Loughborough University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

creative commons

COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

Disclaimer

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

A Model-based Approach to Wind Turbine Condition Monitoring using SCADA Data

W. G. Garlick*, R. Dixon*S. J. Watson*

**Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK*
(Tel: +44(0)1509 227016; e-mail: w.g.garlick@lboro.ac.uk).
(Tel: +44(0)1509 227018; e-mail: r.dixon@lboro.ac.uk).
(Tel: +44(0)1509 63 348; e-mail: s.j.watson@lboro.ac.uk).

Abstract: Modern wind turbines are complex aerodynamic, mechanical and electrical machines incorporating sophisticated control systems. Their design continues to increase in size and they are increasingly being positioned offshore where the environment is hostile and where there are limited windows of opportunity for repair and maintenance activities. Condition monitoring is essential offshore if Wind Turbines (WTs) are to achieve the high reliability necessary for sustained operation. Contemporary WT monitoring systems already provide vast amounts of data, the essential basis of condition monitoring, much of which is ignored until a fault or breakdown occurs. This paper presents a model-based approach to condition monitoring of WT bearings. The backbone of the approach is the use of a least squares algorithm for estimating the parameters of a discrete time transfer function (TF) model relating WT generator temperature to bearing temperature. The model is first fitted to data where it is known no problems exist. It is then used in predictive mode and the estimates of the bearing temperature are compared with the real measurements. The authors propose that significant discrepancies between the two are indicative of a developing problem with the bearings. The promising experimental results achieved so far indicate that the approach is viable.

Keywords: ARX model, condition monitoring, data-based modelling, SCADA data, wind turbine.

1. INTRODUCTION

It is anticipated that electrical power generated from renewable energy will, over the next twenty years, become a significant part of the total generating capacity of the European Union (EU). For example, the installed capacity of wind energy in Europe, between 1995 and 2004, increased from 2.5 GW to 34 GW, Bousseau, P. et al. (2006). The UK has a high level of offshore wind energy resource that when fully developed will present new management and control challenges for network operators. The operation of Wind Turbines (WTs) has not been without its problems of reliability, Tavner, P.J. et al. (2008), but the hostile offshore environment will further increase the demand for high reliability turbines and cost effective condition monitoring systems will have a part to play in achieving high reliability. Contemporary WT monitoring systems already provide vast amounts of data, much of which is ignored until a fault or breakdown occurs. If this data can be used to identify potential failures or breakdowns, then this may well prove to be a very cost effective means of condition monitoring.

Generator bearing temperature is monitored in many WTs and as bearing problems cause significant down times, it seems a good candidate for the demonstration of model-based condition monitoring. SCADA records covering three years operation of a land based wind farm were provided for use by the Supergen Wind project. Preliminary analysis of the bearing temperature data indicated the possible presence of dynamic terms; therefore, a model-based approach using system identification techniques provided an appropriate analysis approach. The use of system identification for condition monitoring is not new, Isermann (1993) and Dixon & Pike (2002). However, it has not previously been applied to WTs nor does it tend to be used with SCADA data – as it is generally assumed dynamics are removed by the nature of the data which is usually averaged.

In this paper, the authors use system identification methods applied to raw SCADA data to develop discrete-time dynamic models which can be used to predict generator bearing temperature – in practice this measurement often provides an early indication of bearing and gearbox faults. The paper is structured as follows: The model-based monitoring approach is outlined in section 2. Section 3 discusses the data-based approach that is used to obtain fault-free process models. In section 4, modelling results are presented with a range of different models discussed. In section 5 the models are applied to data from turbines where faults are known to have occurred and the results are cross-checked with written logs to ensure real faults have been identified. In section 6 conclusions are drawn and suggestions made for future work.

