Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection

J Tautz-Weinert and S J Watson

Centre for Renewable Energy Systems Technology, Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, LE11 3TU, UK

j.tautz-weinert@lboro.ac.uk

Abstract. Effective condition monitoring techniques for wind turbines are needed to improve maintenance processes and reduce operational costs. Normal behaviour modelling of temperatures with information from other sensors can help to detect wear processes in drive trains. In a case study, modelling of bearing and generator temperatures is investigated with operational data from the SCADA systems of more than 100 turbines. The focus is here on automated training and testing on a farm level to enable an on-line system, which will detect failures without human interpretation. Modelling based on linear combinations, artificial neural networks, adaptive neuro-fuzzy inference systems, support vector machines and Gaussian process regression is compared. The selection of suitable modelling inputs is discussed with cross-correlation analyses and a sensitivity study, which reveals that the investigated modelling techniques react in different ways to an increased number of inputs. The case study highlights advantages of modelling with linear combinations and artificial neural networks in a feed-forward configuration.

1. Introduction

Onshore wind turbines are now able to compete with fossil fuel powered plants in terms of the levelised cost of energy achieving 74 EUR/MWh [1]. But unscheduled maintenance, particularly offshore, results in high costs as accessibility is restricted by weather and availability of vessels. Studies of recent offshore projects reported operation and maintenance costs of 40-44 EUR/MWh [2]. Advanced maintenance strategies based on actual condition rather than using corrective or preventive maintenance can reduce these costs. Evaluation of operational data recorded by the Supervisory Control And Data Acquisition (SCADA) system of a wind turbine shows promise for the purposes of condition monitoring as the high cost of additional sensors in a common dedicated condition monitoring system is avoided.

Increased temperatures in bearings or the gearbox can indicate reduced performance or imminent failure as mechanical faults are usually accompanied by increased heat loss [3]. Thresholds of absolute values are generally implemented in control systems to avoid overheating. But wear-related changes in the temperature trends are often hidden by normal operational fluctuations in temperature due to the variable speed nature of modern large-scale wind turbines as shown for a simulated fault in figure 1. Some of the first approaches of condition monitoring using SCADA temperature data used manual trending against power [3–5] or clustering [6,7] to find anomalies. These techniques succeeded for single turbines in historic analyses, but feasible detection in real time is difficult due to the required manual interpretation of results. Another recent approach is normal behaviour modelling, i.e. the prediction of a temperature while assuming that the component is behaving normally [8–17]. This approach appears to be more suitable for automated failure detection due to an easily interpretable indicator, i.e. the residual of measured minus modelled temperature.

In this paper, different approaches for normal behaviour modelling are investigated using historic SCADA data. Extensive tests are conducted to gain not only the most accurate temperature prediction for a single turbine and modelling target, but also the average prediction performance and the robustness of each approach using automated training and testing. Two different drive train temperatures are modelled for more than 100 turbines in a wind farm.

In section 2 of this paper the methodology is presented. Section 3 provides details of the case study. The modelling results are discussed in section 4. The final section summarises the findings and addresses future work.

2. Methodology

In this section the idea of normal behaviour modelling is explained and the settings of the different modelling approaches described. Cross-correlation and performance metrics are introduced for input selection and prediction evaluation, respectively.

2.1. Normal behaviour modelling

Temperature signals in recorded SCADA data can give information about the changing performance of mechanical parts. Temperatures of drive trains fluctuate due to the rapidly changing operation of variable speed turbines as shown in figure 1. Normal behaviour modelling is a way to reveal hidden trends in temperature signals. This type of model can be used to estimate temperature using information from sensors external to the component being monitored. Figure 2 shows the idea of modelling a measured variable by using environmental signals (e.g. ambient temperatures, wind speed etc.) and process parameters (e.g. rotational speeds, other temperatures) as inputs to predict the target temperature. The model learns normal behaviour by training with input and desired output data under healthy conditions. After training, the residual of measured minus modelled temperature acts as a potential indicator of failure: if a fault occurs, the residual will increase. An alarm can be raised if a fixed threshold or confidence band is violated [8,12–14], if a Mahalanobis distance considering temperature and residual distributions from training exceeds a probability threshold [16] or based on an abnormal level index which weights residuals according to their probabilities [17]. Generating warnings on the basis of residuals of one or more days of data has been proposed to provide more confidence in alarms [11,13,17]. Further evaluation of alarms with fuzzy inference systems has been applied [8,13,15,17].

