

Wind turbine fault detection based on SCADA data analysis using ANN

Zhen-You Zhang · Ke-Sheng Wang

Received: 16 December 2013 / Accepted: 27 January 2014 / Published online: 2 March 2014
© Shanghai University and Springer-Verlag Berlin Heidelberg 2014

Abstract Wind energy is one of the fast growing sources of power production currently, and there is a great demand to reduce the cost of operation and maintenance. Most wind farms have installed supervisory control and data acquisition (SCADA) systems for system control and logging data. However, the collected data are not used effectively. This paper proposes a fault detection method for main bearing wind turbine based on existing SCADA data using an artificial neural network (ANN). The ANN model for the normal behavior is established, and the difference between theoretical and actual values of the parameters is then calculated. Thus the early stage of main bearing fault can be identified to let the operator have sufficient time to make more informed decisions for maintenance.

Keywords Artificial neural network (ANN) · Supervisory control and data acquisition (SCADA) · Wind turbine · Fault detection

1 Introduction

Renewable energy source is playing an important role in the global energy mix, as a means of reducing the impact of energy production on climate change. Wind energy is the most developed renewable energy technology worldwide with more than 282.48 GW installed capacity at the end of

2012 [1]. Certain forecasts indicate that the share of wind in Europe's energy production will reach up to 20% in the close future [2]. Today, large wind turbines (2–6 MW) are becoming established as economically viable alternatives to traditional fossil-fuelled power generation. In some countries, such as Denmark, Germany and Spain, wind turbine has become a key part of the national power networks [3].

Condition monitoring of wind turbines is of increasing importance as the size and remote locations of wind turbines used nowadays make the technical availability of the turbine very crucial. Unexpected faults, especially of large and crucial components, can lead to excessive downtime and cost because of restricted turbine accessibility especially for some remote controlled wind farms on mountain and offshore wind farms. However, even smaller issues and faults of auxiliary equipment like pumps or fans can also cause expensive turbine downtime due to the same causes. From an operator's point of view it is therefore worth increasing the effort spent to monitor the turbine condition in order to reduce unscheduled downtime and thus operational costs. The key part of wind turbine monitoring system is to detect fault (fault diagnosis) of turbines as early as possible so that the maintenance staff can manage and prepare the maintenance action in advance.

Most wind turbines installed nowadays are integrated with supervisory control and data acquisition (SCADA) system which can monitor the main components. SCADA system typically monitors parameters such as temperatures of bearings, lubricating oil, windings and vibration levels of driven train [4]. These monitored data are collected and stored via an SCADA system that archives the information in a convenient manner, usually for all of the turbines in the wind farm. These data quickly accumulate to create large and unmanageable volumes that can hinder attempts to

Z.-Y. Zhang (✉)
Department of Wind Park Management, Kongsberg Maritime AS, Trondheim, Norway
e-mail: zhenyou.zhang@kongsberg.com

Z.-Y. Zhang · K.-S. Wang
Department of Production and Quality Engineering, Norwegian University of Science and Technology, Trondheim, Norway

deduce the health of a turbine’s components. It would prove beneficial from the perspective of utility companies, if the data could be analyzed and interpreted automatically to support the operators in identifying defects. One main function of SCADA data analysis is fault detection and diagnosis as early as possible to support the decision of maintenance action and operation.

Methods used for fault diagnosis of wind turbine can be classified as model based or data based. Model based methods require a comprehensive physical or mathematical model which is normally unavailable. Success of data based methods is conditioned by the significance of historical data and the mathematical method to detect the patterns in data. For wind turbine systems where an important amount of data is stored regularly by SCADA system and process model is not available, using the data driven methods is preferred [5].

This paper describes artificial neural network (ANN) that can be used to identify incipient faults in the main component of a turbine, such as main bearing, gearbox and blades, through the analysis of these SCADA data. The SCADA data sets are already collected and stored, and therefore no new installation of specific sensors or diagnostic equipment is required. The technique develops normal behavior model by ANN and SCADA data analysis which can calculate the theoretical value of related parameters and compare to the real measurement of the same parameters. The parameters mentioned above can be indicator of abnormal behavior of incipient component failure. In this way, only interesting information is highlighted to the operator, therefore significantly reducing the volume of data they are faced with. This paper just take the main shaft rear bearing monitoring as an instance to show how the technique works.

The remaining sections of this paper are organized as follows. Sections 2 and 3 describe the principle of ANN and the data of SCADA system. Section 4 describes how to establish ANN normal behavior model for main shaft rear bearing, and Section 5 shows the results of fault detection using established ANN model. Sections 6 and 7 discuss some issues of this method and give conclusion respectively.

