51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition
07 - 10 January 2013, Grapevine (Dallas/Ft. Worth Region), Texas

AIAA 2013-0313

Predicting Faults in Wind Turbines using SCADA Data

Anders Bech Borchersen*

Vattenfall R&D, 7000 Fredericia, Denmark

Jesper Abildgaard Larsen† Jakob Stoustrup‡

Department of Electronic Systems, Section of Automation and Control, Aalborg University, 9220 Aalborg, Denmark

The cost of operation and maintenance of wind turbines is a significant part of the overall cost of wind turbines. To reduce this cost a method for enabling early fault detection is proposed and tested in this paper. The method is taking advantage of the fact that wind turbines in wind farms are located near similar wind turbines. This is done by generating a model for each turbine, the model is then used to evaluate the performance of that turbine and the nearby turbines. The evaluations from the models are then combined and used as votes to identify the faulty turbines. The method is applied and tested on historical Supervisory Control And Data Acquisition (SCADA) data from nine operational turbines over a testing period of nine months. The performance of the fault detection is found to be acceptable based on the testing period. During the testing period several gear related services were performed, some of these were predicted by the proposed fault detection systems. The advantage of the purposed method is that it applicable for operational turbines without requiring any extra measurements, since the used SCADA data is available from most modern wind turbines.

I. Introduction

Placing wind turbines offshore is attractive due to the better wind condition at sea. But when wind turbines are placed offshore Operation and Maintenance (O&M) increase compared to onshore, from 10-20% to 30% of the total cost.[1] This is due to the lower and more expensive accessibility offshore, caused by the weather conditions at sea. Thus to make offshore wind energy competitive to fossil fuel energy the cost of O&M must be reduced.

The reason for the high O&M cost for offshore wind turbines today is to some extent caused by the way the maintenance is conducted. The maintenance today is conducted as reactive maintenance combined with preventive maintenance.[2] This means that whenever a turbine stops, a service team is sent to the turbine (reactive maintenance), and every year a profound service is performed (preventive maintenance). Instead of using the current maintenance strategy, predictive maintenance could reduce the number of services and the downtime of the turbines.[3]

To enable the change to predictive maintenance it is necessary to have predictions about upcoming faults. Theses predictions shall then be used in the planning of upcoming maintenance by giving an alarm before the failures occurs and causes the turbines too stop.

Fault detection of wind turbines has been study extensively for individual turbines[4,5] and exists to some extent in industrial practice. Normally it is based on either Condition Monitoring Systems (CMS), or internal alarms generated by the turbine. The main problem with CMS is the high cost and thus these systems are not included in all turbines.[6] The internal alarms generated from the turbines tends to be either too conservative,

*Industrial PhD Fellow, Vattenfall R&D, Oldenborggade 25-31, 7000 Fredericia, Denmark

†Associate Professor, Section for Automation and Control, Aalborg University, Frederik Bajers Vej 7C, 9220 Aalborg, Denmark

‡Professor, Section for Automation and Control, Aalborg University, Frederik Bajers Vej 7C, 9220 Aalborg, Denmark

and thereby give non or very short reaction time before the failure occurs; or be too sensitive to small changes which are within normal operation, and then flooding the O&M center with false alarms which is undesirable.[7] All turbines sends operational data to the Supervisory Control And Data Acquisition (SCADA) system. The operational data is called SCADA data. SCADA data consist of production numbers, wind speed, and crucial temperature measurements, these are stored for each turbine. The production numbers are used for evaluating the performance of the turbines. The temperature measurements are used in fault detection by checking that they are within specified thresholds, if these thresholds are violated an alarm is given.[8]

In this paper a method for improving the fault detection based on SCADA data is proposed. The structure of the SCADA system is described in Section II. In Section III the method for detection faults in turbines by using the assumption that wind turbines standing nearby each other should be under similar conditions.The proposed method is then tested in Section IV using data from nine operational wind turbines. Over a test period of nine months reasonable result were achieved, by using the temperature measurements available from the SCADA system. The achieved results are finally discussed in Section V.

II. System Structure

A simplified structure of a SCADA system is shown in Figure 1. The work in this paper are related to

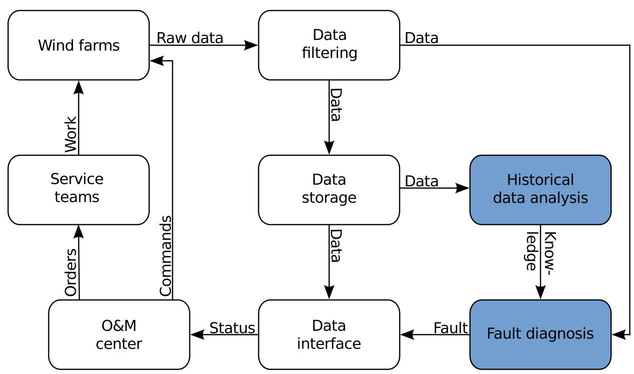

Figure 1. Overview of how the data from the turbines are used in the O&M of the turbines. The data is sent from the wind farms through a data filtering module into a central data storage which then present the data to the O&M center via a data interface. In the O&M center the planning of services is then conducted based on the information from the turbines. To improve the planning of services a fault detection system consisting of a historical data analysis and a online fault diagnosis has been added.