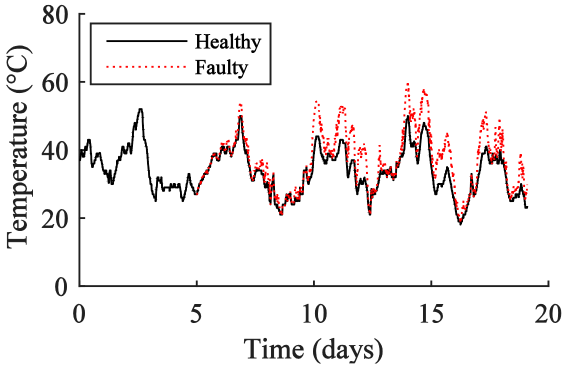

Figure 1. Example of a bearing temperature time series. A time series of a simulated fault is added for visualisation of the detection difficulty.

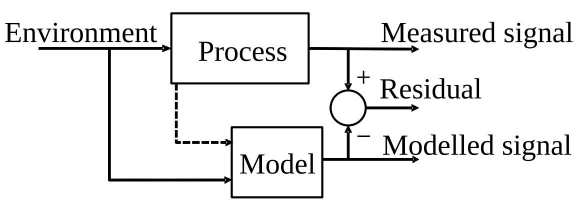

Figure 2. Normal behaviour modelling sketch.

The quality of the failure indicator depends on the accuracy of the modelling. In this study, modelling based on linear combinations ([9,11,14,15]) is compared with artificial neural networks ([8,11,12,15–17]) and adaptive neuro-fuzzy inference systems ([13]). Two novel techniques for modelling of SCADA temperatures are added: support vector machine regression and Gaussian process regression.

Autoregressive modelling approaches and state estimation techniques ([18]) are not considered due to their more likely adaption to new behaviour in the case of a failure as a result of using the target temperature itself for prediction. This is not necessarily desirable for condition monitoring when changes in a physical state which may be an indicator of failure need to be detected through an increased residual. Due to this reason all approaches are here applied in a strictly non-autoregressive way without using any historic values of the target signal (in contrast to approaches in [8–10,16,17]).

2.2. *Cross-correlation*

The sample cross-correlation (CC) gives a measure of the similarity of two signals and can be used as a basis for selecting suitable inputs for modelling. CC at lag k is defined for two (real-valued) signals x and y as:

$$\mathrm{CC}(k) = \sum\nolimits_i (x_{i+k} - \bar{x})(y_i - \bar{y}) \tag{1}$$

with an over-bar denoting the mean value. The summation uses all possible samples for the particular lag of interest. Usually CC is normalised to a value of one for auto-correlation at lag zero, i.e. the correlation of the target with itself at zero lag.

2.3. *Model approaches*

Linear modelling is conducted with a least squares fit of first order polynomials. Simple linear modelling (LIN) uses a linear regression model consisting of a sum of all inputs with individual weights and an interceptor. Linear modelling with interactions (LIN-I) uses a model with intercept, linear terms and products of pairs of inputs (without squared terms).

Artificial neural networks (ANNs) are applied in two configurations: feed-forward (ANN-FF) and layer recurrent (ANN-LR). ANN-FF describes a network with only connections from inputs and layers to the next layer without any feedback or recurrence and has been used in [11,12] for normal behaviour modelling. In contrast, ANN-LRs have a delayed feedback from layer outputs to the input of the layer and are used as a novel way to include system inertia in SCADA normal behaviour modelling. For both configurations a hyperbolic tangent sigmoid (tansig) transfer function is used for neurons in the hidden layer and a linear transfer function for the output layer. Initial tests resulted in an architecture consisting of one hidden layer with six neurons. ANN-LRs are set up with a delay of two steps for the recurrence. Training of neural networks is conducted by Levenberg-Marquardt backpropagation and the mean squared error as performance function. Convergence criteria are minimum performance gradient of 10^{-7}, 1000 epochs or 6 successive iterations with validation performance failing to decrease. Selected training data are randomly split using 80% for real training and the rest for validation.

Gaussian process regression (GPR) [19] is configured with a squared exponential kernel for the covariance function and a constant basis matrix. Standardisation of inputs is applied. Fitting uses a subset of data points approximation.