2. CONDITION MONITORING APPROACH

A basic model-based condition monitoring scheme is shown in Fig. 1. It shows a time dependent process, $G(t)$ and an associated model of that process $\hat{G}(t)$. Let's assume that the model $\hat{G}(t)$ has been chosen to accurately predict the measured output $y(t)$ based on a known input $u(t)$. In this case any differences between the output $y(t)$ and the predicted output, $\hat{y}(t)$, should be due to unmeasured disturbances or changes in the process. So by monitoring this difference (also known as the *error* or *residual*) it should be possible to infer whether changes in the condition of the plant *G(t)* have occurred. This is the foundation of the approach being pursued in this paper. The hypothesis is that it is possible to detect wind-turbine faults, and in particular bearing faults, at an early stage by comparing measured bearing temperatures with those predicted by a model of the system. If the error is large this indicates a potential fault in the WT or bearing.

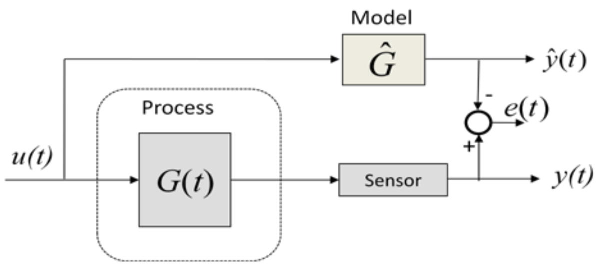

Fig. 1 Model-based monitoring

3. PROCESS MODELLING

3.1 Introduction

The general approach taken in this paper is to use system identification techniques to fit discrete time transfer function (TF) models which relate a suitable measured variable to generator bearing temperature. This section describes the WT data available and the system identification approach used in the paper.

3.1 The WT Data

The data available took the form of 10 minute averaged SCADA records for a total of 12 turbines. Fifty four key parameters, including generator bearing temperature, are recorded and the complete sets of data for each WT covered a period of 3 years starting from February 2004. It should be noted that the data was not all contiguous and contained gaps of varying lengths due to high/low wind speeds and turbine outages caused by faults or maintenance activities. Finally, written logs recording significant maintenance activities or issues with each turbine were also available; these are used later in the paper to cross-check potential faults identified from the data against what was recorded as actually happening.

3.2 System Identification Approach

There are a profusion of algorithms available for obtaining estimates of model parameters for a system, e.g. the well known method of least squares (LS), as well as other, more complex approaches, such as: extended least squares (ELS), instrumental variables (IV) and refined instrumental variables (RIV). Detailed information about these and other methods can be found in the many texts on the subject, such as Young (1984), Ljung (2006). All of these can be used to estimate the parameters of linear, discrete-time transfer function models (with a variety of different assumptions made about the structure for the different algorithms). In this work, the LS method will be used and an AutoRegressive with eXogenous input (ARX) model structure is assumed. The ARX model, Fig. 2, benefits from simplicity but has the disadvantage that disturbances are part of the system dynamics. However, the effects of this may be reduced with a good signal-to-noise ratio, which should be the case for time averaged SCADA data. Therefore, it seems a suitable choice.

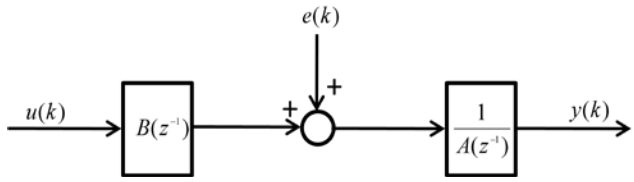

Fig. 2 General ARX Model

2.3 ARX Process model

The process model can be written in TF form as,

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})}u(k) + \frac{e(k)}{A(z^{-1})} \tag{1}$$

where $y(k)$ and $\underline{u}(k)$ are the output and input at the k^{th} sample point, z^{-1} is the backward shift operator and the parameter polynomials are of the form,

$$A(z^{-1}) = 1 + a_1 z^{-1} \cdots + a_n z^{-n} \tag{2}$$

$$B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} \cdots + b_m z^{-m} \tag{3}$$

finally, $e(k)$ is a zero mean white noise sequence with variance σ^2, representing the unmeasured stochastic inputs, measurement noise or disturbances that may influence the system.

