance, which lead to non productive downtimes. Especially, when considering offshore WPP, these downtimes produce high costs.

To reduce these downtimes a precise proactive scheduling of maintenance task is needed. This is achieved through condition monitoring (CM) systems [4]. Those systems try to reason about the inherent system states such as wear, although these conditions cannot be measured directly, but the growing amount of sensors in modern WPP enable an adequate description of the machines state. To make use of this, CM systems need a model of the WPP, which describes the system behavior based on observed data.

Existing CM solutions for WPP rely on specific sensors and are specialized to monitor single parts of the system. The gearbox [5], the bearing [6], the generator [7] or the blades [8] have been monitored in order to perform proactive maintenance. Here, specific sensors are needed as a requirement for these specialized methods.

This article presents a system-level solution which handles heterogeneous WPP architecture regardless of installed sensor types. Also, an algorithm for modeling a WPP on system level and another algorithm for anomaly detection are stated. To achieve this, three challenges are tackled and their solutions are presented:

I. *Logging data* from available sensors of a WPP, using existing infrastructure independent of the architecture. Additionally, the opportunity must be given to add new sensors and sensor types on demand.

II. *Automatic modeling* of a WPP, by combining existing and generic data-driven methods. Such a model must be able to learn the complex sensor interdependencies without extra manual effort.

III. *Anomaly detection* for a WPP regardless of its kind of architectures, especially with no assumptions on available types of sensors.

The article is structured as follows. Section 2 deals with state of the art technology in WPP CM. Hardware and data acquisition for the presented solution are specified in section 3, here point I is the central issue. Data-driven models realizing point II and the analyzed machine learning approaches are the purpose of section 4. Anomaly detection and its general approach, according point III is stated in section 4.2. The results of an evaluation of the presented methods is content of section 5. Finally, this paper concludes in section 6 and describes future aims of the presented work.

2 Related Work

The core task of a CM system is anomaly detection. As stated in [9], the models used for anomaly detection of complex systems should be learned automatically and data-driven approaches to learning such models should be moved into the research focus.

A wide range of data-driven algorithms that deal with modeling the system behavior for anomaly detection are available in the literature.

Because of its simplicity in processing huge amounts of data, the Principal Component Analysis (PCA) based algorithms are widely applied in the condition monitoring of WPP [10][11].

As one of the classic density based clustering method, DBSCAN shows its advantages over the statistical method on anomaly detection in temperature data [12].

Piero and Enrico proposed a spectral clustering based method for fault diagnosis where fuzzy logic is used to measure the similarity and the fuzzy C-Means is used for clustering the data [13].

Due to the high complexity of a WPP and its harsh working environment, the modeling of WPPs on system level is very challenging. Most data-driven solutions to WPP condition monitoring concentrate on the errors of one particular component (in component level) [4]. These methods are designed to detect specific faults (e.g. fault in gearbox, generator).

The application of such methods is available in different studies. In [6], a shock pulse method is adapted for bearing monitoring. A multi-agent system is developed in [5] for condition monitoring of the wind turbine gearbox and oil temperature. In [8], the ultrasonic and radiographic techniques are used for non-destructive testing of the WPP blades. Using these methods can prevent the WPP breakdowns caused by the particular faults. For enhancing the availability and the reliability of the whole WPP, a method for monitoring the WPP on system-level is desired.

In this work, a PCA-based algorithm for condition monitoring of WPP is presented. This approach is aimed to model a WPP on system-level in order to perform automatic anomaly detection. As a comparison, DBSCAN and spectral clustering are utilized for the same purpose. To the best of our knowledge, no application of either DBSCAN or spectral clustering in condition monitoring of WPP exists.

