Journal of Sound and Vibration (2002) **253**(3), 529–544
doi:10.1006/jsvi.2000.3486, available online at http://www.idealibrary.com on IDEAL®

DETECTING LONG-TERM TRENDS IN TURBO-GENERATOR STATOR END-WINDING VIBRATIONS THROUGH NEURAL NETWORK MODELLING

E. M. P. van Wyk and A. J. Hoffman

School for Electrical and Electronic Engineering, Potchefstroom University for CHE, Private Bag X6001, Potchefstroom 2520, *South Africa. E-mail: eeiajh@puknet.puk.ac.za*

(*Received* 8 *May* 2000, *and in final form* 24 *October* 2000)

The accurate assessment of remaining useful life based on condition monitoring variables is not a trivial task, since long-term trends are often obscured by short-term fluctuations. Short-term variations in such variables also tend to overshadow the long-term drift in magnitude. Stator end-winding vibrations are one of the key indicators of the remaining useful life of turbo-driven generators. In this paper, a technique is developed to separate long-term drifts in stator end-winding vibrations from short-term fluctuations. The technique rests on the fact that short-term variations in winding vibrations are largely affected by operational variables measured on a turbo generator, including load and temperature. These dependencies can be captured in a model reflecting the short-term behaviour of the vibration amplitudes. The long-term trend in vibration amplitude is, however, not governed by the same relationships. It is hence possible to extract the long-term trend from the overall behaviour by subtracting the short-term effects of operational variables from the overall behaviour. In this way, a reliable long-term trend is obtained, from which remaining life assessments could be made.

1. INTRODUCTION

Condition monitoring (CM) traditionally means acquiring data from various classes of a plant which gives an indication of the condition of a machine [1]. Data acquired typically include vibration signals, values obtained from oil analysis, temperatures, pressure, etc. These sets of data are processed by various digital signal-processing techniques in order to present them in a more understandable format. Maintenance engineers will make suggestions on the operation of the unit in question based on an examination of these results [2]. This process is an error-prone and lengthy one. In recent years, certain changes have been made to this process; in particular, an attempt was made to automate the decision-making part (that is, to remove the need for experienced people to evaluate the data).

Condition monitoring is an essential element of predictive maintenance. An ideal condition monitoring system would accept measured data as input and will produce the operational status, a possible mode of failure and time to failure as outputs. Since the data used are statistical in nature, no certainties exist, and thus only a *probable* mode of failure and a *predicted* time to failure (PTTF) can be produced.

The work reported in this paper is based on data obtained from the Lethabo Power Station of ESKOM in Viljoensdrif, South Africa. Power is generated by six 650 MW

0022-460X/02/$35.00

GEC-Alsthom turbo-driven generators. Should the stator end-winding (SEW) vibrations on such a machine become out of control the generator can be seriously damaged. A generator hence needs rewinding when the SEW vibrations reach a certain maximum limit. This is a very costly process, and is furthermore accompanied by hidden costs including severe penalties to National Control for loss in production.

SEW vibrations are measured by equipment (originally set up by Alsthom) that includes a narrow band-pass filter around 100 Hz (2 times line-frequency). Values are reported every 5 s to an on-line system called visual automation. This system also captures a large number of other variables reflecting various aspects of the operational status of the generators.

SEW vibrations are currently regarded as the most important condition monitoring variable from which the remaining life of a generator can be assessed. The problem in determining a reliable long-term trend is that these vibrations tend to vary a great deal over the short term. Medium-term variations are also present, including daily and weekly cycles. The variation of interest is the long-term trend extending over several weeks or months. The primary goal of this paper is to propose a method of identifying a long-term trend and give advance warning of impending failure.

The condition monitoring procedure often requires the manipulation of extremely large sets of data. For example, the visual automation system delivers 305 + variables simultaneously. Working with such a large data set not only requires significant computational power, but also complicates the process of identifying underlying causes of undesirable behaviour. The secondary goal of this paper is hence to explore ways of reducing a data set by removing redundant data and noise. As will be shown later, the solution to the secondary problem presents itself naturally in the solution of the primary one.

The methodology used for the previously mentioned goals includes the development of a model for the SEW vibrations. From this a further long-term goal can be derived: to identify the mechanism or mechanisms that are responsible for low-frequency variations in the stator end-winding vibrations.

This paper is structured as follows: section 2 describes the condition monitoring variables used for monitoring the condition of the turbo-generators; section 3 provides a description of the type of neural modelling techniques used in the paper; in section 4 different modelling techniques are applied to the data in order to separate short-term variations from long-term trends; section 5 deals with the issue of redundancy in the data set; section 6 provides conclusions on the topics covered.

2. CONDITION MONITORING VARIABLES FOR TURBO-DRIVEN GENERATORS

2.1. BACKGROUND

Narakesari *et al.* [3] describe a system implemented for condition monitoring at a thermal power plant. This system reads 1400 analogue, 400 dynamic and 100 digital signals. These time-domain data sets are then converted to frequency domain prior to analysis. The system was tested over a 16-month period and showed excellent results in predicting possible failure states; some of these failure predictions were purposely ignored to check its validity.

Monostori [4] describes several condition monitoring techniques, as well as difficulties encountered in previous monitoring systems. These revolve mostly around the fact that previous systems use different data sets to examine different aspects of a machine—usually only in the time domain. Monostori's guidelines regarding requirements for a condition monitoring system include the ability to measure large numbers of digital and analogue signals, using pre-processing methods and employing artificial intelligence to make decisions based on the resulting measurements.