3.3 Model Structure Selection

Having decided on an ARX model of the form given by equation 1, it is necessary to decide on the number of numerator and denominator parameters, that is: *m* and *n* in (2) and (3). A variety of statistical evaluation criteria are discussed in the literature, Ljung (2006), and can be used, alongside prior knowledge of the system and engineering judgement, as a means of evaluating a set of candidate model structures. The coefficient of determination, R_T^2, is defined in Young (1984) as:

$$R_T^2 = 1 - \frac{\sigma^2}{\sigma_y^2} \tag{4}$$

where σ^2 is the sample variance of the model residuals $e(k)$ and σ_y^2 is the sample variance of the measured system output $y(k)$ about its mean value. As a result this goodness of fit criterion tends to unity as the fit of the model to the data improves.

A second figure of merit is Akaike's Information Criterion (AIC), Ljung (2006), which takes account of the number of parameters and their estimated variance as well as the fit. For the AIC the largest negative value is said to indicate the "best" model. Here engineering judgement dictates that a proper model structure without any delays is appropriate. This assumption is used alongside the AIC and R_T^2 to select from the full set of candidate models.

3.4 Input Data Selection and Data Length

As the generator bearings (and bearing faults) are the point of interest for this study, selection of the generator bearing temperature as the output, $y(k)$, is judged to be essential. However, the choice of model input, $u(k)$, is less straightforward. To the engineer there are two natural choices the power and the wind-speed. However, because there were many variables available it was decided to use correlation analysis to isolate the best candidate inputs.

The accuracy of the model parameters depends upon the amount of data used to produce the model and on whether that data contains sufficient excitation (movement of the signals). In general, more data leads to a more accurate model. As the model is intended to be used for condition monitoring, in a predictive mode, the data used to construct the model needs to be as small as possible while still giving good prediction capabilities (on fault free data) in order to minimise initialisation (or training) of the condition monitoring model. Models constructed from data ranging from 1 day to 5 months were initially built to explore the variation in accuracy.

4. BEARING TEMPERATURE MODELLING

4.1 Introduction

In this section, models are fitted to data in order to identify a suitable model structure in terms of input and number of parameters (section 4.2). Then the models are used in a predictive mode and the results compared with unseen data (section 4.3) to see how they perform assuming no faults.

4.2 *Model Identification*

As stated earlier, the output, y(k), will be the WT bearing temperature and there were three initial questions to answer: the number of model parameters, the range of error free data to produce the model and which data to use as a model input.

a) Fault free data check

A check of the fault logs of all 12 WTs quickly identified suitable fault free data ranges. It was possible to select at least 1 day of error free data from all WTs, but only one WT had the first 5 months operation error free, turbine reference WT17658 and the data from this machine was used in all subsequent modelling.

b) Data correlation check

At this stage of the research it was not clear how much data to use in building a process model and therefore three models were initially built, using 1 day, 3 and 5 months data from 2004. A correlation check was made, of data used for each model, between the bearing temperature and all the other SCADA data for WT17658. The data that gave the highest correlation were found to be: maximum power, generator winding temperature, and average wind speed. Table 1 shows some of the correlation results for WT 17658.

Table 1 Correlation between generator bearing temperature and other SCADA data for WT 17658

SCADA data	1 day data correlation	3 month data correlation	5 month data correlation
Max. power	0.755	0.733	0.667
Generator winding temp.	0.949	0.922	0.876
Average wind speed	0.787	0.761	0.677

The highest correlation with the generator bearing temperature was expected to be either maximum power or wind speed. However, the generator winding temperature data generally had the highest correlation and the 1 day model provided the highest correlation in each of the three input data sets and was therefore used in all subsequent work. Note that lower correlations may be caused by the lack of or intermittent data or an indication that the WT has a bearing problem already.

c) Number of model parameters

Initially, parameters were estimated for a range of models with the number of parameters, the same in *A* and *B*, ranging from 1 – 6. The criteria used to select the best model were as mentioned in section 3.3. Table 2 shows that models with 3-6 parameters had similar results in terms of AIC and R_T^2 , thus, a 3-parameter model (3 parameters in *A* and *B*) was chosen for subsequent modelling. This model structure is applied to the data from each WT in the remainder of the paper.