the historical data analysis, since this provides the foundation for the online fault diagnosis. The assumption is that faults happened in the past are likely to happen in the future. Thus methods that can detect historical faults, would also be able to detect future faults. Therefor when a method working well on the historical data has been found, knowledge from the historical analysis of the method are then used to ease the implementation in the online fault diagnosis system.

III. Method

To find a suitable method for detecting faults in the wind turbines based on SCADA data a look into the data is needed. The data available in this study is temperature measurements from the SCADA system given as 10 minute averages. An example of the SCADA data is given in Figure 2. From the figure it is clearly seen that the SCADA data contains periods with no data. The length of these periods can vary from one sample to several weeks if a profound service is performed at the turbine.

The structure of the FDI method is shown in Figure 3 for N turbines. The fault detection system is based on the assumption that a majority of the turbines are non-faulty. Thus a fault can be detected if three or more turbines are grouped together. For each turbine in the fault detection system a model following the individual turbine is derived. This gives a fault detection system with N models, each of these models are then fed with the data from all turbines. The predicted output of each model is then compared to the measured output, to generate the residuals. Thereby each model generates N residuals, one for each turbine.

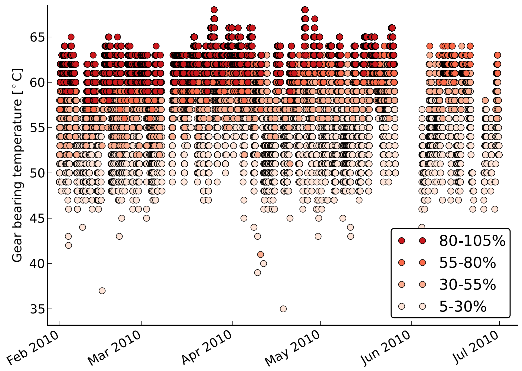

Figure 2. View of the raw SCADA data from the gear bearing. The color is indicating the relative power production from the turbine.

The residuals are then evaluated through a change detection method. Based on the residual evaluation a vote for each turbine is then given, 0 for fault free and 1 for faulty. Thereby each model votes on all N turbines. The votes are summed up for the individual turbine to generate the final result.

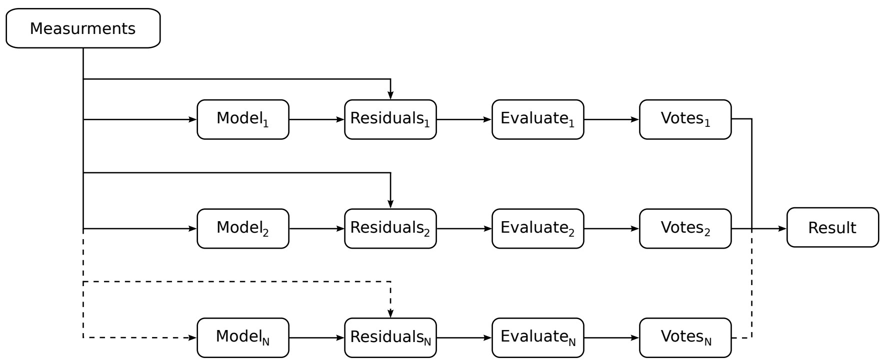

Figure 3. Overview of how the fault detection system for N turbines, all measurements fed into all N models, and thereby the models evaluate the performance of all turbines.

Since the data contains period where there are missing samples, thus making it difficult to use models that depend on previous measurements since this can not be ensured. Therefore the model should be able to handle longer periods without any data. Furthermore there is no temperature model available for the turbine, thus a data driven model is proposed using Partial Least Squares (PLS).

The residuals are evaluated using the CUmulitive SUM (CUSUM) algorithm. CUSUM is selected since it is well suited for detection change in the mean value.[9] The assumption is that a turbine is faulty when the mean value of the residual is non zero.

The design of the fault detection can be summarized in the following steps:

1. Find a group of N similar turbines.
2. Generate a PLS model fitted to each individual turbine.
3. Use the model of each turbine to generate residuals using measurements from all N turbines.
4. Evaluate the residuals from the models and vote for faulty turbines.
5. Combine the N votes from each model into a result.
6. Mark the turbines according to the number of votes they get.
7. Go to step 2 if a new model is needed, else go to step 3.