Two of the most frequently used variables in the continuous monitoring of generators or large rotating machines are mentioned in Weiss *et al.* [5]. These are vibration and temperature. Vibration is described as being an indication of the percentage of degradation in performance. Other signals to be monitored include: H_2 and lubrication oil inlet and outlet temperatures, H_2 and oil pressure, fluid levels, metal temperatures, running speed, differential expansion, eccentricity, stator winding temperature, conductivity, active power, reactive power, voltage, current and power factor [6].

When dealing with large rotating machines like generators, vibration analysis can shed light on problems as well as give early indication of impending failures. Vibration is measured by placing an accelerometer inside the machine on selected positions, the problems being that the machine has to be off-line to install these accelerometers and due to high speeds and strenuous conditions they may loosen or go out of calibration. Although vibration frequency trends are still in use on large machines such as generators, most condition monitoring programmes on smaller machines today are based on vibration spectrum analysis. Several different machine problems can be diagnosed from peaks occurring at certain frequencies. A few of them are listed below [7]:

- *Looseness/bearing distress*: unusual number of harmonics of running speed (up to 6 times).
- *Misalignment*: two times running speed.
- *Rotor imbalance*: unusual vibration at running speed.
- *Mechanical looseness*: directional vibration (i.e., vibration only in one direction).
- *Gear problems*: gear mesh frequency peaks (number of teeth times running speed)—these may also appear at lower magnitudes under normal conditions.
- *Blades and vanes*: high fundamental vibration and large number of harmonics near blade/vane passing frequencies (number of blades/vanes times running speed).

Some common but high-risk problems can be solved by detailed vibration analysis. A few common problems are listed below [3]:

- *Thrust problems*: an analysis of metal temperatures and especially rate of change of temperature will give an indication of where the problem originates. Vibration analysis, on the other hand, will give early warning signs of failure.
- *Instability problems*: this is a problem arising from subsynchronous vibration making operation of high-pressure pumps all but impossible. Detailed cascade and spectral analysis permits early detection as well as a criterion for evaluating fixes.
- *Misalignment problems*: misalignment can also be detected by automated vibration analysis. The ability to correlate these analyses with process variables indicates relationships between operating conditions and misalignment.

2.2. VARIABLES USED IN THIS STUDY

The system under investigation is a GEC-Alsthom 618 MW turbo-driven generator. The generators each form part of a totally autonomous plant called a unit. The six units connect to each other only on the power grid. A unit mainly consists of a boiler plant and a turbine plant, the latter containing the generator in question. The generator is physically extremely complex and therefore only a general description can be given.

The stator is the heaviest component of the generator and consists of grain-orientated sheet steel (GOSS). The stator windings are in actual fact semi-hollow bars to allow water to

Figure 1. A real-life stator—notice the stator end-windings.

flow through them. At the physical ends of the stator the windings need to cross over each other; these crossover windings are called the stator end-windings. A picture of these windings is shown in Figure 1.

Over 300 condition and state variables are recorded at each unit, some with rather obscure connections to the generator. These recordings are made on a system called visual automation (VA), which reads these variables every 5 s and stores it for 30 days, after which it is discarded. This causes a major problem to discern long-term trends or extremely low-frequency variations.

The VA system sometimes loses communication with the sensors or even with its storage medium. This causes either stuck bytes or ridiculous values to be stored. To reduce the effect of these error conditions, the sets have to be checked against their standard deviations to remove outliers. Statistical pre-processing of the data included the removal of outliers through visual inspection.

VA data can be exported to Microsoft Excel format. One problem that appears is the fact that the VA can only handle about 720 readings per variable at any one time. This translates to around 1 h if the 5-s sampling rate is used. The system, however, allows lower sampling rates. For this paper two different rates were used, namely 18 min (roughly 9 days' worth of data) and 1 h (exactly 30 days' data).

While the list of possible variables on a generator can number up to 500, the list on VA is only 300 long. The VA set is too long to list here but this set will not always be used; in fact, normally a reduced set of 10 state of operation variables and five condition variables will be used:

2.2.1. *State variables*

- Volts: generator output voltage.
- MVAR: reactive load.
- PF: power factor.
- Current: generator output current.
- Pressure: actual steam pressure.
- Absolute expansion: main turbine absolute expansion.
- Speed: revolutions per minute.
- H_2 inlet temp: hydrogen coolant inlet temperature.
- H_2 outlet temp: hydrogen coolant outlet temperature.
- Coolant temp: coolant tank temperature.

TABLE 1

List of state variables

Name	Description	Correlation	Time lag (samples)
Volts	Stator potential	0·2645	0
MVAR	Megavolt ampere reactive	0·3106	21
PF	Power factor	0·4102	0
Current	Generator current	0·516	0
Pressure	Actual steam pressure	0·3072	153
Exp.	Turbine absolute expansion	0·2930	4
Speed	Turbine speed (r.p.m.)	0·1164	36
H_2 in	Generator H_2 inlet temperature	0·2105	1
H_2 out	Generator H_2 outlet temperature	0·1058	152
Tank	Stator cooling tank temp	0·4746	0

The 10 short-term signals shown above were chosen because they show some similarity with the vibrations when inspected visually. These signals are correlated with the output (vibration) signals to determine whether they can be used to model the short-term fluctuations in the vibrations can be modelled do contain some of the information present in the vibrations, was actually correct.

The peak correlations (amplitude and position) with respect to one of the vibration signals (the others exhibit similar characteristics) are displayed in Table 1. Note that a positive time lag indicates that the vibration follows whatever signal it was correlated with—for example, if MVAR changes, there will be a resulting change in vibration 21 samples later. Also note that although a lag may be present, the zero lag correlation will still be high.