Table 2 Examples of a process model with increasing number of parameters

No. of parameters	Coefficient of determination	AIC
1	0.6905	0.3075
2	0.7964	-0.6528
3	0.8018	-0.6971
4	0.8	-0.7022
5	0.7992	-0.7024
6	0.7999	-0.7023

In order to have the most data to compare with that predicted by the model, the model based on the least amount of data with the highest correlation was preferred. But as a further check of which model size produced an error distribution close to a normal distribution, the distribution of errors for each model was compared to each other and to a normal distribution curve using generator temperature data (for the model data range) as input to the model. The distribution of model errors, Figs. 3-5, was significantly different and counter-intuitively, the 1 day model had the worst fit to a normal distribution curve in spite of having the highest correlation.

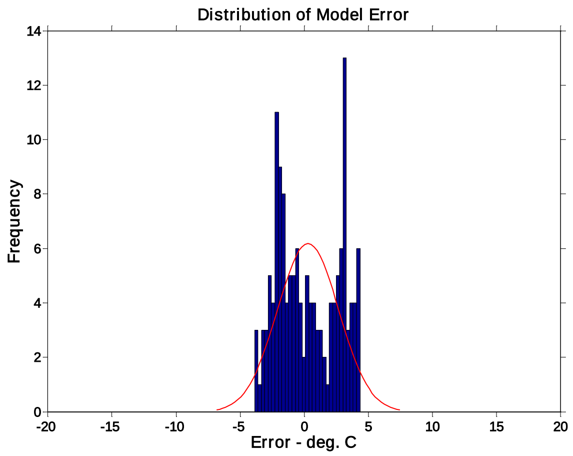

Fig. 3 Distribution of errors for a 1 day model

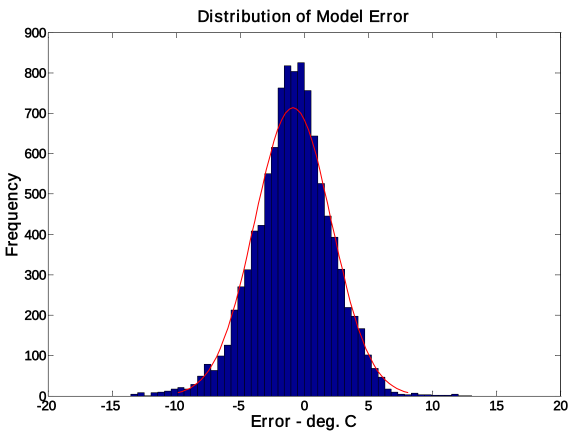

Fig. 4 Distribution of errors for a 3 month model

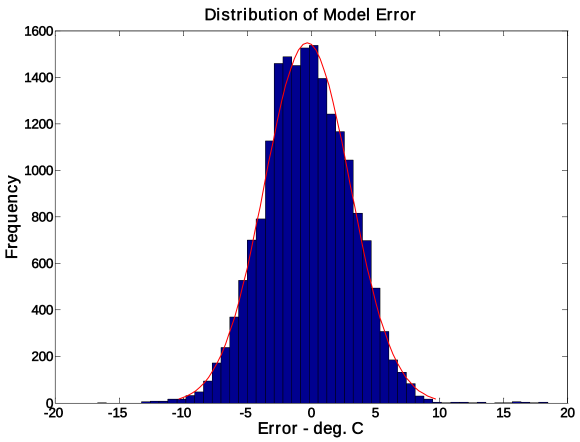

Fig. 5 Distribution of errors for a 5 month model

The results in Fig. 3 are not close to a normal distribution and for an ARX model this would usually indicate that the error was not wholly stochastic, i.e. there were still some unmodelled dynamics in the error. In this case, scarcity of data is the likely cause (this is backed-up by the normal distribution obtained for the 1day models when used in predictive mode – see section 4.3).