2 Artificial neural network

The pattern classification theory has become a key factor in fault detection and diagnosis for wind turbines. Currently, ANN is one of the most popular methods in this domain. ANN is a model that emulates a biological neural network [6]. The origin of ANN can be traced back to a seminar paper by McCulloch and Pitts [7] that demonstrated a collection of connected processors, loosely modeled on the

organization of brain, and could theoretically perform any logical or arithmetic operations. Then, the development of ANN techniques is very fast which is extensive to many categories containing back-propagation (BP), self-organization mapping (SOM) and radial basis function (RBF), etc. The application of ANN model lies in the fact that it can be used to infer a function from observations. This is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical. This attribution is very nontrivial in diagnostic problems for wind turbine.

ANN deal with complex non-linear problem without sophisticated and specialized knowledge of the real systems. It is an effective classification technique and low operational response times needed after training. The relationship between the condition of component and the features is not linear but non-linear. ANN does not need to know the exact form of analytical function on which the model should be built. This means neither the functional type nor the number and position of the parameters in the model-function need to know. It can deal with multi-input, multi-output, quantitative or qualitative, complex system with very good abilities of data fusion, self-adaptation and parallel processing. Therefore, it is very suitable to select as a method of fault detection and diagnosis for wind turbine.

The standard BP (SBP) neural network which is the most widely used neural network model currently was proposed by Rumelhart and McClland in 1986 [8]. It is a multilayer feed-forward network usually containing the input layer, hidden layer and output layer (see Fig. 1), which is trained by an error back propagation algorithm. The biggest advantage of ANNs trained by back propagation is that there does not need to know the exact form of analytical function on which model should be built. Thus it’s not necessary to have neither the function type nor even

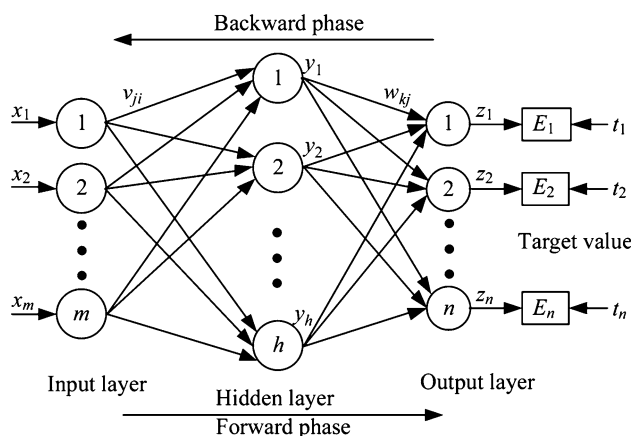


Fig. 1 BP neural network with single hidden layer

the number and position of the parameters in the model function. Moreover, BP network can learn and store a lot of input-output model mappings without mathematical equations describing these mappings. The learning method of BP is the steepest descent method which is adjusting the weights and thresholds of the network to minimize the sum of squared errors. The general procedure of BP network training can be summarized as follows [6]:

- (i) Initialize the weights to small random values $(-1, 1)$;
- (ii) Select a training vector pair (input and the corresponding desired output) from the training set and present the input vector to the inputs layer of the ANN;
- (iii) Calculate the actual outputs (forward phase);
- (iv) Adjust the weights to reduce the difference according to the error between actual output and target (backward phase);
- (v) Return to Step (ii) and repeat for each pattern p until the total error has reached an acceptable level;
- (vi) Stop.

3 SCADA dataset description

An operational wind farm typically generates vast quantities of data which are well known SCADA data. (i) The SCADA data contain information about every aspect of a wind farm, from power output and wind speed to any errors registered within the system. Thus by keeping track of both wind speed and power output parameters, the overall health of the turbine can be supervised. (ii) SCADA data may be effectively used to “tune” a wind farm, providing early warning of possible failures and optimizing power output across many turbines in all conditions.

It is common for “condition monitoring” to be applied to a wind farm. However, this involves the addition of extra instrumentation, including wind farm down time, extra cost and potential warranty implications. As distinct from condition monitoring, performance monitoring using existing instrumentation to analyze SCADA data of wind turbines is no extra instrumentation, no down time and no cost. It has the advantage of using data already routinely gathered. By making use of specially-designed software tools, a great deal of information may be gathered and analyzed to provide a detailed look at the performance of the wind farm.

Typical parameters recorded by SCADA on wind turbines can be broadly categorized into following types which can be used in fault detection and diagnosis activity [9]:

- (i) wind parameters, such as wind speed and wind deviations;

- (ii) performance parameters, such as power output, rotor speed, and blade pitch angle;
- (iii) vibration parameters, such as tower acceleration and drive train acceleration;
- (iv) temperature parameters, such as bearing temperature and gearbox temperature.