3 Data Acquisition Solution

A WPP includes different types of sensors, actuators and controllers installed to monitor and control the different devices and components as shown in Figure 1. To monitor the condition of a WPP, it is necessary to collect process data from its sensors and components accurately and continuously feed this data to the diagnosis algorithms. To maximize accuracy, data should be acquired directly from the sensors and components or via the existing communication systems. Despite the fact that IEC 61400-25 [14] addresses a variety of standards and protocols in WPP, lots of proprietary solutions exist today. A general approach to accurate data acquisition in an uniform way implies protocol adapters or data loggers (DL) to connect the diagnosis framework. This is done not only for IEC 61400-25 conformant WPP, but also for proprietary ones using e.g. the MODBUS protocol or a direct connection via general-purpose input/output (GPIO) [15]. Also the data logger should model data based

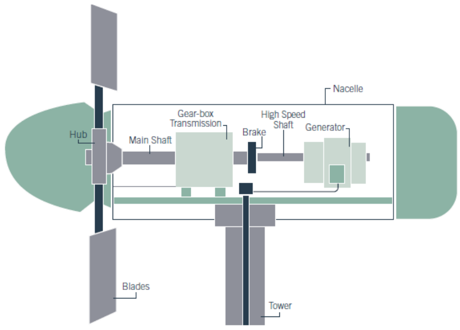

Figure 1: Diagram showing the inside of a nacelle and main components [4]

on generic industrial standards (IEC 61400-25) and transfer them to a database for storage and processing. Such a data logger meets point I (see section 1). In addition, the timestamp of the data should be synchronized between data loggers, database and application accurately.

In this work we followed a three layer architecture for data acquisition as shown in Figure 2 which covers all of the CM system components. In layer 1, the physical machine components are connected to the data logger hardware using different industrial connections and protocols e.g. digital GPIO, RS485, MODBUS, etc. The data loggers are time synchronized using global positioning system (GPS) or network time protocol (NTP) time references via an embedded time client running in the data logger. Collected sensor data is attached to their accurate timestamps by an embedded OPC UA server inside the data logger. The sensor data is categorized based on an OPC unified architecture (OPC UA) data model (e.g. conformant to IEC 61400-25) for a standalone WPP.

The communication between data logger, OPC UA server and layer 2 is realized with a secure general packet radio service network (GPRS) or a virtual private network (VPN), while it can be accessed for widely distributed WPPs in different geographical locations. The layer 2 comprises a middleware to collect and host the sensor data coming from distributed data loggers. It mainly covers a database with support of historical data and also an OPC UA server aggregating the data incoming from distributed WPP data loggers and pushes them to the database using an OPC UA database wrapper. As shown in Figure 2, the main component of layer 3 is an analysis engine. This engine applies algorithms on the database. Based on the learned machine models an output about the machines condition is presented to the operator by a human machine interface (HMI).

4 Modeling Solution

The main idea of the presented solution is to automatically learn a model of normal system behavior from the observed data using data-driven methods. Classical manual modeling utilizes expert process knowledge to build a simulation model as a reference for anomaly detection. But a process such as a WPP contains numerous continuous sensor values, which make it difficult to model the system manually. Therefore, as first step of the solution a model is learned from a set of data. The second step utilizes this model as reference to perform anomaly detection. This section considers these two steps.

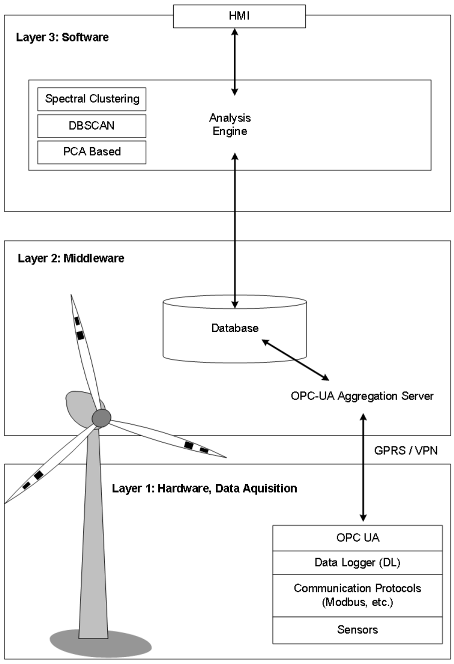

Figure 2: Architecture overview of the presented system-level condition monitoring solution for a WPP

4.1 Step 1: Data-Driven Modeling

In order to automatically compute a system model, the presented solution use generic methods to analyze training data and aim for process knowledge. These methods from the field of machine learning reduce effort of time for generating a system model caused by the complex sensor interdependencies. Additionally, a WPP is influenced by seasonal components and a normal state of work cannot be declared as precise as for a machine that works in a homogeneous environment of a factory. This meets the requirement in point II (see section 1). In this solution, step 2 detects anomalies as deviation between an observation and the learned reference model of the system, this is described in section 4.2.