2.2.2. *Condition variables*

Five different vibration signals are measured by accelerometers at various positions inside the stator. The raw output of these accelerometers is the instantaneous acceleration of the particular winding. These measurements are integrated twice to produce the relative displacements. Data may be collected during different stages of operation, including start-up, run-down, normal steady operation or outages. For the case of this study only data collected during steady state operation were used.

Histograms of all the variables were plotted to test for normality (neural techniques usually perform best with approximately normally distributed data). It was found that the short- and the long-term data for both the input and the state variables were close to normal in most cases. No transformation of the data is hence required.

Frequency plots of the data displayed the expected periodic behaviour; a weekly cycle due to reduced electricity consumption over weekends, a daily cycle due to morning and evening peaks, and an hourly cycle since National Control of the power grid make major adjustments every hour.

3. ARCHITECTURE FOR NEURAL NETWORK MODELLING

Architecture refers to the structure of a specific neural network. The simplest architecture is that of the perceptron (a single neuron). The first modification normally made to

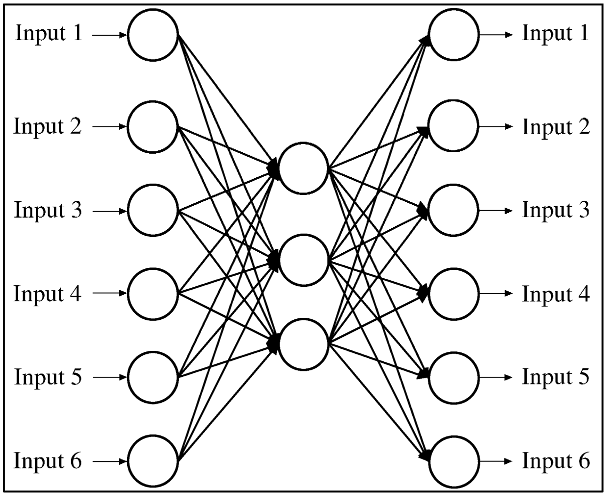

Figure 2. Karhunen–Loeve network.

a perceptron is the addition of another input that is always one; this input is called the bias and its weight is the bias weight. Two basic architectures exist, namely feedforward and feedback/recurrent. Feedforward networks consist of a number of sequential layers characterized by the fact that a layer only receives inputs from a previous layer. Feedback neural networks refer to networks in which neurons can receive inputs from both the previous and the following layers.

In this paper, neural networks are used in two different applications: non-linear regression and data reduction. In the case of non-linear multiple regression a condition variable (vibration) is modelled in terms of a number of state variables. This application may include historic values of the state variables as inputs. For this application a normal feedforward network, typically with sigmoidal activations and trained with back propagation, will be utilized.

The second application involves the reduction of the number of input variables while retaining as much as possible of the variation contained in the inputs (similar to principle component analysis). For this application an auto-associative network, called the Karhunen–Loeve network, as shown in Figure 2, will be used.

A number of neural network techniques have been proposed to solve the PCA problem. Most of these networks consist of a basic feedforward network with a specialized training algorithm. One such network was proposed by Mao *et al.* and is extensively used in engineering [8]. This network calls for a double-train algorithm; that is, for each sample of the data set, the network has to be trained by two different training algorithms. The recirculation or Karhunen–Loeve neural network, in its simplest form, consists of a three-layer network: an input layer, an output layer and one hidden layer. The inputs and training set outputs are the same (both have d neurons). The hidden layer has m neurons, and the network is trained by normal back propagation or other feedforward rules. If the network trains to within certain limits (the outputs should have high correlation with the corresponding inputs) the hidden layer should output a set of data that has the same properties as a PCA set. This set of data contains enough information to regenerate the total set of original data and should be relatively uncorrelated [7].

4. MODELLING THE SHORT- AND LONG-TERM BEHAVIOUR OF END-WINDING VIBRATIONS

The primary objective of condition monitoring is to perform remaining life assessments on equipment. In the case of turbo-generators this important parameter will be reflected mainly by the long-term trend in variables like end-winding vibration amplitudes. It has been mentioned before that short-term fluctuations in the vibration amplitude far exceed the long-term trend in magnitude. It is, however, expected that the short-term behaviour is mostly caused by, or at least correlated with, operational variables like the state variables described in section 2.

The aim of this section is hence to investigate the relationship between end-winding vibrations and the operational variables described in section 2. Based on these relationships models will be extracted for both the short- and the long-term behaviour of vibration amplitudes. The short-term models will be expected to be dominated by behaviour that is not related to long-term trends (should any trend be present over the corresponding time period). The idea is then to isolate the trend component in the long-term model by subtracting the effect of operational variables, as reflected by the short-term model.

4.1. LINEAR REGRESSION MODELS

Linear regression models using operational variables with no time lags as input were fitted to the five vibration amplitudes. The results (as reflected by the correlations between the actual and the modelled variables) are described in the tables, both for the short- and for the long-term data sets. The exercise was repeated for various transformations of the output variables, in order to accentuate either variations close to zero (logarithmic transformation) or far away from zero (exponential transformation). Differential and integral transformations were also investigated to explore different types of potential relationships between the input and output variables.