Specifically the SCADA data recorded and used for condition monitoring from wind turbines are as follows: active power output (10 min max/min/average), anemometer-measured wind speed (10 min max/min/average), turbine speed (10 min max/min/average), nacelle temperature (10 min max/min/average), turbine rear bearing temperature (10 min max/min/average), turbine rear vibration (10 min RMS max/min/average), turbine front bearing temperature (10 min max/min/average), turbine front vibration (10 min RMS max/min/average).

The technique presented utilizes only some types of the data mentioned above. The parameters listed above are typical of data collected and stored by commercial wind turbine SCADA systems. This means that the approach developed in this paper can be widely applied by wind farm operators.

4 Modeling of SCADA parameter normal behavior

SCADA data from wind farm of Hundhammerfjellet owned by NTE - Nord-Trøndelag Elektrisitetsverk are used in this paper. A parameter of main shaft rear bearing in the SCADA data, i.e., turbine rear bearing temperature, gives an indication of how hot of the bearing are running, and therefore offers the possibility to detect rear bearing overheating. The straightforward threshold check which has already been applied in real wind farm, could be used to flag up temperature exceeding a certain limit, and this might be too late to avoid significant damage to the main shaft rear bearing. The desired functionality should be taken into consideration any relevant aspects of turbine operation. This approach would allow temperatures to be detected that are too high in the context of the concurrent level of power generation, leading to a quicker and more effective identification of abnormal behavior.

4.1 Parameter selection

To establish the normal behavior ANN model of main shaft rear bearing temperature, the variables that can affect this temperature must be taken into consideration to build an accurate model. Wind turbines can only aerodynamically capture a proportion of the energy in the incident wind [10]. This energy is converted by the rotor blades into mechanical power and is transmitted directly to the

Table 1 Input and outputs of ANN model

Model output	Input
Rear bearing temperature	Rear bearing temperature ($t - 1$)
	Active power output (t)
	Nacelle temperature (t)
	Turbine speed (t)

generator by the main shaft because the turbine monitored in this paper is direct-driven turbine without gearbox. Zaher et al. [11] established the normal behavior model of

gearbox bearing temperature and cooling oil temperature, and Garcia et al. [12] built the same model in addition to cooling oil thermal difference. The former model utilized active power and nacelle temperature while the later also utilized the operation of fans to cool the gearbox. However, there is no gearbox in direct-driven wind turbine, which avoids the faults of comprehensive gearbox. The main shaft bearings are the key components of this type of wind turbine. Therefore, one of the key components, main shaft rear bearing, is main monitoring object in this paper.

Accordingly, the parameters that may affect the rear bearing temperature contain active power output, nacelle