Common strategies for data-driven modeling are supervised and unsupervised learning methods. Supervised methods such as Multilayer Perceptron, Support Vector Machines or Naive Bayes Classifier (see [16] for more information) can be used to directly classify data according to learned hyperplanes in the data space. To be reliable, those methods need a-priori knowledge from labeled data of possible faults and the normal state. Gathering those precise data for a continuous production system like a WPP is hard to realize, as faults are rare and environmental conditions increase the number of possible faults dramatically.

In comparison, unsupervised learning methods (e.g. Clustering, Self Organizing Maps) seek to model data without any a-priori knowledge. Therefore, they are able to extract knowledge from unlabeled data sets and generate a model out of this knowledge. In this article, two types of unsupervised learning methods are investigated to model a WPP using unlabeled data. The PCA based modeling is compared against cluster based modeling methods, which are used as reference.

Clustering based modeling

The goal of cluster analysis is to partition data points into different groups. Similarity of points is defined by a minimal intra-cluster distance, whereas different cluster aim for a maximum inter-cluster distance. Thus, cluster analysis can be utilized to find the pattern of a system direct using the multi-dimensional data without explicit descriptions about the system features. This is the main advantage in using cluster analysis for modeling complex systems with seasonal components, e.g. WPP.

In the presented solution, a system model for anomaly detection should characterize the normal system behavior and can be used to identify unusual behavior. For most complex system, the normal behavior might consist of multiple modes that depend on different factors, e.g. work environments, operations of the systems. When the cluster analysis is performed on a data set representing the normal behavior of a system, multiple clusters can be recognized. Each cluster (group) represents a particular status of the system. Then such multiple clusters can be used as the normal behavior model of a system for anomaly detection.

In this paper, two well-known clustering algorithms, DBSCAN and spectral clustering, are utilized to model the normal behavior of a WPP on system level. Each of them has advantages in clustering the data with complex correlations.

DBSCAN is resistant to noise and can recognize patterns of arbitrary shapes. In DBSCAN, the density for a particular point is defined as the number of neighbor points within a specified radius of that point [17]. Two user-defined parameters are required: Eps - the radius; $MinPts$ - the minimal number of neighbors in the Eps. DBSCAN uses such center-based density to classify the data points as core point (Eps-neighbors $\geq MinPts$), border point (not core point but the neighbor of minimal one core point) or noise point (neither a core nor a border point). Two core points that are within Eps of each other are defined as density-reachable core points. DBSCAN partitions the data into clusters by iteratively labeling the data points and collecting density-reachable core points into same cluster. As result, DBSCAN delivers several clusters in which noise points are also collected in a cluster. DBSCAN is not suitable to cluster high dimensional data because density is more difficult to define in high dimensional space. Therefore, a method to reduce dimensionality should be applied to the data before using the DBSCAN. This leads to a density based description of the normal behavior.

This method assumes that the training data perfectly describe the distribution of system normal states. For WPP, some special states of the plant occur so rarely that the recorded data can not represent such special states very well. In addition, environmental influences lead to noise points within the data set. Therefore, a complete coverage of the normal states of a WPP in learning data set is unrealistic to achieve.

Compared to the traditional approaches to clustering (e.g. k-means, DBSCAN), spectral clustering can generally deliver better results and can be solved efficiently by standard linear algebra methods[18]. Another advantage of spectral clustering is the ability to handle the high dimensional data using spectral analysis. Thus, extra dimensionality reduction method is not required. The idea of spectral clustering is to represent the data in form of a similarity graph $G(V, E)$ where each vertex $v_i \in V$ presents a data point

Algorithm 1 PCA based modeling

1: **Input**: $\mathbf{X}$ ▷ learning data set
2: **Output**: $\mathbf{Model_X}$ ▷ model of input data

3: **procedure** PCA_BASED_MODELING ($\mathbf{X}$)
4: l: *reduced dimensionality*

5: $\mathbf{PCA\_Matrix}$ = *performPCA*($\mathbf{X}$)
6: $\mathbf{X}_{PCA}$ = *mapToLowDimension*($\mathbf{X}$)
7: $\mathbf{Model_X}$ = *generate_N-Tree*($\mathbf{X}_{PCA}$)
8: **end procedure**