Tables 2 and 3 display correlations between linear regression models and actual vibrations for the short- and the long-term data sets. In these tables R8, V9, R2, R7 and V6 refer to five different vibration sensors (accelerometers) on the stator end-windings. It is clear that the operational variables have a much superior ability to model the vibrational amplitudes over the short term compared to the longer term. This confirms the expectation that short-term fluctuations in vibration amplitudes are dominated by the effect of operational variables such as temperature, load, voltage and power factor. The fact that these operational variables could not accurately model the long-term behaviour indicates that in this case other factors come into play, affecting the longer term trends in vibrational

TABLE 2

Correlations between linear regression models and actual vibrations for the short-term data set

	R8	V9	R2	R7	V6
No transformation	0·9528	0·7971	0·8989	0·9406	0·9192
Exponential TX	0·9445	0·7960	0·8789	0·9318	0·9062
Logarithmic TX	0·9197	0·7608	0·8757	0·9426	0·9205
Differential TX	0·6939	0·6758	0·4209	0·6940	0·7206
Integral TX	0·8099	0·6035	0·6190	0·8947	0·8521

TABLE 3

Correlations between linear regression models and actual vibrations for the long-term data set

	R8	V9	R2	R7	V6
No transformation	0·2135	0·1558	0·2069	0·2033	0·3231
Exponential TX	0·2152	0·1509	0·1943	0·1917	0·3293
Logarithmic TX	0·1955	0·1483	0·2096	0·2059	0·2828
Differential TX	0·1273	0·1399	0·1398	0·1784	0·1940
Integral TX	0·1788	0·1825	0·1926	0·1700	0·2795

amplitudes, which in turn determine the remaining lifetime. It can also be seen that none of the transformations experimented with resulted in significant improvement of the linear regression modelling accuracies. Such transformations are hence omitted in further analyses.

Whether the accuracy of the linear models could be improved by including time-lagged values of the input variables into the regression model was also investigated. The regression model could then be written as

$$\hat{y}(n) = \sum_{j=1}^{P} \left[\sum_{k=0}^{M-1} h_j(k) x_j(n-k) \right],$$

where $\hat{y}$ indicates the modelled value, $x_j(n-k)$ the input values at time lag k, and $h_j(k)$ the regression coefficients.

This model was tested for filter lengths of up to 100 to determine if a significant increase in modelling accuracy was possible. The results indicated that only a very moderate improvement in accuracy could be obtained. The maximum correlation between the actual and modelled variables was obtained at a filter length of about 40 time lags when using 128-bit data; for longer filter lengths the correlation decreased due to quantization errors. When using 16-bit data the optimal filter length was between 10 and 15 time lags. The improvements obtained were, however, only of the order of about 1%, and hence not considered worthwhile to include. The conclusion is hence that most of the explanatory power is contained in the zero-lag input variables.

To explore further the predictive power of the operational variables for future behaviour of the vibration amplitude, an ARMA predictive filter was designed to predict the vibration amplitude D time periods into the future. The equation for this filter can be described as

$$\sum_{j=1}^{P} \left[\sum_{k=0}^{M-1} h_j(k) r_{x_j x_i}(l,k) \right] = r_{yx}(l+D), \quad i = 1, \dots, P;\ l = 0, \dots, M-1,$$

where P indicates the total number of input variables and M the filter length. The results obtained with this prediction filter for vibration amplitude R7 are described in Table 4, both for the training and for the test sets. SSE represents the sum of squared errors for the respective model.

It can be seen that the ARMA filter possesses some prediction ability, but that it starts to fail for more than about five time periods (approximately 90 min) into the future. This is, however, not long enough to be of real practical use in a condition monitoring context. The alternative approach to predict long-term behaviour, by separating the long-term trend from short-term behaviour, is further explored in the next section.

TABLE 4

Performance of the ARMA prediction filter for different values of D

D	Train set correlation	Train set SSE	Test set correlation	Test set SSE
0	0·9541	0·001782	0·9146	0·00349
1	0·9175	0·003141	0·8221	0·005459
2	0·8777	0·004566	0·7129	0·007521
5	0·7747	0·007964	0·5324	0·009122
10	0·6691	0·01105	0·4578	0·008354
20	0·6517	0·01252	0·139	0·1432

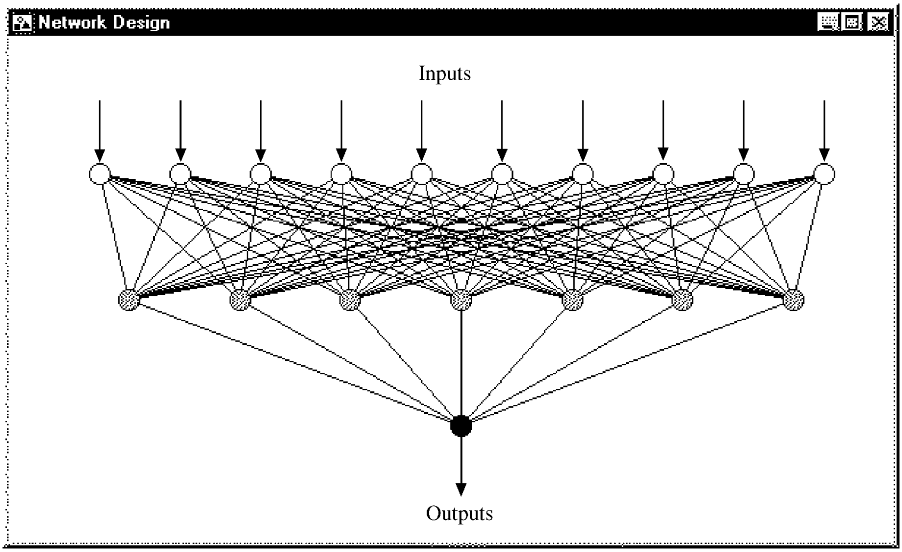

Figure 3. Neural network architecture used for modelling.