Linear epsilon-insensitive Support Vector Machine (SVM) [20] regression is applied with a Gaussian kernel. A Sequential Minimal Optimisation solver is used with training until a feasibility gap of 10^{-3}, a zero gradient or 10^{6} iterations are reached.

Adaptive Neuro-Fuzzy Inference System (ANFIS) modelling is conducted in a similar manner to [13]. Two Gaussian membership functions are associated with each input. A linear membership function is used for the output. Training uses a hybrid algorithm utilising backpropagation and least squares in 20 epochs.

For easier comparison of the modelling accuracy, a 'trivial' modelling approach is added, where the target temperature is set to the mean value of the training period. The prediction is constant and unaffected by any input signals.

2.4. Evaluation metrics

Modelling is conducted in Matlab 2015b on a 64-bit operating four core CPU with 2.8 GHz clock rate and 32 GB memory. The runtime for training and testing of each model is recorded to compare the computational effort of the investigated approaches.

Performances of the modelling approaches are evaluated in terms of the mean absolute error (MAE), the root mean squared error (RMSE), standard deviation of error (STDE) and the Coefficient of Determination (R^2), as defined in equations (2)-(6) with n as the number of samples, y as the measured and $\hat{y}$ as the modelled target temperature. Results of the individual models for all turbines are summarised for the farm by calculating the median of the metric values as a measure for the average performance ignoring extreme outliers. Although outliers will be of interest in the condition monitoring stage, here the focus lies on the average performance of normal behaviour modelling. Residual distributions of all turbines are merged by selecting the median for each bin count.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2}$$

$$\mathrm{RMSE} = \left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)^{\frac{1}{2}} \tag{3}$$

$$\mathrm{STDE} = \ \sigma(\hat{y} - y), \text{ with standard deviation } \sigma(x) \text{ as:} \tag{4}$$

$$\sigma(x) = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{1}{2}} \tag{5}$$

$$R^2 = 1 - \frac{\sigma(\hat{y}-y)^2}{\sigma(y)^2} \tag{6}$$

3. Case study

The different approaches are tested with data from a US wind farm with more than 100 turbines. Six months of SCADA data from variable speed turbines with a rated power of 1.5 MW are analysed. The data consist of temperatures as well environmental and control parameters in 10 minute averages.

Two different drive train temperatures are selected to be modelled: a bearing temperature and a generator winding temperature. Due to a lack of maintenance information, all turbines are assumed to operate normally with only short stops for minor repairs. Distributions of operational status are analysed to exclude turbines with significantly long downtimes. The analysis is carried out by visual interpretation of status codes and distributions of power output. Six turbines are excluded from modelling due to unusually high frequencies of non-operational status codes and downtime.

The investigated SCADA data are not always of a high quality. Unfeasible sensor values are found to occur and temperature records show a non-physically high frequency of discrete whole numbers. Although this is a limitation to achieving good modelling accuracy, the aim of comparing different approaches is not hindered. Pre-processing of data is conducted in terms of applying valid sensor ranges similar to [13]. Detailed investigations of SCADA uncertainties by means of sensitivity studies are left for future research.

3.1. Cross-correlation results

All possible inputs for predicting the two chosen target temperatures are analysed with a cross-correlation (CC) calculation up to a maximum lag of $\pm$ 20 ten-minute time-steps. The results in table 1 indicate that the Bearing A temperature correlates with the other bearing, the ambient and generator temperature. Power, currents, wind and rotational speeds have a delayed impact on the bearing temperature. Blade angles and generator voltages do not correlate with the bearing temperature. A comparison of the maximum CC with the CC without any lag reveals the most significant difference for the ambient temperature with a value of 0.82 for a signal lagging 17 time-steps behind compared to 0.76 for the simultaneous signals. The results for the Generator 1 temperature as the target show that the two generator temperatures are statistically identical. The target temperature is highly correlated with the Bearing B temperature, the power, phase currents and wind and rotational speeds. The ambient temperature has a low cross-correlation value of 0.48. Power, currents and wind speed have a delayed correlation with the generator temperature.

Table 1. Highest normalised cross-correlation (CC) (as a function of lag) with Bearing A (a) and Generator 1 (b) temperature. Median of all turbines for first 7500 samples.