4.3 Process models used in predictive mode.

The value of the parameters in each model were compared, Table 3, and found to be similar, which may be of future interest in the identification of a generic model for all WTs of a similar design.

Table 3 Model parameters

Parameter	**1-day model**	**3-month model**	**5-month model**
$\mathbf{a_1}$	-1.1929	-1.0854	-1.0893
$\mathbf{a_2}$	0.1047	0.0528	0.0161
$\mathbf{a_3}$	0.1694	0.0637	0.1003
$\mathbf{b_1}$	0.2774	0.4176	0.3816
$\mathbf{b_2}$	-0.2838	-0.5078	-0.4547
$\mathbf{b_3}$	0.0350	0.0973	0.0802

To use the process models in a predictive mode, data from outside the range used to create the process model was applied to the model. The model output was then compared to the real bearing temperature data. Generator winding temperature data for a one year period, 2004, was applied to each WT process model and the distribution of model error results compared. In spite of having dissimilar distribution of errors over the model period, the predictive results were very similar with standard deviations of 4.3, 4.6 and 4.4.

The initial conclusion drawn from these result are that the correlation of data used to produce the process model may be a better guide to the model performance than the distribution of errors when the model is built from sparse data.

These results confirm that the process models are similar, particularly over the dominant terms, a_{1-3}. The 3 and 5 month models were very similar over all their parameters.

5. MONITORING RESULTS: IDENTIFICATION OF GENERATOR BEARING FAULTS

5.1 Introduction

A 1-day model for WT17658 with generator temperature data as an input was used to produce the results in this section. The approach adopted to identify bearing faults was first to examine the process error and distribution of error plots, from the predictive mode studies, for any identifiable features that may indicate a change or deviation from normal operation and secondly, to compare the change to the fault log.

5.2 2004 Data

The process errors plot for the 2004 data, Fig. 6, shows four significant features:

a) the solid central part indicates that the error in most bearing temperature predictions appear in this band of approximately 10º C

b) narrow positive going spikes indicate short excursions to temperatures outside the normal range. These spikes do not appear to be randomly distributed but occur for two periods: 22 June – 18 July and 11 November – 11 December. The fault log for these periods in 2004 records problems with high generator temperatures and slip rings fan.

c) large negative spikes indicate zero temperatures or null readings caused when the generator is not operational. Removing these from the data resulted in a negligible change to the model and therefore they were not removed from the data.

d) a generally rising mean level with a sudden reduction in August, shortly after the end of the first group of high temperatures, followed by a lower mean level for the rest of the year.

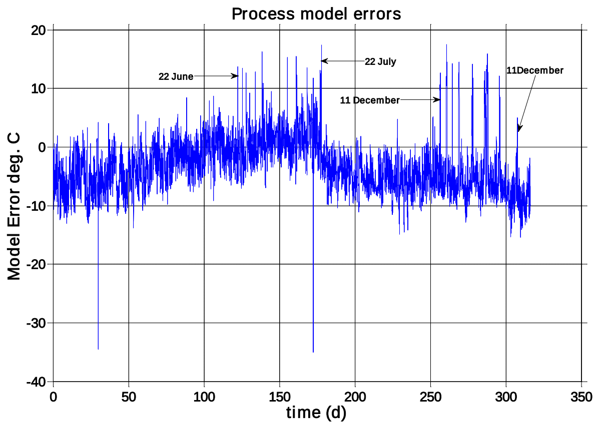

Fig. 6 Process model errors – WT17658 2004 data

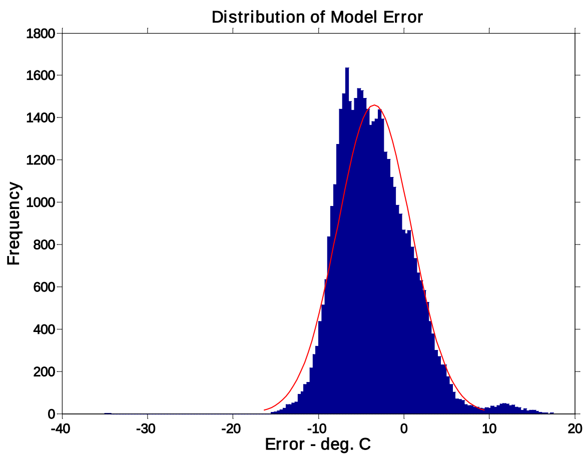

Fig. 7 Distribution of errors – WT 17658 2004 data

The distribution of errors plot, Fig. 7, shows two significant features:

a) generally a reasonable fit to the normal curve but with a small shift to the left, -7 or -8º C.

b) a small secondary peak at approximately +12º C and above the normal distribution curve at that point.

These results do not help to identify any bearing problems but serve to indicate generator problems, problems that are in the data used as the model input.

5.3 2004 Data

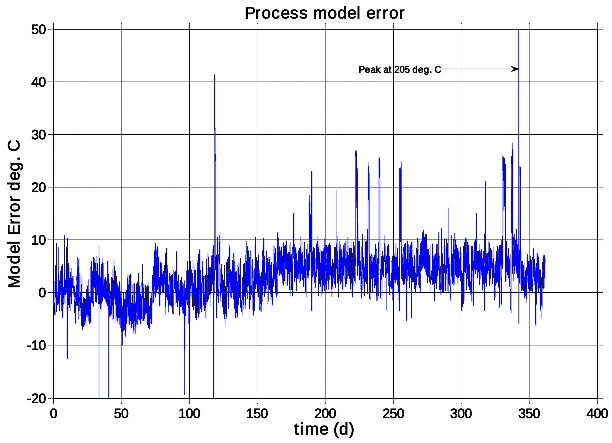

Fig. 8 Process model error – WT17658 2005 data

The process errors plot for the 2005 data, Fig. 8, shows two significant features:

a) a variable mean band for the first four months. The fault log records high generator temperatures and thermo error slip ring fan, which appear to be a continuation of the problems from 2004.

b) Positive going spikes, sometimes quite broad spread over the period May – December. The fault log records that the generator bearings were changed on 6, 12 and 13 December, but do not indicate whether this was one or three changes. Periods of generation can be seen between these changes and notably, on 12 December bearing temperatures of 205º C were recorded. This indicates a severe problem, possibly a shaft misalignment or lubrication failure, so one is lead to believe that possibly three changes took place in the month.

The distribution of errors plot, Fig. 9, shows two significant features:

a) the accumulation of high temperatures above the normal distribution curve, in this case at approx24º C, is again evident. For clarity and ease of comparison with other results, the very high bearing temperatures at 205º C are not shown.

b) the distribution of errors has a very peaky and narrow distribution at the centre of the normal curve.

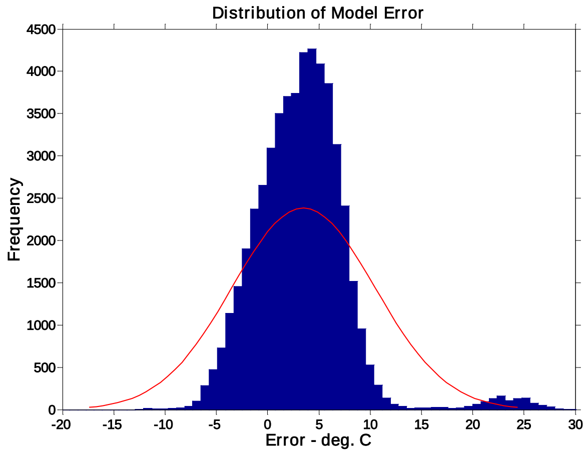

Fig. 9 Distribution of errors – WT17658 2005 data

5.4 2006 Data

The process errors plot for the 2006 data, Fig. 10. The positive spikes are again evident commencing towards the middle of March 2006 and continue to the end of the year. Once again the mean level starts to change towards the end of September and the fault log records problems with the generator fan.