9: **function** GENERATE_N-TREE($\mathbf{X}_{PCA}$)
10: $\mathbf{Tree}$: *List with length* 2^l
11: **for** ($\mathbf{x_{pca}}$ in $\mathbf{X}_{PCA}$) **do**
12: $\mathbf{i}$ = *determine_orthant*($\mathbf{x_{pca}}$)
13: $\mathbf{Tree_i}$ = append($\mathbf{Tree_i}$, $\mathbf{x_{pca}}$)
14: **end for**
15: **for** (*leaf* in *Tree*) **do**
16: **if** (*sizeOf(leaf)* > 1) **then**
17: *leaf = generate_N-Tree(leaf)*
18: **end if**
19: **end for**
20: **return** ($\mathbf{Tree}$, $\mathbf{PCA\_Matrix}$)
21: **end function**

in the dataset. Each edge $e_{ij} \in E$ between two vertices v_i and v_j carries a non-negative weight (similarity between the two points) w_{ij}. Then, the clustering problem can be handled as graph partition[19]. G will be divided into smaller components, such that the vertices within the small components have high connection and there are few connections between the small components. These small components correspond to the clusters in the results of spectral clustering and can be used as normal status model for anomaly detection.

PCA based modeling

Algorithm 1 presents the stated modeling solution for a system-level approach to a WPP. The algorithm utilizes the Principal Component Analyses (see, line 5 algorithm 1) as a very first step to achieve a dimensional reduced description of the training data set. Although a part of the information is lost due to the reduction, the sensor correlations in the low dimensional space are reduced drastically, which minimizes the computational effort.

The PCA is based on the assumption, that most of the information is located in the direction of most variance. Therefore, this method aims to project a data set to a subspace with a lower dimension by minimizing the sum of squares of $\mathbf{y_i}$ and to their projections $\theta_\mathbf{i}$ following cost function:

$$\sum_{i=1}^{m} = ||y_i - \theta_i||^2.$$

Let $\mathbf{x_1}, \dots, \mathbf{x_m}$ be the data point of m sensor values and $\mathbf{X}$ is a historical dataset of N scaled data points.

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,m} \end{bmatrix} \in \boldsymbol{R}^{N \times m}$$

Then as first step for computing the PCA, the covariance matrix is formed as

$$\boldsymbol{\Sigma}_0 \approx \frac{1}{\boldsymbol{N}-1} \boldsymbol{X}^T \boldsymbol{X}$$

By means of EVD (eigen value decomposition) or the equivalent SVD (singular value decomposition) the covariance matrix is decomposed as follows:

$$\boldsymbol{\Sigma}_0 = P^T \Lambda P, \Lambda = \begin{bmatrix} \Lambda_{pc} & 0 \\ 0 & \Lambda_{res} \end{bmatrix}$$

With $\Lambda = diag(\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_i^2)$ where σ_i, $i = 1, \cdots, m$ is the i-th eigenvalue and P is a matrix of the eigenvectors, sorted according to the eigenvalues of Λ. Λ_{pc} are the chosen principal components according to a threshold l and Λ_{res} denotes the less informative rest. l is a parameter which depends on the eigenvalues proportion of total variance and determines the dimension of the reduced normal space.

$$\boldsymbol{Y} = \boldsymbol{P}^T \boldsymbol{X}$$

Transforms the p-dimensional dataset $\mathbf{X}$ into a dataset $\mathbf{Y}$ of a lower dimension l, with a minimum of information loss. The axes of the dimensionally reduced data space are orthonormal and aligned to the maximum variance of data. Prerequisite for modeling a WPP with this kind of transformation is the input data to calculate eigenvalues and the rotation matrix. Therefore, the presented data set of a WPP needs to describe a period of fault free operation, which is denoted by the term 'normal state'. Using this data set as a learning base, the PCA described above spans a reduced normal state space, where signal covariances are taken into account due to the eigenvalues of the covariance matrix as the basis for transformation. The input variables are transformed within the algorithm 1 in line 6.