When designing PLS models the inputs and the outputs must be found. Then the model order and the amount of data used when generation the model must be determined. These are normally found by sweeping each parameter and then evaluated over a predefined prediction period.

IV. Test Case

To test if the proposed method is suitable for predicting faults in wind turbines, it is tested on data from nine turbines located near each other in a wind farm. The test period is from October 2009 to June 2010. In this period ten gear related services were conducted, in which three of them were exchanges of the gear. The reason for predicting gear faults in particular that it has come up as one of the main faults in the wind turbines.[10] To detect the faults related to the gear the gear bearing temperature is selected as the output of the PLS model. The inputs to the model are the other 22 temperature measurements available in the SCADA data.

A. Determining Model Parameters

The parameters that must be found beside the inputs and the outputs for the PLS model, are the number of principle components and the length of the modeling period. These parameters are found by sweeping each of the parameters, which are shown in Figure 4 and Figure 5. The sweeping has been done for three turbines and the averages are taken. The prediction period used for finding the RMSEP is the ten days following the modeling period.

A reasonable model order are found to be 6 based on the plot of shown in Figure 4. Based on the plot shown in Figure 5 a sample size of 1500 samples seems reasonable. This corresponds to approximately 10 days of data. The modeling parameters which will be used are then 6 principle components and the modelling period will be set to 10 days. These are found as a good compromise between model complexity and RMSEP.

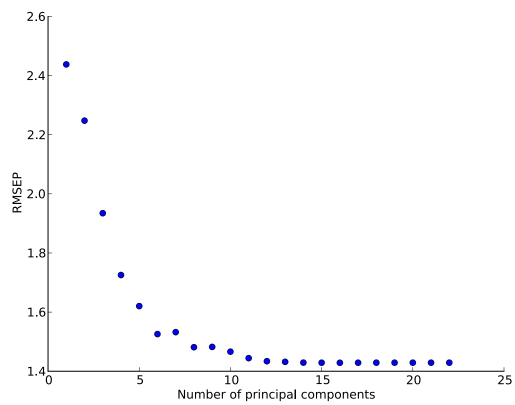

Figure 4. Plot of RMSEP with respect to the number of principle components used in the model.

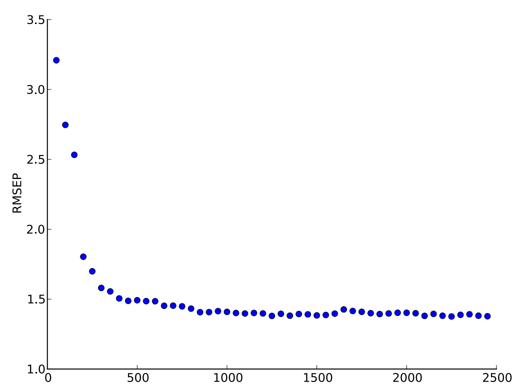

Figure 5. Plot of how the RMSEP is changing according to the number of samples included in the model.

B. Fault Detection Output

A five day output of the fault detection system together with the faults related to the gears are shown in Figure 6. The output of the fault detection system is indicated by a colored dot (the darker the red the more votes for the turbine to be faulty). The dot is indicating the evaluation over a five day period, for every ten day period a new PLS model are generated using the previous 10 days of data.

The performed services are marked in the figure with the text "Gear". Three of them where exchanges of the gear boxes. These happened in turbine 1, 4, and 8. The exchanges are identified by the periods with no output from the fault detection system in April for these turbines. The missing data is caused by the turbines are stopped when the gear is exchanged. The period from end of May to start of June with no data from all the turbines are caused by a loss of SCADA data in the period (the turbines are still running). The other periods with missing data are caused by missing data from the turbines which can come from other services or simply a bad data connection to the turbine.

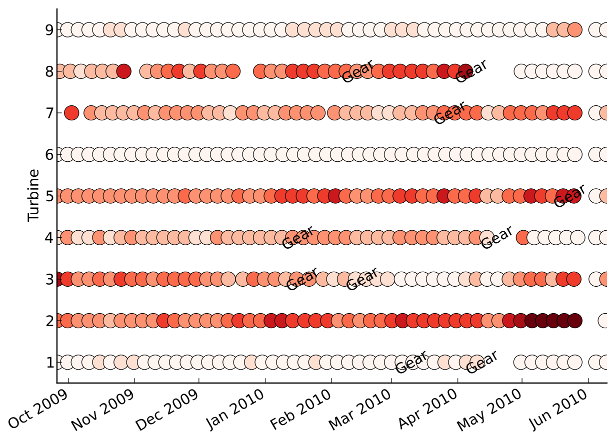

Figure 6. Output from the fault detection system over the test period.