4.2. NEURAL NETWORK MODELS

As mentioned earlier, the data set used is actually a combination of three different sets, each set 2 weeks long, that have been filtered to remove outliers and abnormal conditions. The first set was collected in December 1996, the second in January 1997 and the third in May 1997. The reason for this selection of data was to ensure a learning set (the first two sets mentioned above) that did not contain significant long-term trends. The three separate sets, however, span enough time to ensure a noticeable trend over the total period, allowing the third set to be used for demonstrating trend-finding.

To allow a fair comparison with other modelling techniques, the full set will be used (with some random test points) to compare modelling effectiveness. The same output will be used, namely R7. However, when the trend-finding is investigated, two-thirds of the first two sets will be used to train and the rest of those sets used to test the short-term model. This is to ensure that any long-term trend is not trained into the model.

A neural network was set-up and trained using the same short-term data that were used earlier for the ARMA and linear regression techniques. The model consists of a 10-node input layer, a 7-node hidden layer and a single-node output layer (R7), as shown in Figure 3.

The network was trained for 2300 epochs; the resulting correlation between actual and modelled outputs was 0·997, which is far better than that for the ARMA models. The SSE was reduced to 2·3191. Since SSE will be larger for larger sets, the mean SSE will be used from now on. This value is 0·0431 for this particular network. It is also important to note that the training took around 4 min to complete. Although the time is almost four-fold that of the ARMA model, the network would use almost the same time to train five outputs simultaneously.

4.3. DETECTION OF LONG-TERM TRENDS IN END-WINDING VIBRATIONS

The vibration data set mentioned earlier was broken up into two separate sets. The first set consists of the steady data (first 2 weeks). The set was divided into two; the first was used as the steady training set, and the second was used as the steady state model validation set. The second set was used to test the previously trained network. This network should not be able to recognize any long-term trends (as it was trained on steady state data), and the error values should reflect such trends.

By experimenting with the fast propagation coefficient and the learning rate the network trained 16 000 epochs in around 15 min. Correlation (of the normalized set) reported was 0·9825 with an SSE of 0·0267 for the training set while the test set achieved a correlation of 0·9377 and an SSE of 0·0495. From the test set correlation versus epochs, shown in Figure 4, it is apparent that the best training time is about 4000 epochs. The values achieved after 4000 epochs are shown in Table 5.

To display the effectiveness of the modelling, the output of the neural network is shown in Figure 5 along with the targets for the total data set. The first 450 points reflect the training data, the next 450 points the test set for the short-term model, and the last 450 points the set of data that included a long-term trend component. While the modelled output closely

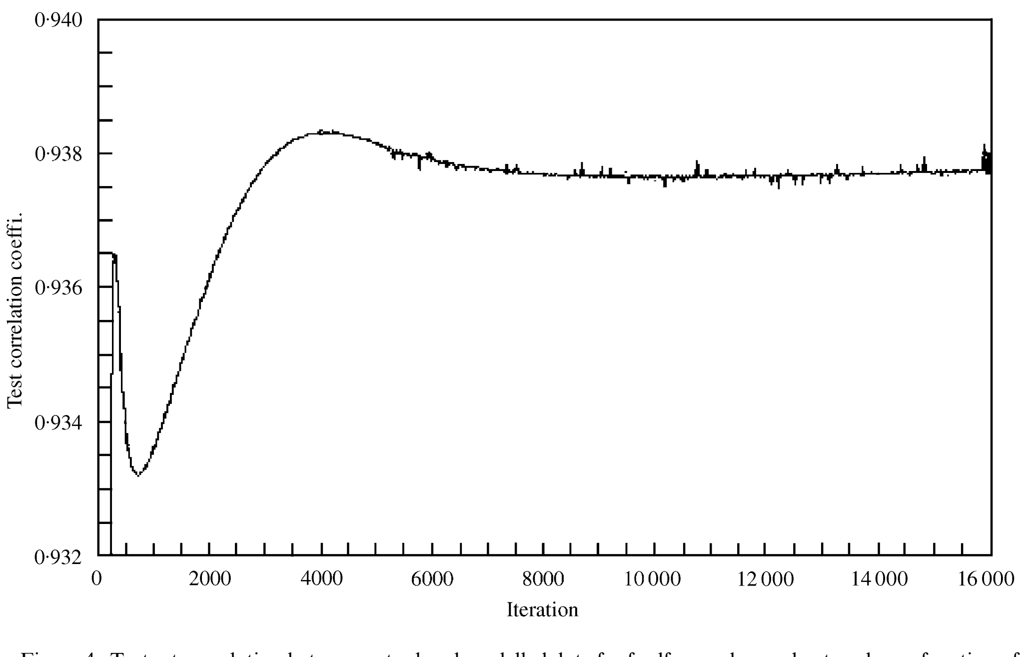

Figure 4. Test set correlation between actual and modelled data for feedforward neural network as a function of the number of epochs.