(a)

Signal	CC(best lag)		CC(0)
Bearing A temperature	1.00	(0)	
Bearing B temperature	0.86	(1)	0.86
Ambient temperature	0.82	(-17)	0.76
Generator 1 temperature	0.76	(0)	
Generator 2 temperature	0.76	(0)	
Power	0.45	(-5)	0.42
Phase current A	0.45	(-5)	0.42
Phase current B	0.45	(-5)	0.42
Phase current C	0.45	(-5)	0.43
Wind speed	0.45	(-5)	0.42
Generator speed	0.36	(-3)	0.35
Rotor speed	0.36	(-3)	0.35

(b)

Signal	CC(best lag)		CC(0)
Generator 1 temperature	1.00	(0)	
Generator 2 temperature	1.00	(0)	
Bearing B temperature	0.95	(3)	0.94
Power	0.83	(-2)	0.80
Phase current A	0.83	(-2)	0.80
Phase current C	0.83	(-2)	0.80
Phase current B	0.82	(-2)	0.80
Wind speed	0.81	(-3)	0.79
Bearing A temperature	0.76	(0)	
Generator speed	0.72	(-3)	0.71
Rotor speed	0.72	(-3)	0.71
Ambient temperature	0.48	(-20)	0.39

3.2. Model input sensitivity study

The selection of inputs for the normal behaviour modelling is based on CC results. However, for condition monitoring purposes not only the prediction accuracy is important, but also the visibility of a fault in the residual [13]. Therefore, previous temperature measurements of the target component are excluded from the model as inputs, as these would be affected by any change in condition of the component. As the premise of normal behaviour modelling is that the system is not changing then including previous measurements may mask systematic changes in the residuals.

A basic configuration (1a) is defined by selecting the two strongest signals in the CC as inputs without any lag. Using more inputs and the optimal lag could increase the prediction accuracy. Therefore, a sensitivity study is conducted in which the signal lag is changed and further inputs are added according to their CC value. Table 2 summaries the input selection for the configurations with different inputs and their sub-variations a-c with different lags.

3.3. Training and testing selection

Models are trained with 7,500 samples, which is equivalent to 52 days. A further 10,000 samples (69 days) are used for blind testing of the models. A two-fold cross validation is applied by partitioning of the measured SCADA time series into a training period and a testing period. These are then reversed for a second run. It has to be emphasised that down times and start or stop manoeuvres are not excluded in order to test the robustness of the approaches.

Table 2. Inputs and lags for modelling the different configurations.

Configuration	Target: Bearing A temperature									Target: Generator 1 temperature								
	1			2			3			1			2			3		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Generator 1 temperature (t)	x	x	x	x	x	x	x	x	x									
Ambient temperature (t)	x		x	x		x	x		x									
Ambient temperature (t-17)		x	x		x	x		x	x									
Bearing B temperature (t)										x	x	x	x	x	x	x	x	x
Power (t)				x		x	x		x	x		x	x		x	x		x
Power (t-2)											x	x		x	x		x	x
Power (t-5)					x	x		x	x									
Phase current A (t)							x		x				x		x	x		x
Phase current A (t-2)														x	x		x	x
Phase current A (t-5)								x	x									
Wind speed (t)																x		x
Wind speed (t-3)																	x	x

4. Results
The results of the case study confirmed that modelling of a temperature with information from other sensors results in a time series signal, which reliably follows the transient trends of the measured signal, as shown for an example in figure 3.

4.1. Baseline results
The results of the bearing temperature modelling with the baseline configuration (1a), table 3, indicate that linear, ANN and ANFIS approaches perform each with a similar small error. The best approach cannot be found as different approaches perform differently for each of MAE, RMSE, STDE and R^2 metrics and for the two tests. GPR and SVM techniques, however, do not perform as accurately as the other models. The results in table 4 for the generator temperature modelling show a similar pattern, although the ANN approaches give the least errors for all metrics.

Figure 4 gives an insight into the distribution of model performance across the wind farm. Although ANN-LR modelling results in the lowest minimum and median MAE, more than 20 % of the turbines have a distinctly larger MAE compared with the other approaches. A model with an inferior minimum accuracy, but more constant prediction errors for the whole farm will be preferred for failure detection purposes. Maximum errors in the farm should be interpreted with care, as they could also denote a problem in the particular turbine, since normal operation is not guaranteed.