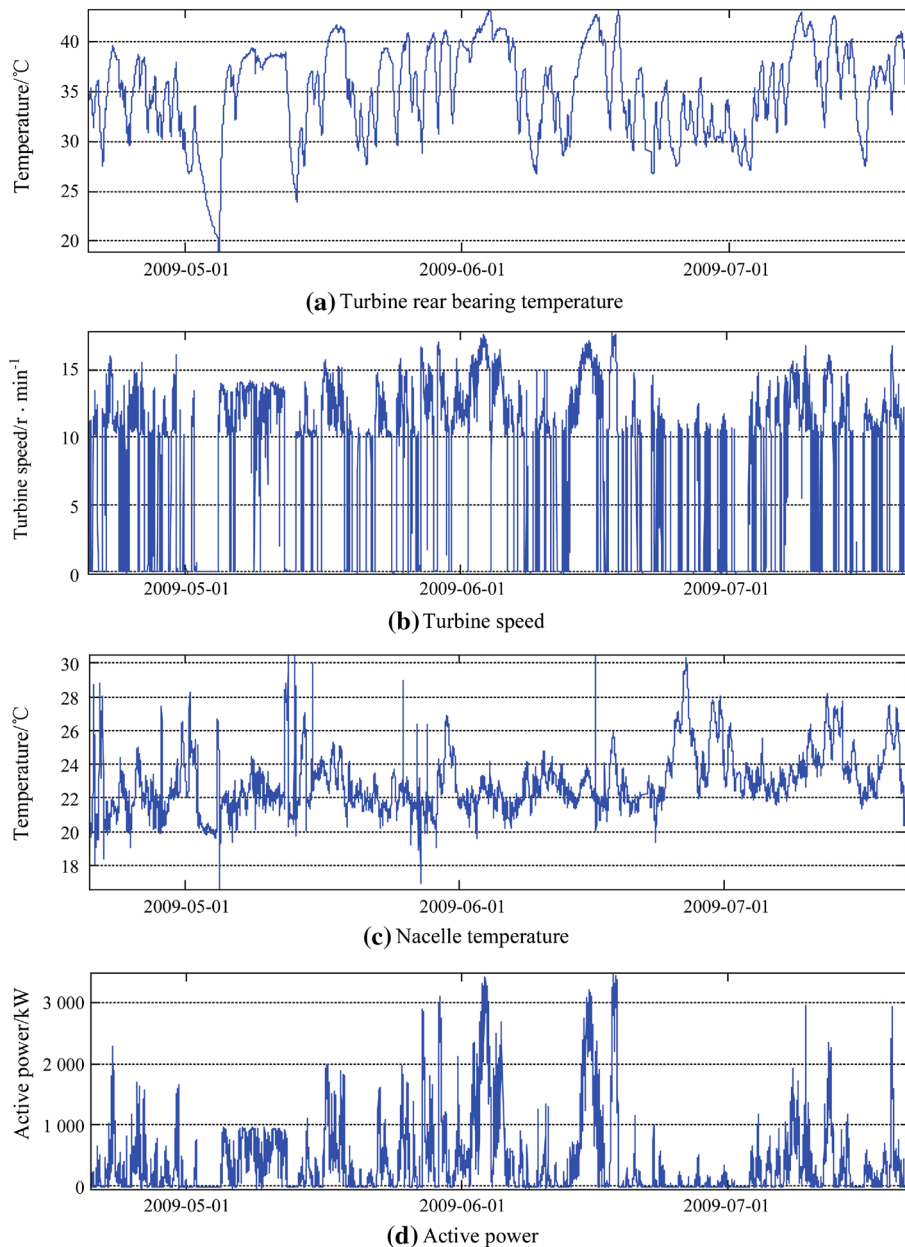


Fig. 2 Neural network turbine rear bearing temperature model training data

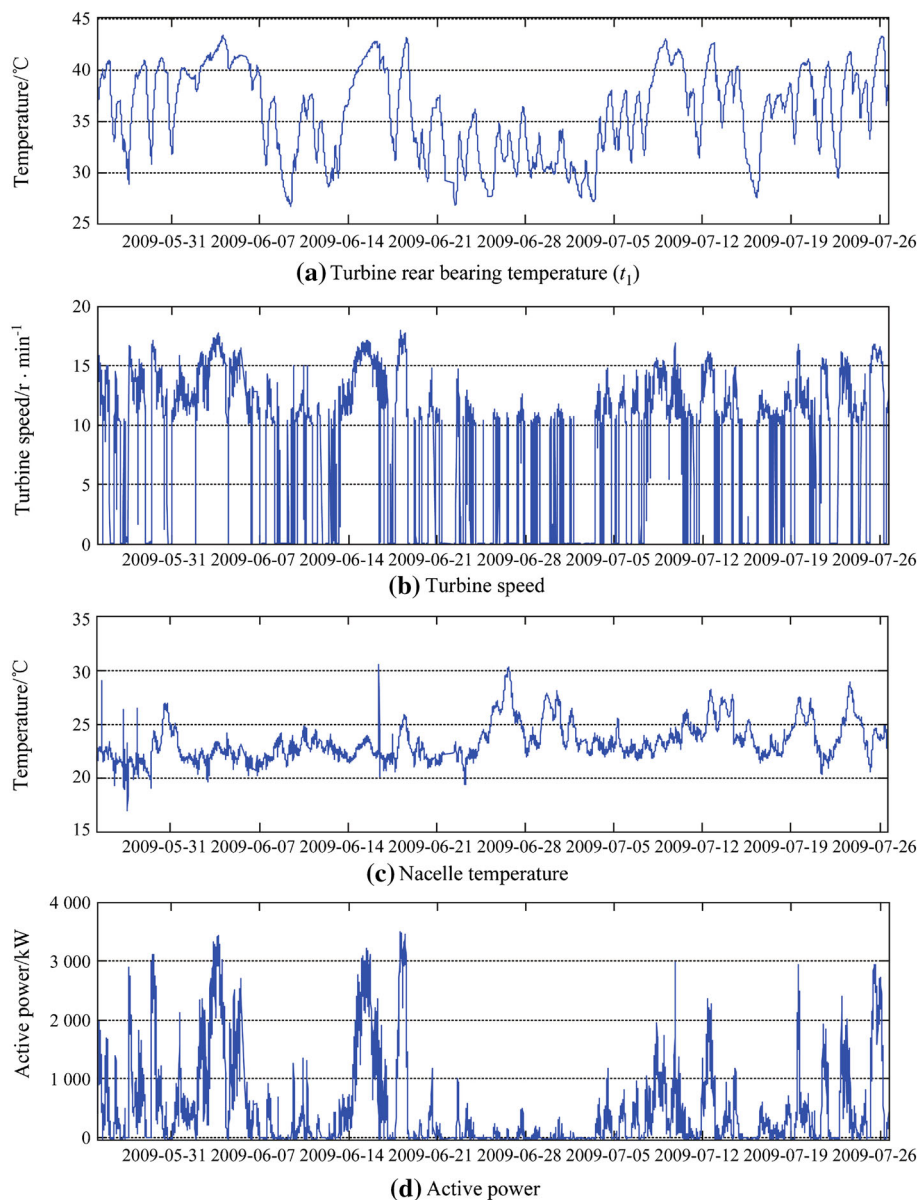


Fig. 3 Rear bearing model testing input data

temperature, turbine speed and cooling fan status. Unfortunately, the cooling fan status is not available in current SCADA data. Thus the parameters selected to establish ANN model for the parameter of main shaft rear bearing temperature can be chosen as seen in Table 1.

4.2 Training ANN model

The models are trained using the parameters discussed in Table 1. In order to get an accurate representation for the parameter under study, the range of the parameters as inputs to ANN should be as varied as possible while still ensuring that the turbines are in normal operational condition in ANN training process. This is achieved through selection the period of training data with many conditions

as possible including the starting and stopping of turbine, big changes of turbine speed, with and without active power output. Therefore, SCADA data from April 22, 2009 to July 21, 2009 are chosen as training data for ANN rear bearing normal behavior model as seen in Fig. 2.

This amounts to roughly 13,000 data points for each input. The training process then attempts to capture the nonlinear relationship between these parameters, i.e., the associated rear bearing and nacelle temperatures for the turbine speed and corresponding power output. The number of training cycles used, also known as epochs, was 1,000. Determining the architecture for the network is an iterative process and depends solely on the structure that yields the best accuracy when tested. The final architecture used for rear bearing model was 5-10-1.