TABLE 5

Test set correlations between actual and modelled data after training feedforward neural network for 4000 *epochs*

	Training set	Testing set
Correlation	0·98184	0·9383
SSE	0·02722	0·05

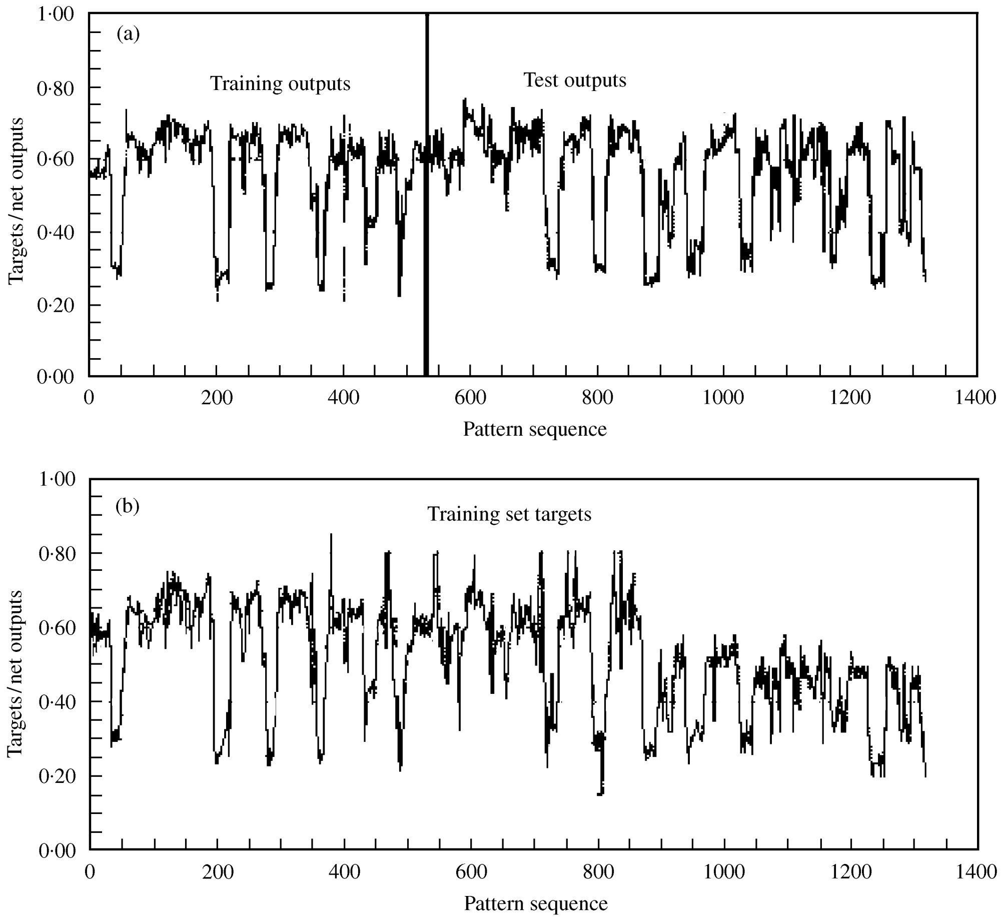

Figure 5. Graphs of training set outputs, test set outputs and target outputs for feedforward neural network.

follows the target over the first test set, the target data over the second test set clearly contains a trend component that was not captured by the short-term model.

The error made by the short-term neural network model (difference between the modelled and target outputs) was investigated to determine if a clear trend was visible. The downward trend is clearly visible as shown in Figure 6, although it is still obscured by some noise. The result is quite satisfactory though, if it is considered that only 10 out of a possible 300

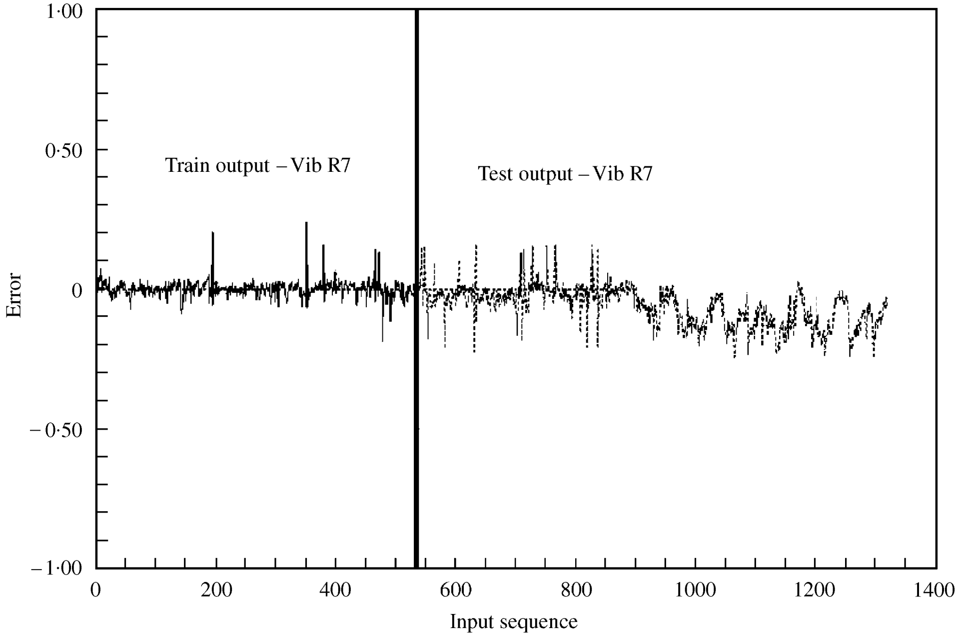

Figure 6. Graph of error between target output and modelled output over training set and test sets feedforward neural network.

variables were used in this model. If a smoother error trend is required, the application of a moving average or low-pass filtering technique may be considered.

5. REDUCING DATA REDUNDANCY

There are two major reasons for removing redundant information from the set of state variables. The first would be that it is difficult to manipulate large sets of variables, and when training neural networks, it is also time consuming. The second reason is the fact that if redundant information is removed the activation of different neurons could be used to ascertain what effect each input has on the output. In other words, if the input set could be transformed to a set of orthogonal inputs, it would be possible to find cause–effect relationships between input and output variables.