An analysis of the median distribution of the residual time series for the two tests, as given in figure 5 and figure 6 for the generator temperature, reveals bell shaped distributions with a slightly skewed behaviour. The residuals of nearly all approaches are shifted to negative values for the chronological training and test sequences (test 1), but to positive values for the reversed sequence (test 2). The residual distributions of the linear modelling are skewed in an ambiguous way. If the modelling approaches are compared, it can be noted that linear modelling results in the broadest and ANN-LR in the sharpest peaks. The skewness trends of the residual distributions are reversed for the bearing temperature modelling, i.e. positively skewed for test 1 and negatively skewed for test 2. A seasonal influence is assumed to cause the skewness, which is already visible for the trivial model. However, the effect cannot be explained completely by this hypothesis.

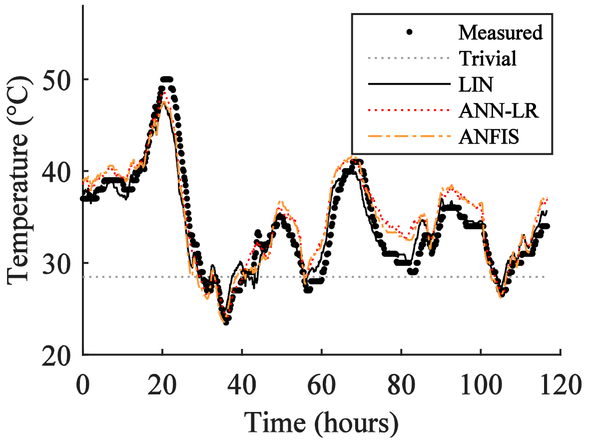

Figure 3. Bearing temperature modelling example.

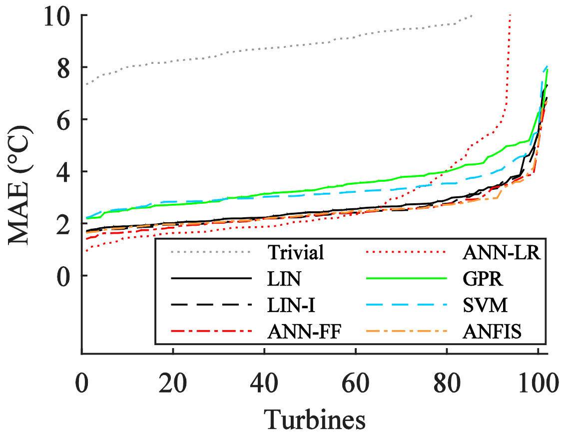

Figure 4. Sorted MAE of all turbines for bearing temperature modelling in configuration 1a, test 1.

Table 3. Performance of different approaches for bearing temperature modelling in basic configuration (1a). Median values are given from all turbines' models.

	MAE (°C)		RMSE (°C)		STDE (°C)		R^2 (-)	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
Trivial	8.92	8.78	10.89	10.62	8.37	7.37	0.00	0.00
LIN	2.43	**2.22**	3.30	**2.93**	2.96	**2.88**	**0.88**	**0.85**
LIN-I	2.30	2.28	**3.16**	3.08	**2.89**	2.99	**0.88**	0.83
ANN-FF	2.32	2.34	3.28	3.24	3.08	3.04	**0.88**	0.84
ANN-LR	**2.15**	2.35	3.34	3.82	3.02	3.52	0.87	0.78
GPR	3.30	3.26	5.28	5.83	4.89	5.32	0.65	0.46
SVM	3.12	3.15	5.02	5.69	4.53	5.32	0.72	0.48
ANFIS	2.36	2.27	3.19	3.01	2.94	2.89	**0.88**	**0.85**

Table 4. Performance of different approaches for generator temperature modelling in basic configuration (1a). Median values are given from all turbines' models.

	MAE (°C)		RMSE (°C)		STDE (°C)		R^2 (-)	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
Trivial	16.02	16.42	20.27	19.81	18.75	18.43	0.00	0.00
LIN	3.56	3.42	4.69	4.57	4.60	4.38	0.94	**0.95**
LIN-I	3.44	3.36	4.58	4.42	4.52	4.21	0.94	**0.95**
ANN-FF	3.04	2.90	4.30	**4.26**	4.19	**4.11**	**0.95**	**0.95**
ANN-LR	**2.49**	**2.38**	**4.25**	4.68	**4.10**	4.55	**0.95**	0.94
GPR	3.16	3.25	4.77	5.25	4.64	5.09	0.94	0.93
SVM	3.40	3.44	5.20	5.91	5.05	5.83	0.93	0.90
ANFIS	3.35	3.30	4.83	4.37	4.68	4.31	0.94	**0.95**