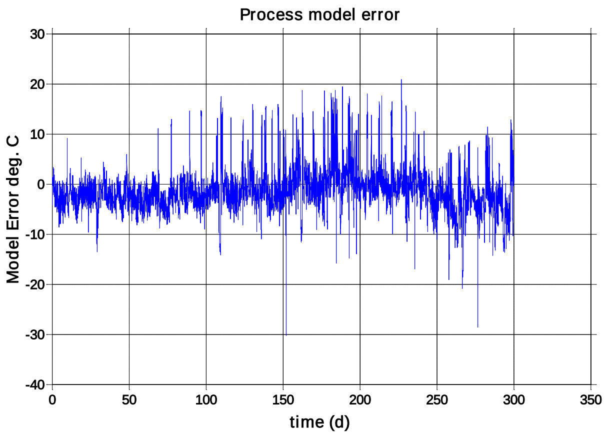

Fig. 10 Process model error – WT17658 2006 data

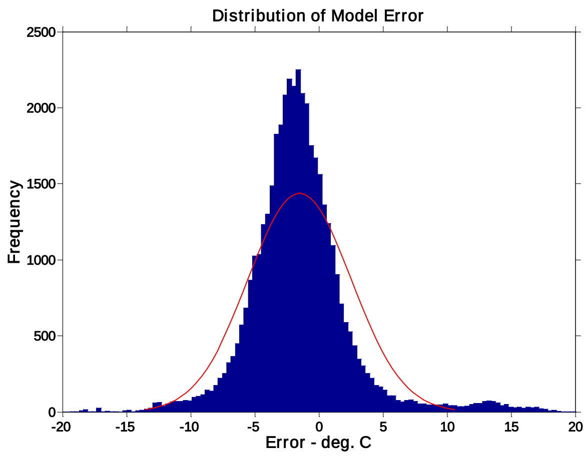

Fig. 11 Distribution of errors – WT17658 2006 data

The distribution of errors, Fig. 11, shows the response becoming narrower and peaky with a secondary peak at +12-14° C. The conclusion drawn is that high generator temperatures were again evident in 2006, but also that rising bearing temperatures may indicate that in 2007 the bearing problems from 2005 re-occurred, but data for 2007 had not been made available to verify this conclusion.

6. SUMMARY

The results of the analysis show that it is possible to build an ARX process model of WT generator bearing temperature from a relatively short SCADA data set and to use the model to identify the onset of the significant features and failures in the fault log concerned with the generator and the bearings. The process used raw unfiltered SCADA data.

Currently evaluation of the monitoring results needs human intervention. Further work needs to identify a means of doing this more automatically. The results to-date suggest an approach that uses a sliding window over some nominal period (say one month or one year) and looks for three factors: the change in mean prediction error (in the window); a threshold for the secondary peak and a running count of high bearing temperatures recorded, with an alarm flag raised after a pre-set number of large errors in a period.

REFERENCES

Bousseau, P., Belhomme R., Monnot R., Laverdure N, Boëda D., Roye D., Bacha S., (2006). Contribution of wind farms to ancillary services. *CIGRE Session 2006*, paper C6-103.

Tavner, P.J. Spinato, F. van Bussel, G.J.W. Koutoulakos, E. (2008). Reliability of different wind turbine concepts with relevance to offshore application. *European Wind Energy Conference (EWEC)*, Brussels, 31 March - 3 April 2008

Isermann, R. (1993). Fault diagnosis of machines via parameter estimation and knowledge processing. Tutorial paper, *Automatica*, volume (29), pp 815-835.

Dixon, R., Pike A.W. (2002). Application of condition monitoring to an electro-mechanical actuator: a parameter estimation based approach. *IEE Computing & Control Engineering Journal*, volume (13), pp. 71-81.

Ljung, L. (2006). *System Identification theory for the user*, Prentice Hall, Englewood Cliffs, N.J., ISBN 0-13-656695-2

Young, P.C. (1984). *Recursive Estimation and Time Series Analysis.* (Communication and Control Engineering Series), Springer-Verlag, Berlin, 1984.