In comparison to clustering methods only the covariance matrix stores explicit shape informations. This leads to the necessity of taking into account all data points for classifying a new observation. That is why computational effort for this model increases with the number of data points in the data set and their dimension. To overcome this issue, the model is extended with an N-Tree as geometrical data structure (see function $generate\_N-Tree$ in algorithm 1). The axis of the PCA transformed normal state space divides the data into 2^l subspaces. Centering these subspaces in each iteration divides the subspaces recursively until each leaf of the tree contains one data point or is empty. Note, that the mean of each subspace needs to be stored.

4.2 Step 2: Anomaly Detection

To comply with point III (see section 1), the prerequisite for a system-level anomaly detection is a data-driven model as stated above. Given such a model, a distance measure is needed to calculate the deviation between a new system observation and the model in order to identify anomalies. Therefore, an observation vector needs to be transformed into the dimensionally reduced space of the model. Then the deviation of an actual observation and the learned model can be calculated using a distance metric, such as Euclidean distance, Mahalanobis distance or Manhattan distance.

DBSCAN generated cluster provide a discrimination of core and border data points. Distance computation in DBSCAN use the euclidean distance metric. Only core points are used to measure the distance between an observations

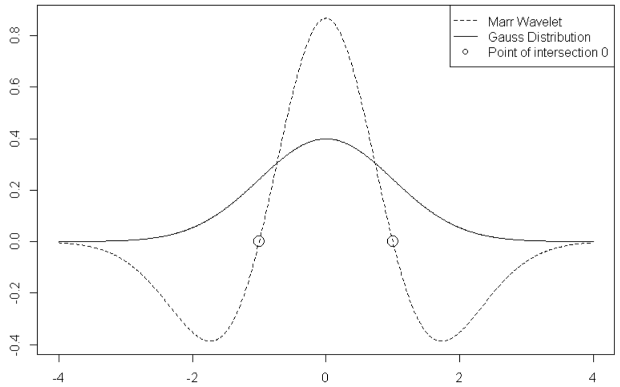

Figure 3: Characteristics of Gaussian distribution in comparison to Marr Wavelet (dashed). Spots are marked where the Marr Wavelet reach zero

and the core points. This leads to the decision whether an observation is part of the models cluster or not.

Spectral clustering computes clusters in a dimensionally reduced space but gives no further information about core or border points. Measuring the distance between such clusters can be achieved by a prototype, for example the cluster center. Then, for computing the distance, a metric like the Mahalanobis distance is used, which is sensible for the multidimensionality of such cluster. Representing a cluster based on a prototype is a generalization.

The PCA based modeling approach uses the dimensionally reduced input data as description of the multidimensional normal state space. Algorithm 2 shows how the model, computed with algorithm 1, is used for anomaly detection. At first a new observation is mapped to the low dimensional space of the model, using the rotation matrix from the PCA (see line 5). Then the mapped observation is compared with the normal state space. Therefore, the N-Tree is searched for its corresponding subset first (see function get_subset). If an empty leaf is found, all neighbor leafs are aggregated to a most relevant subset of data points. As the data is not generalized by border points or cluster means as prototypical points it is necessary to measure distance of the observation to each point of this subset. Now the distance is computed (see line 7).

Absolute distance measuring is missing a threshold to decide when an observation meets the model or not. Even when utilizing a Gaussian density function to provide an indicator for classification, a threshold needs to be estimated for classification. In this project, a Marr wavelet function is used to decide whether a new observation is part of the learned normal space. Instead of a Gaussian distribution the characteristic form of a Marr wavelet [20] allows a classification where the threshold can be set to zero, see figure 3. Taking into account the Marr wavelet and the euclidean distance function the process of distance measuring is computed as follows.

Let $X_{pca} = [x_1, \cdots, x_l]$ be a vector of the models' principal normal-space and $O_{pc} = [o_1, \cdots, o_l]$ a transformed observation, where l denotes the number of principal components. Then the distribution function to measure if a new observation is part of the normal state space is formed as:

$$\psi(X_{pca}, O_{pca}) = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}} \cdot 1 - \frac{k}{\sigma^2} \cdot \exp\left(-\frac{k}{2\sigma^2}\right)$$