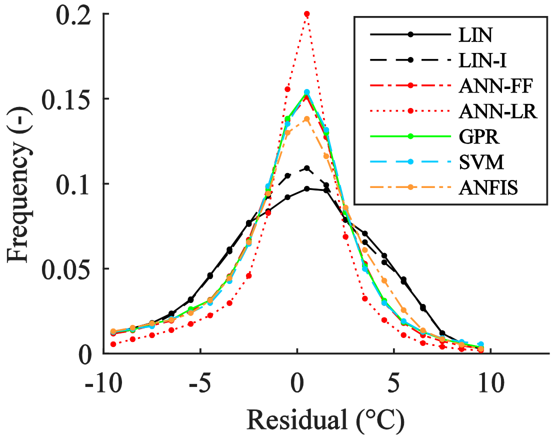

Figure 5. Median residual distribution for generator temperature prediction for test 1.

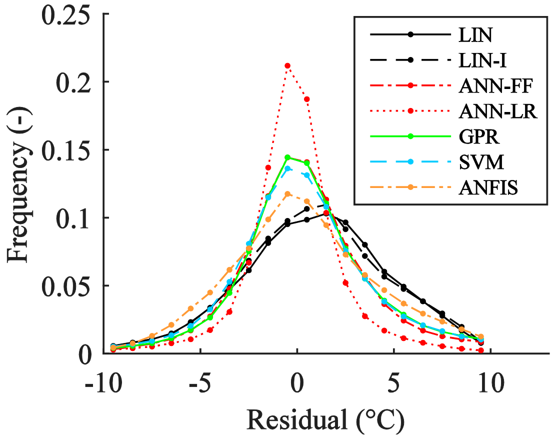

Figure 6. Median residual distribution for generator temperature prediction for test 2.

4.2. Sensitivity to input selection

The RMSE is plotted for the different input configurations in figure 7 and figure 8 for bearing and generator temperature prediction, respectively. For simplification, results from test 1 and 2 are merged by presenting the inferior value from both tests. Using the optimal lag from the CC instead of simultaneous inputs is not beneficial in general. Also, more inputs do not lead to higher accuracy for all approaches. A moderate RMSE reduction trend is visible for linear and ANN modelling approaches if more inputs are used. SVM and ANFIS tend to have larger errors with more inputs. If both simultaneous and lagged inputs are used, the error is only smaller for all results in linear and ANN-FF modelling. There is no clear trend for the other approaches. The optimal setting is found for the RMSE metric in configuration 3c and ANN-FF modelling, although LIN, LIN-I and ANN-LR show very similar accuracy. A detailed comparison for this configuration is given in table 5 and table 6 for bearing and generator temperature prediction, respectively. For the bearing temperature modelling ANN-FF performs best in all median values of the metrics. However, LIN-I shows similar accuracy for the median value and is generally a better performer in terms of the mean values of the metrics, suggesting that it is less affected by outliers. ANN-LR performs well in terms of the median metric, but significantly less well with respect to mean values. The results of the generator temperature modelling show that both ANN approaches perform most accurately here if the median metric values are compared. For ANN-LR the inferior performance in terms of the mean indicates again that the approach performs poorly for some turbines.

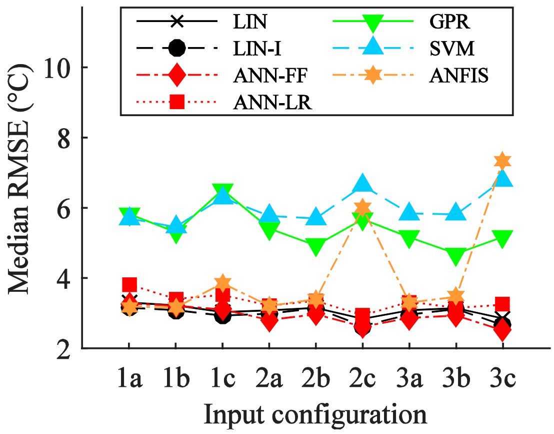

Figure 7. Input sensitivity study for bearing temperature prediction.