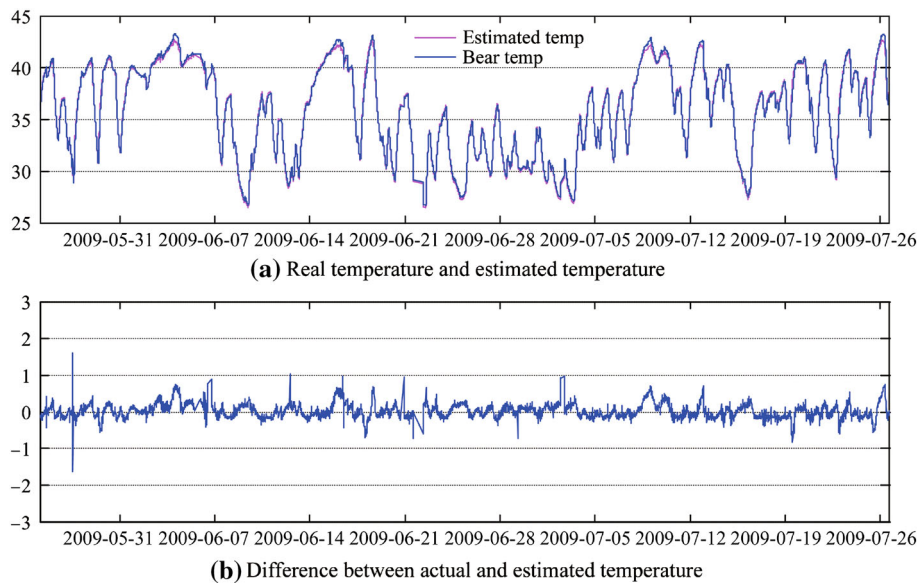


Fig. 4 Rear bearing model output in normal condition

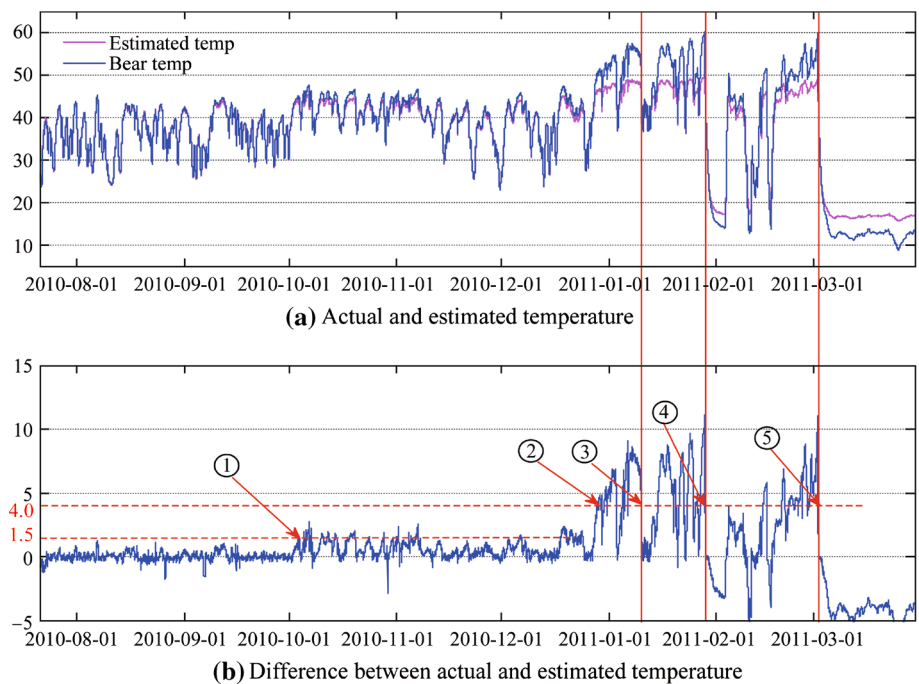


Fig. 5 Fault detection results of rear bearing

The trained model was tested in new data from a healthy wind turbine which had not been used in training process of ANN. Figure 3 shows the input data used to test rear bearing ANN model from the same turbine of the date from May 26, 2009 to July 26, 2009. The test data shown here was also very varied. Figure 4 shows the output of rear bearing ANN model (Estimated temp), the real temperature of that (Bear temp) and the difference between these values. The average difference between actual and estimated

value is 0.026 °C, and the root mean square error is 0.2, which is considered to be an acceptable level for diagnosis requirement for successful fault detection. This means that the output of ANN model can be used directly as a comparison with the actual temperature to assess if a fault is present. If the difference between the estimated value and the actual value increases for a continuous number of instances, i.e., a prolonged period of time and not a minor fluctuation, then this will flag as a fault.