The most widely used technique to achieve this objective is that of principle component analysis. As described in section 3, a neural network technique can also be implemented to reduce the number of signals. Although the neural technique does not guarantee orthogonality among the reduced set of signals, it can be used to obtain a reduced set of vectors that have non-linear dependencies on the input set. To ensure a non-linear set, there need to be at least three hidden layers in the network. Since neural network techniques provide more flexibility they were the preferred method for implementing data reduction.

A critical aspect of data reduction is to find the optimal number of signals to retain, corresponding to the number of hidden nodes in the network. The only reliable method is to use trial and error until the smallest set that performs satisfactorily is found. The short data set used earlier (10 variables) was used as inputs and targets for 10 recirculating networks with 1–10 hidden nodes. The correlations between input and output for both the test sets

TABLE 6

Performance of the Karhunen–Loeve network with different numbers of hidden nodes

	Training set		Test set	
Nodes	Correlation	SSE	Correlation	SSE
1	0·8186	0·0924	0·8102	0·0949
2	0·8778	0·0773	0·8863	0·0744
3	0·9161	0·0646	0·9117	0·0663
4	0·9440	0·0529	0·9439	0·0540
5	0·9713	0·0386	0·9702	0·0385
6	0·9870	0·0257	0·9890	0·0242
7	0·9970	0·0126	0·9971	0·0120
8	0·9973	0·0119	0·9969	0·0128
9	0·9979	0·0105	0·9979	0·0105

TABLE 7

Performance of the Karhunen–Loeve network with seven hidden nodes

Volts	MVAR	PF	Current	Press	Exp	Speed	H_2 in	H_2 out	Coolant
0.9969	0·9966	0·9962	0·9954	0·9974	0·9979	0·9970	0·9969	0·9964	0·9953

and the training sets were recorded and are displayed in Table 6. The test was run for 1250 epochs with a learning rate of 0.01 and then for another 1250 epochs with learning rate control.

From the result it was decided that a safe number of reduction (hidden layer) nodes to lose a very small fraction of the variation in the input set is seven. One would, however, expect to also obtain reasonably good results by retaining only five nodes. The 10 outputs were correlated with the inputs using seven hidden nodes to determine the effectiveness of the reduced set to regenerate the inputs. The correlations were as indicated in Table 7.

It is clear that all inputs are generated with a high degree of accuracy. It would be desirable for the resultant signals to have small cross-correlations. When this was investigated it was found that this condition is largely satisfied. The zero-lag cross-correlations between the reduced set of signals are plotted in Figure 7 where the unity autocorrelations appear along the diagonal region. Note that the absolute values were used. Also note that the cross-correlations were not exactly zero; this suggests that further reduction in the number of nodes may still be possible.

The principle components (the outputs of the hidden nodes in the previous network) were then used to model the stator end-winding vibrations. The short-term set was used since the best modelling accuracy was obtained for this set using the complete set of input variables. The network had seven input nodes, four hidden nodes and a single output, set to target vibration R7. The test was run for $\pm$ 20 000 epochs taking less than 8 min (compared to the 15 min for the full set and 16 000 epochs). The test set correlation is shown in Figure 8. The training correlation after 19 765 epochs was 0·981032 and the SSE was 0·027806. For the test set the correlation was 0·941656 and the SSE was 0·047864. From Figure 8 it can be seen that the maximum test set correlation is almost reached after 8000 epochs. The results achieved with the full set of 10 inputs and with the reduced set of seven inputs are compared

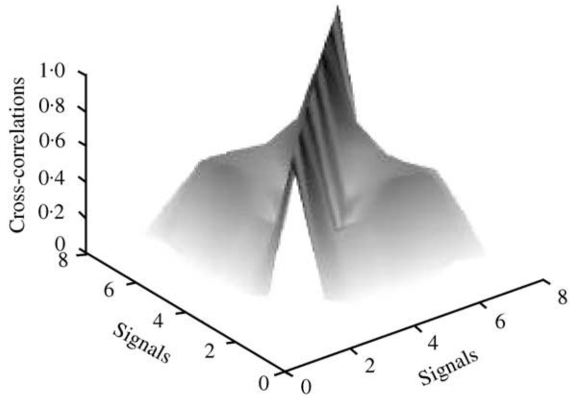

Figure 7. Cross-correlations between the reduced set of variables.