V. Discussion

It is seen that the selection of the number of modeling parameters is easy since beyond a certain number of model parameters, the model does not improve in the prediction performance. The output of the fault detection system is more difficult to evaluate since the assumption about less than half of the turbines are faulty at the same time might not hold in all circumstances in practice. This hypothesis can only be tested statistically by counting the number of consistent votes. In specific instances it can not be determined if a voting process has a correct outcome. The problem is that it is only known when the service was performed not whether the fault actually occurred. Further, it can never been known when the faults occur since this will require the fault to be inserted at purpose at known time instances. Therefore the output of the fault detection system will be evaluated for each turbine against the known services that has been performed.

Turbine 1 The fault detection system does not detect anything for this turbine even though one service and one change of gear is performed during the testing period. Thus the performance criterion is set as not satisfied.

Turbine 2 The fault detection system detects this one as becoming more and more faulty until the state is change in mid June. This could be a clear indication of a non-reported service since the turbine has a longer out-take period than the others in this period. The performance is set to unknown since further investigation are needed to clarify what happened in June for this turbine.

Turbine 3 This turbine has two services related to the gear during this period. The fault detection system reacts to some extent to these, but not very clear detection is seen. Thus the performance is acceptable.

Turbine 4 This turbine has both a service and a gear exchange during the period and there is a weak detection of both. Acceptable performance.

Turbine 5 One gear service which is clearly identify long time ahead. Acceptable performance.

Turbine 6 No service and no detection. Acceptable performance.

Turbine 7 One service which is weakly detected but the system keeps identifying the turbine as faulty. This could indicate that the turbine would fail in the future but it is not registered in the service reports. Thus the performance for this turbine is unknown.

Turbine 8 One service and one gear exchange both clearly detected ahead and after gear change the turbines is detected as non faulty as expected. Very good performance.

Turbine 9 No service and only a weak identification end the end of the testing period. Acceptable performance.

So overall the fault detection system showed identification of five of the seven services, and the identification of the faults where approximately one month ahead of the services. The detection system had acceptable performance for six of the turbines and one with an average performance, and one with unknown performance and 1 false alarm. This is overall seen as acceptable result for the proposed method.

VI. Conclusion

In this paper the structure of a typical SCADA system used for O&M of turbines has been presented and extended with a fault detection system, which is based on the assumption that wind turbines near each other in a wind farms should be under the same operating conditions. The fault detection system is designed such that it only uses available SCADA data. This is achieved by using PLS models for modelling the relationship between different measurements. CUSUM algorithm are used for evaluating the residuals from the turbines. The method has then been tested using real measurements from nine operational turbines with acceptable results. Thereby a method for detecting faults in wind turbines using the existing SCADA system has been proposed and tested. The method can easily be extended to use other measurements as outputs to detect other types of faults in the turbines. The next step will then be to use the method for fault detection on-line and use the predictions in the planning of future O&M.

Acknowledgments

Anders Bech Borchersen would like to thanks Vattenfall R&D and the Danish Agency for Science Technology and Innovation for their financial support.

References

[1] Blanco, M. I., "The economics of wind energy," *Renewable and Sustainable Energy Reviews*, Vol. 13, No. 6-7, Aug. 2009, pp. 1372–1382.

[2] Walford, C. A., "Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs," Tech. Rep. March, 2006.

[3] Swanson, L., "Linking maintenance strategies to performance," *International Journal of Production Economics*, Vol. 70, No. 3, 2001, pp. 237–244.

[4] Odgaard, P., Stoustrup, J., and Kinnaert, M., "Fault tolerant control of wind turbines a benchmark model," *Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical processes*, Vol. 1, 2009, pp. 155–160.

[5] Odgaard, P., Stoustrup, J., Nielsen, R., and Damgaard, C., "Observer based detection of sensor faults in wind turbines," *Proceedings of European Wind Energy Conference*, 2009, pp. 1–10.

[6] Amirat, Y., Benbouzid, M., Al-Ahmar, E., Bensaker, B., and Turri, S., "A brief status on condition monitoring and fault diagnosis in wind energy conversion systems," *Renewable and sustainable energy reviews*, Vol. 13, No. 9, 2009, pp. 2629–2636.

[7] Qiu, Y., Feng, Y., Tavner, P., Richardson, P., Erdos, G., and Chen, B., "Wind turbine SCADA alarm analysis for improving reliability," *Wind Energy*, 2011.

[8] Kusiak, A. and Li, W., "The prediction and diagnosis of wind turbine faults," *Renewable Energy*, Vol. 36, No. 1, 2011, pp. 16–23.

[9] Gustafsson, F., *Adaptive Filtering and Change Detection*, John Wiley & Sons, Ltd, Chichester, UK, Oct. 2001.

[10] Borchersen, A. B., Stoustrup, J., and Larsen, J. A., "Fault Analysis of Wind Turbines Based on Error Messages and Work Orders," *10th European Workshop on Advanced Control and Diagnosis*, 2012.