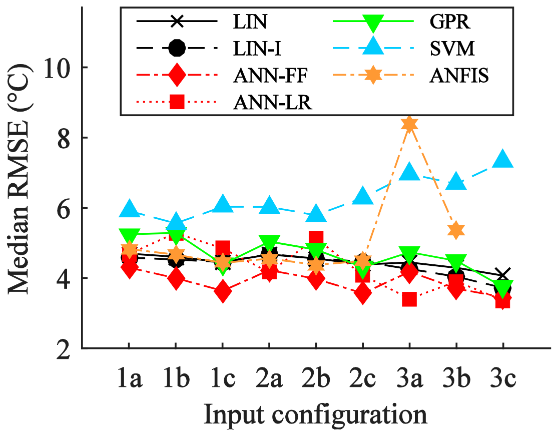

Figure 8. Input sensitivity study for generator temperature prediction. ANFIS modelling was not completed for configuration 3c due to excessive runtimes.

Table 5. Performance for bearing temperature modelling in configuration 3c.

	MAE (°C)		RMSE (°C)		STDE (°C)		R^2 (-)	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
LIN	2.04	2.31	2.85	3.16	2.62	2.90	0.87	0.86
LIN-I	1.84	2.09	2.66	**2.88**	2.48	**2.69**	0.89	**0.87**
ANN-FF	**1.68**	**1.99**	**2.51**	3.00	**2.39**	2.81	**0.90**	0.84
ANN-LR	1.83	2.87	3.24	4.89	2.98	4.34	0.84	0.04
GPR	2.67	2.93	5.18	5.47	4.99	5.09	0.56	0.52

Table 6. Performance for generator temperature modelling in configuration 3c.

	MAE (°C)		RMSE (°C)		STDE (°C)		R^2 (-)	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
LIN	3.19	3.41	4.07	4.37	4.00	4.23	0.96	0.95
LIN-I	2.50	2.80	3.71	4.01	3.64	3.87	0.96	**0.96**
ANN-FF	2.33	**2.61**	3.46	**4.00**	3.29	**3.85**	**0.97**	0.95
ANN-LR	**2.05**	3.45	**3.37**	6.29	**3.22**	5.77	**0.97**	0.47
GPR	2.31	2.64	3.74	4.09	3.60	3.96	0.96	0.95

4.3. Comparison of computational effort

Table 7 gives the training and testing runtimes for the different modelling approaches and different configurations. The computational effort is insignificant for linear modelling. SVM and ANN-FF are trained in about two seconds, but ANN-LR and GPR require 10 and 20 seconds, respectively. The runtime for ANFIS increases significantly with the number of inputs with less than a second for a configuration with 2 inputs and nearly 30 minutes per turbine for 7 inputs.

Table 7. Average runtime for model training and testing per turbine in seconds.

Configuration	Bearing, 1a, test 1	Generator, 1a, test 1	Bearing, 2a, test 2	Bearing, 2b, test 1	Bearing, 2c, test 1	Generator, 3a, test 2	Bearing, 3c, test 1
Inputs	2	2	3	3	5	4	7
LIN	0.02	0.02	0.02	0.02	0.02	0.02	0.02
LIN-I	0.02	0.02	0.02	0.02	0.03	0.02	0.05
ANN-FF	2.36	2.28	2.17	2.26	2.43	2.24	2.61
ANN-LR	11.84	10.14	15.42	10.99	14.38	11.10	18.63
GPR	18.86	19.86	17.42	17.99	17.50	21.05	17.94
SVM	2.49	1.92	2.51	2.52	3.12	2.18	3.54
ANFIS	0.67	0.65	1.60	1.62	32.93	6.41	1702.63

5. Conclusion

Normal behaviour modelling of two wind turbine drive train temperatures has been investigated with modelling approaches based on linear systems, ANNs, ANFIS, SVM and GPR. In a case study with real SCADA data from more than hundred turbines' inputs for modelling were selected following a detailed correlation analysis. All investigated approaches predict the target temperatures with good accuracy. Best results are obtained for linear, ANN and ANFIS modelling in a basic configuration with two input signals. GPR modelling works well for the generator temperature prediction, but less well for the bearing temperature prediction. Modelling with SVM results in distinctly higher errors for both targets. Results of a two-fold cross-validation indicate that there is a seasonal impact in the modelling since the residuals are differently skewed for different training periods. In a sensitivity study, the impact of adding inputs and introducing time lagged signals is investigated. The results indicate that most approaches perform better with more inputs, except for SVM and ANFIS. The computational effort is significant for ANN-LR and GPR independent of the number of inputs and for the ANFIS model if five or more inputs are used. If the different variants of approaches are compared, it can be noted that adding interactions to linear models is beneficial, whereas introducing recurrence in the ANN model seems to be only helpful for some turbines, but leads to inferior performance for others. Adequate input selection with appropriate delay may be a better way to increase the accuracy, although the simple selection of delay based on averaged correlation analysis results does not work in general.