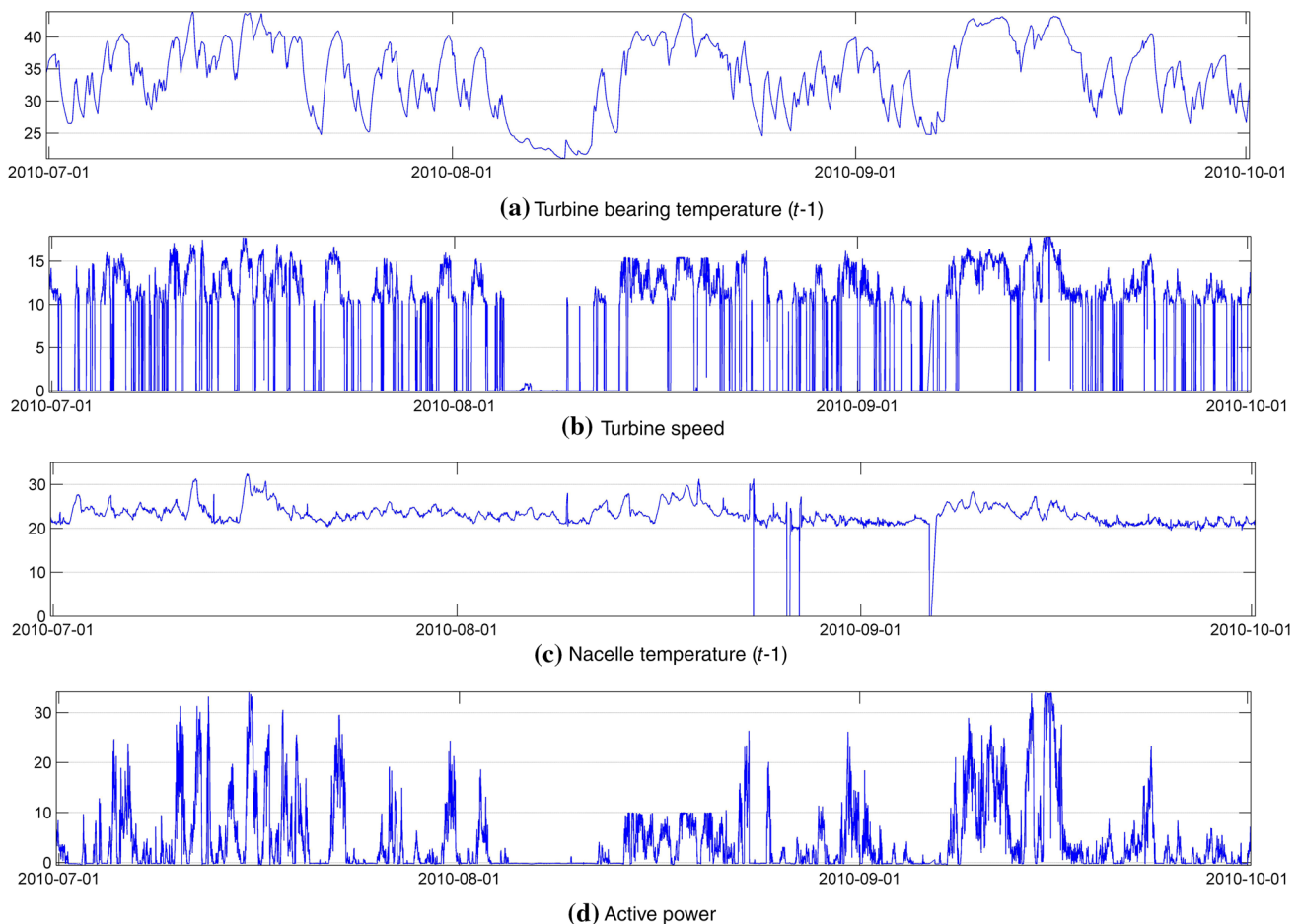


Fig. 6 Rear bearing model testing input data of new turbine

5 Detection of rear bearing fault

Once the normal behavior of rear bearing ANN model is trained, it can be used to detect the corresponding fault of rear bearing by comparing estimated and actual temperature. Figure 5a shows the evolution of rear bearing temperature from the period of July 2010 to March 2011 which contains eight months where it eventually fails. Figure 5b shows the difference trend between the estimated and actual temperature of rear bearing in this period. The first important deviation from the model estimation occurred from the start of October 2010, i.e., point ①. The frequency of deviation and their duration increased in the following months. From point ②, the deviation from the model estimates increased to 4 °C and lasted to point ③ where the turbine was stopped because of overheating. Then, the operator of wind farm tried to solve the problem two times in point ③ and point ④, but not successful and finally the turbine completely stopped because of the same overheating. From Fig. 5, the method can give the operator a warning as early as three months in point ① before the failure happens. With the evolution of the failure, the

deviation from model estimation increases and the alarm can be given to operator when the deviation reaches the level of point ②. Therefore, the alarm can be given as early as 10 d before the failure happens.

The results produced by ANN model for rear bearing fault detection are very positive. They can provide an early warning of problems developing in the bearing before the absolute temperature becomes apparently high. The results of the fault detection can be used to help the operator to make the schedule of maintenance actions before the failure happens to reduce the maintenance cost, the unanticipated downtime and improve the reliability of the wind turbine.

6 Discussion

This section mainly discusses whether the model established from the SCADA data of one turbine as in Fig. 2 can apply in fault detection for other turbines. Then, another turbine is selected to test for this purpose. However, there is no same fault in this wind farm and thus the SCADA data are from a new turbine in normal condition. If the