Algorithm 2 Anomaly detection

```
1:  Input: Tree          ▷ (Learned model, see algorithm 1)
2:  Input: O                            ▷ Input observation
3:  Output: Boolean                              ▷ Anomaly
4:  procedure ANOMALY_DETECTION(Tree, O)
5:      O_PCA = mapToLowDimension(O)
6:      subset = get_subset(Tree, O)
7:      dist = calculate_distance(O, subset)

8:      if ( dist > 0 ) then
9:          anomaly: TRUE
10:     else
11:         anomaly: FALSE
12:     end if
13:     return ( anomaly )
14: end procedure

15: function GET_SUBSET( Tree , O_PCA )
16:     i = determine_orthant(x_pca)
17:     if (size(leaf_i) > 1) then
18:         get_subset(leaf_i)
19:     else
20:         subset = neighbors(leaf_i)
21:     end if
22:     return ( subset )
23: end function
```

Where l denotes dimensions of reduced normal-space and

$$k = \sqrt{\sum_{i}^{l} (O_{pca_i} - X_{pca_i})^2}$$

k is the l-space euclidean distance. For $\psi > 0$ an observation in principal space O_{pca} is denoted part of the normal state space (see line 17).

5 Results

The data used in the evaluation is collected over a duration of 4 years from 11 real WPPs in Germany with 10 minutes resolution. The dataset consists of 12 variables which describe the work environment (e.g. wind speed, air temperature) and the status of WPP (e.g. power capacity, rotation speed of generator, voltage of the transformer).

For evaluation, a training data set of 232749 observations of the 10 minutes resolution was used to model the normal behavior of a WPP. The evaluation data set of 11544 observations contains 4531 reported failures and 7013 observations of normal behavior. Table 1 shows the confusion matrix [21] as a result of the evaluation. Here, true negative denotes a correct predicted normal state and true positive a correct classified failure For this use case, the F1-score is used to analyze the system's performance in anomaly detection. Also, the runtime for the evaluation is denoted in Table 1 to compare speed performance of the different analyzed methods.

As can be seen, the presented PCA based algorithm outperforms the standardized spectral clustering. Especially a significant performance boost in computation time is achieved due to the extended N-Tree data structure.

Both, DBSCAN and Spectral Clustering, rely on complete sensor information for clustering the data set. A defect sensor leads to a maintenance action. The delay for this

	True Pos.	True Neg.	False Pos.	False Neg.	Bal. Acc.	F-Measure	elapsed Time
DBSCAN	1812	6827	186	2719	68.66%	55.50%	3s
Spectral Clustering	3832	6328	685	699	87.40%	84.71%	6637s
PCA based	3970	6517	496	561	90.27%	88.25%	68s

Table 1: Evaluation results of wind power station data.

maintenance is based on the localization of the WPP and cause missing sensor values for a certain time. To be operable in the use case of WPP such a model needs a fall back strategy in case of missing sensor values. Here, redundancy and correlation of different sensors comes in handy. By extending the PCA to a Probabilistic Principal Component Analyzes (PPCA), missing values can be estimated according to the data learned from the data set. Tipping and Bishop [22] extend a classic PCA by a probability model. This model assumes Gaussian distributed latent variables which can be inferred from the existing variables and the matrix of eigenvectors from the PCA. With the use of a PPCA, the solution for a system-level is robust enough to stay reliable even when sensors are missing. This was tested by training the model with a defect data set, containing 10% missing sensor values. While evaluating this model, also 10% of the data was damaged, simulating missing sensor values. The result of this evaluation is presented in table 1.

6 Conclusion

In this work a solution for system-level anomaly detection was presented. Three main requirements are identified and satisfied: At first a hardware concept for sensor data acquisition in the heterogeneous environment of WPPs was developed. This hardware logs existing sensor values and offers an adaptive solution to integrate new sensors on demand. Second, generic data-driven algorithms to automatically compute a system-level model out of minimal labeled, historical sensor data is presented. At last an anomaly detection method has been shown, which reaches an F-Measure of 89.02% and a ballanced accuracy of 91.46%. This solution is not specialized for specific parts of a WPP and can be trained in a short period. With an extension of the standard PCA to a probabilistic PCA, the robustness of the algorithmic solution against sensor failures is ensured.

In the future, this solution will be evaluated using data from more WPPs with different working environment. Beyond the task of anomaly detection, diagnosis of the root cause of an anomaly is also a sensible functionality of a CM system. The presented solution will be extended by a root cause analysis. Such an extension can support maintenance personal to trace the detected anomaly. Another focus will be the prognosis of anomalies in a WPP. To achieve this, an appropriate algorithm will be developed to predict the future system status using the learned model of the system behavior.