Investigations of filtering out non-operational times (similar to [17]) and a sensitivity study of the training length and number of neurons in an ANN have been the subjects of preliminary investigation, but are not included in this paper due to length limitations. Further research will continue to find suitable approaches for optimal selection of inputs for temperature modelling. In particular ANN-FF and linear modelling will be investigated with more inputs using approaches such as stepwise adding and removing of inputs. Finally, the failure detection capabilities of the normal behaviour models developed in this research need to be tested with real data containing recorded failures. It is intended to compare the approaches with auto-regressive and state estimation techniques to evaluate the differences in failure detection in more detail. Different advanced alarm concepts like the Mahalanobis distance [16] and abnormal level index [17] will be compared.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642108 (Advanced Wind Energy Systems Operation and Maintenance Expertise, http://awesome-h2020.eu/).

References

[1] Bloomberg New Energy Finance 2015 [cited 2016-04-28] *Wind and solar boost cost-competiveness versus fossil fuels* [Internet] available from: http://about.bnef.com/press-releases/wind-solar-boost-cost-competitiveness-versus-fossil-fuels/

[2] Milborrow D 2014 [cited 2016-03-29] *The tide turns on offshore maintenance costs* [Internet] available from: http://www.windpoweroffshore.com/article/1314299/tide-turns-offshore-maintenance-costs

[3] Feng Y, Qiu Y, Crabtree C J, Long H and Tavner P J 2013 Monitoring wind turbine gearboxes *Wind Energy* **16** 728–40

[4] Wiggelinkhuizen E, Verbruggen T, Braam H, Rademakers L, Xiang J and Watson S 2008 Assessment of Condition Monitoring Techniques for Offshore Wind Farms J. *Sol. Energy Eng.* **130** 031004–1 – 031004–9

[5] Yang W, Tavner P J and Court R 2013 An online technique for condition monitoring the induction generators used in wind and marine turbines *Mech. Syst. Signal Pr.* **38** 103–12

[6] Catmull S 2011 Self-Organising Map Based Condition Monitoring of Wind Turbines *EWEA Annu. Conf. 2011*

[7] Kim K, Parthasarathy G, Uluyol O, Foslien W, Sheng S and Fleming P 2011 Use of SCADA data for failure detection in wind turbines *ASME 5th Int. Conf. Energy Sustain.* 2071–9

[8] Garcia M C, Sanz-Bobi M A and del Pico J 2006 SIMAP: Intelligent System for Predictive Maintenance Application to the health condition monitoring of a windturbine gearbox *Comput. Ind.* **57** 552–68

[9] Garlick W G, Dixon R and Watson S J 2009 A Model-based Approach to Wind Turbine Condition Monitoring using SCADA Data *20th Int. Conf. Syst. Eng.*

[10] Zaher A, McArthur S D J, Infield D G and Patel Y 2009 Online wind turbine fault detection through automated SCADA data analysis *Wind Energy* **12** 574–93

[11] Schlechtingen M and Santos I F 2010 Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection *Mech. Syst. Signal Pr.* **25** 1849–75

[12] Kusiak A and Verma A 2012 Analyzing bearing faults in wind turbines: A data-mining approach *Renew. Energy* **48** 110–6

[13] Schlechtingen M, Santos I F and Achiche S 2013 Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description *Appl. Soft Comput.* **13** 259–70

[14] Wilkinson M, Harman K, van Delft T and Darnell B 2014 Comparison of methods for wind turbine condition monitoring with SCADA data *IET Renew. Power Gen.* **8** 390–7

[15] Cross P and Ma X 2015 Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines *Int. J. Autom. Comput.* **12** 25–34

[16] Bangalore P and Tjernberg L B 2015 An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings *IEEE Trans. Smart Grid* **6** 980–7

[17] Sun P, Li J, Wang C and Lei X 2016 A generalized model for wind turbine anomaly identification based on SCADA data *Appl. Energy* **168** 550–67

[18] Wang Y and Infield D 2012 Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring *IET Renew. Power Gen.* **7** 350–8

[19] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (London: MIT Press)

[20] Smola A J and Scholkopf B 2004 A tutorial on support vector regression *Stat. Comput.* **14** 199–222