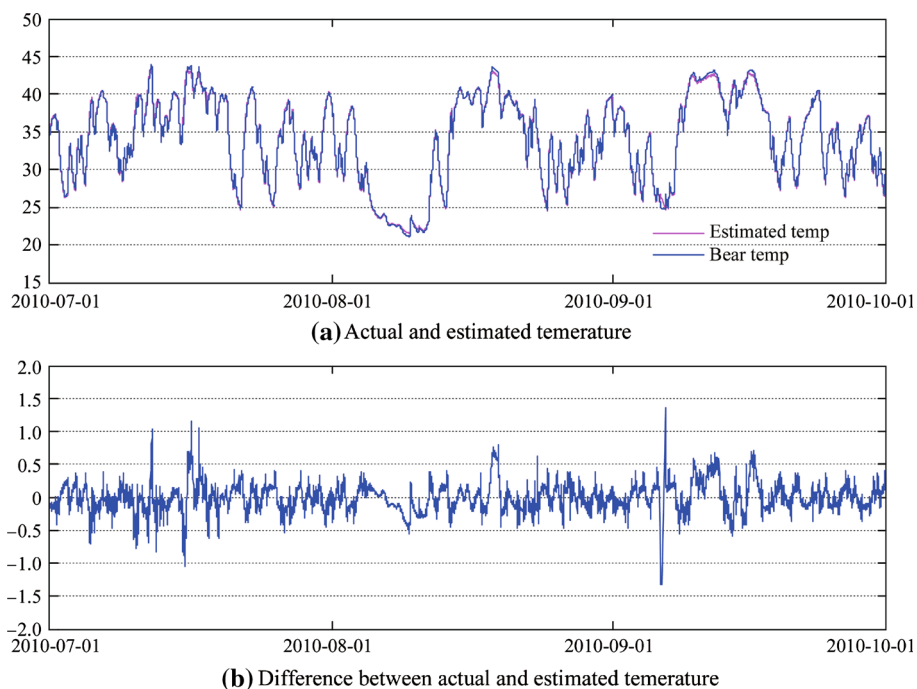


Fig. 7 Rear bearing model output in normal condition of new turbine

differences between actual and estimated temperature are located within $1.5\text{ }^{\circ}\text{C}$ as in Fig. 4, the conclusion can be drawn that the ANN model for rear bearing can be applied for fault detection for other turbines.

Figure 6 shows the three months' SCADA data of a new turbine in same wind farm in normal condition. The data presented in Fig. 6 are also very varied: the turbine speed is varied with starting and stopping; the active power changes from 500 kW to more than 3,000 kW, and the temperatures are also varied. These three months' SCADA data are supposed to be large varied of the parameters in normal condition. Figure 7 shows the results of ANN model using the SCADA data from the new turbine. The estimated rear bearing temperature is very close to the actual value. The maximum difference between estimated and actual temperatures is less than $1.5\text{ }^{\circ}\text{C}$ which is an early warning level as shown in Fig. 5. This means that the new turbine is in normal condition. Therefore, the ANN model of rear bearing using SCADA data from one turbine can be applied in fault detection for other turbines in same type.

7 Conclusions

This paper has described ANN technique for early fault identification for the main components of wind turbines especially for the bearings based on the existing SADA data collected by commercial SCADA system. The result shows that it can deal with large volume SCADA data, give the operators of wind farm very early warning and close alarm to

help them make the right maintenance schedule and action decision in advance. In this way, the information presented to the operator is dramatically reduced without omitting useful information. The maintenance and operation cost can also be reduced by optimizing the maintenance plan, staff and preparation of tools according to the early warning and alarm.

This paper only established a normal behaviour ANN model for one component, i.e., main shaft rear bearing. In the future, more components are needed to establish normal behaviour model. In this paper, the ANN model established by one turbine is tested in new one only in normal condition, and in the future, the test shall be done in different conditions containing normal level, warning level and alarm level.

Acknowledgments This paper is supported by the research project Windsense, Add-on Instrumentation System for Wind Turbines, which is financially supported by the project participants and Research Council of Norway. Thanks also give to Nord-Trøndelag Elektrisitetsverk for supporting and providing SCADA data from wind farm of Hundhammerfjellet.

References

1. Global Wind Energy Council (2013) Global wind statistics 2012, pp 1–4
2. Blanco MI (2009) The economics of wind energy. *Renew Sustain Energy Rev* 13(6–7):1372–1382
3. Pinar Pérez JM, García Márquez FP, Tobias A et al (2013) Wind turbine reliability analysis. *Renew Sustain Energy Rev* 23: 463–472

4. Becker E, Poste P (2006) Keeping the condition monitoring of wind turbine gears. *Wind Energy* 7(2):26–32
5. Laouti N, Sheibat-Othman N, Othman S (2011) Support vector machines for fault detection in wind turbines. In: The 18th IFAC world congress, Milan, Italy, pp 7067–7072
6. Wang K (2005) Applied computational intelligence in intelligent manufacturing systems. Advanced Knowledge International Pty Ltd, Australia
7. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115–133
8. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) *Parallel distributed processing: explorations in the microstructure of cognition*. MIT Press, Cambridge, pp 318–362
9. Verma A, Kusiak A (2012) Fault monitoring of wind turbine generator brushes: a data-mining approach. *J Sol Energy Eng*, doi:[10.1115/1.4005624](https://doi.org/10.1115/1.4005624)
10. Hansen MOL (2007) *Aerodynamics of wind turbines*. 2nd edn. Earthscan, London
11. Zaher A, McArthur SDJ, Infield DG et al (2009) Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy* 12(6):574–593
12. Garcia MC, Sanz-Bobi MA, del Pico J (2006) SIMAP: intelligent system for predictive maintenance application to the health condition monitoring of a wind turbine gearbox. *Comput Ind* 57(6):552–568