Acknowledgments

Funded by the German Federal Ministry for Economic Affairs and Energy, KF2074717KM3 & KF2074719KM3

References

[1] Global Wind Energy Council. Global wind statistics 2014. *Avaiable online at http://www.gwec.net/wp-content/uploads/2015/02/GWEC_GlobalWindStats 2014_FINAL_10.2.2015.pdf*, March 2015.

[2] European Wind Energy Association et al. Wind in power, 2012 european statistics, 2013. *Available online at http://www.ewea.org/fileadmin/files/library /publications/statistics/Wind_in_power_annual _statistics_2012.pdf*, March 2015.

[3] Julia Nilsson and Lina Bertling. Maintenance management of wind power systems using condition monitoring systems life cycle cost analysis for two case studies. *Energy Conversion, IEEE Transactions on*, 22(1):223–229, 2007.

[4] Wenxian Yang, Peter J Tavner, Christopher J Crabtree, Y Feng, and Y Qiu. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, 17(5):673–693, 2014.

[5] AS Zaher and SDJ McArthur. A multi-agent fault detection system for wind turbine defect recognition and diagnosis. In *Power Tech, 2007 IEEE Lausanne*, pages 22–27. IEEE, 2007.

[6] Li Zhen, He Zhengjia, Zi Yanyang, and Chen Xuefeng. Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme. *Mathematics and computers in simulation*, 79(3):318–338, 2008.

[7] Saad Chakkor, Mostafa Baghouri, and Abderrahmane Hajraoui. Performance analysis of faults detection in wind turbine generator based on high-resolution frequency estimation methods. *arXiv preprint arXiv:1409.6883*, 2014.

[8] E Jasinien, R Raiutis, A Voleiis, A Vladiauskas, D Mitchard, M Amos, et al. Ndt of wind turbine blades using adapted ultrasonic and radiographic techniques. *Insight-Non-Destructive Testing and Condition Monitoring*, 51(9):477–483, 2009.

[9] Oliver Niggemann and Volker Lohweg. On the diagnosis of cyber-physical production systems: State-of-the-art and research agenda. In *Association for the Advancement of Artificial Intelligence (AAAI)*, 2015.

[10] O. Bennouna, N. Heraud, and Z. Leonowicz. Condition monitoring and fault diagnosis system for offshore wind turbines. In *Environment and Electrical Engineering (EEEIC), 2012 11th International Conference on*, pages 13–17, May 2012.

[11] Ning Fang and Peng Guo. Wind generator tower vibration fault diagnosis and monitoring based on pca. In *Control and Decision Conference (CCDC), 2013 25th Chinese*, pages 1924–1929, May 2013.

[12] M. Celik, F. Dadaser-Celik, and A.S. Dokuz. Anomaly detection in temperature data using dbscan algorithm. In *Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on*, pages 91–95, June 2011.

[13] P. Baraldi, F. Di Maio, and E. Zio. Unsupervised clustering for fault diagnosis. In *Prognostics and System Health Management (PHM), 2012 IEEE Conference on*, pages 1–9, May 2012.

[14] Karlheinz Schwarz and Im Eichbaeumle. Iec 61850, iec 61400-25 and iec 61970: Information models and information exchange for electric power systems. *Proceedings of the Distributech*, pages 1–5, 2004.

[15] Richard A Zatorski. System and method for controlling multiple devices via general purpose input/output (gpio) hardware, June 27 2006. US Patent 7,069,365.

[16] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley, 2006.

[17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD96)*, 1996.

[18] Ulrike von Luxburg. A tutorial on spectral clustering. In *Statistics and Computing, 17 (4)*, 2007.

[19] Shifei Ding, Liwen Zhang, and Yu Zhang. Research on spectral clustering algorithms and prospects. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, volume 6, 16-18 2010.

[20] Lei Nie, Shouguo Wu, Jianwei Wang, Longzhen Zheng, Xiangqin Lin, and Lei Rui. Continuous wavelet transform and its application to resolving and quantifying the overlapped voltammetric peaks. *Analytica chimica acta*, 450(1):185–192, 2001.

[21] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley, 2006.

[22] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.