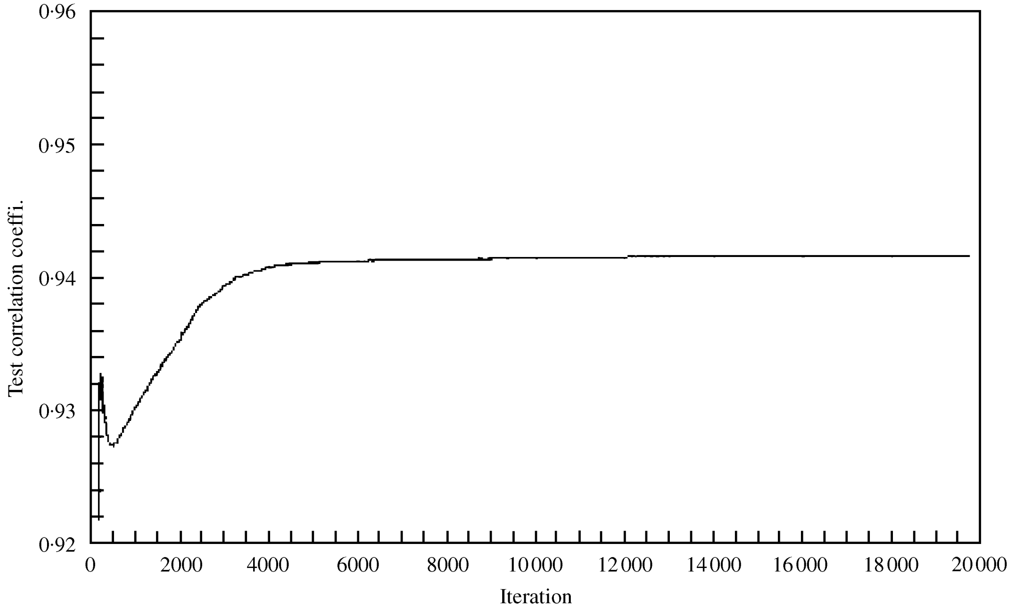

Figure 8. Correlation between target and output values for the test set using a reduced set of seven input variables.

in Table 8. Figure 9 displays the error graph for the long-term data obtained with the reduced set of seven input variables.

Some observations can be made:

- While the training set delivered slightly worse results for the reduced set of inputs, the test set results were slightly improved, indicating better generalization.
- The full set showed a definite peak in test correlation at about 4000 epochs, while the reduced set showed a continued ascent, indicating a lesser inclination to overfit with the reduced set.
- The reduced set trained 8000 epochs in 3 min while the full set trained 4000 epochs in about 7·5 min, indicating the computational advantages when using a reduced set.

Table 8

Comparison between modelling efficiencies of the full set and the reduced set of input signals

	Full set (4000 epochs)		Reduced set (8000 epochs)	
	Training set	Testing set	Training set	Testing set
Correlation	0·98184	0·9383	0·980696	0·940965
SSE	0·02722	0·05	0·028047	0·048528

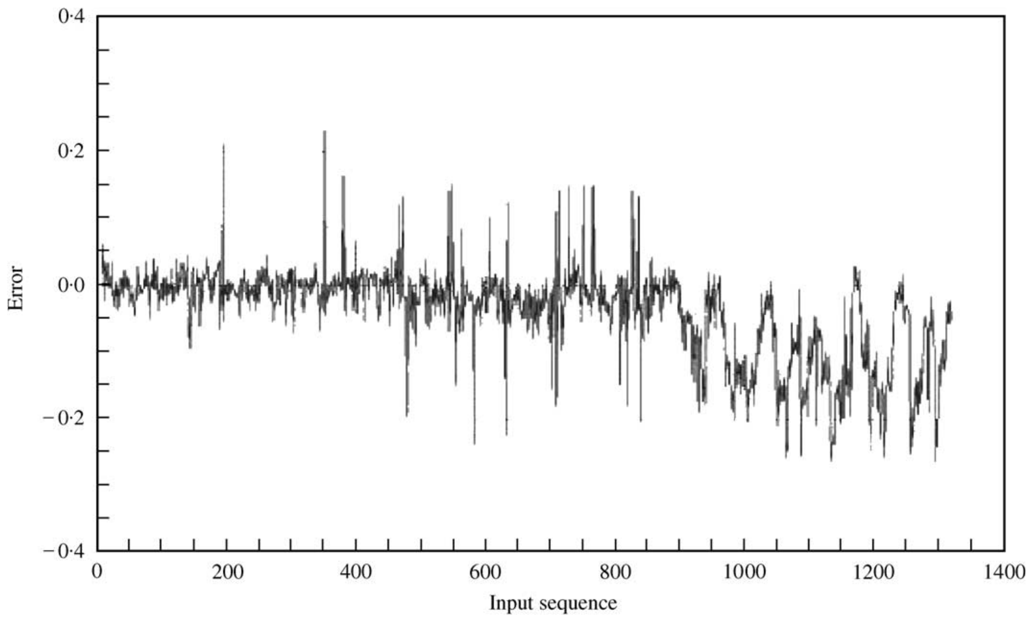

Figure 9. Error graph for the long-term data obtained with the reduced set of seven input variables. Train Output Vib R7

6. CONCLUSIONS

The degree to which the initially set goals have been achieved is briefly summarized below:

- *Identificaion of a long-term trend from short-term fluctuations in order to give advance warning of impending failure*: by developing a short-term neural network model, the long-term trend was found by applying the model to long-term data, and plotting the error signal.
- *Introduce a way of reducing a data set by removing redundant data and noise*: by utilizing a neural network realization of the Karhunen–Loeve transform, a set of principle components was derived. The networks showed great effectiveness with reductions of between 30 and 50%. It provided a fast, intuitive and easy method for removing redundant data and apparently even noise to some degree.
- *Assist in identifying the mechanism or mechanisms that are responsible for low-frequency variations in the stator end-winding vibrations*: the result here was not definite, but rather an indication of possibilities that can be further explored. By examining the connection

strengths from each principle component to each output, the inputs with the most influence could be found. By examining in turn the correlations of these inputs with the original inputs, the input data with the most influence could be found.

It should also be borne in mind that these tests and simulations have all been done with relatively small training sets of data. This was done due to time constrictions as well as computational restrictions and lastly a shortage of long-term data sets. These results could hence definitely be expected to be improved upon in future.

What is, however, of crucial importance is for data storage to be well co-ordinated. Important events which took place at each generator at a specific time should be recorded. This would allow anomalies in the data to be explained. Should these obstacles be overcome it would open the way for neural networks to recognize reliably impending situations and failures.

REFERENCES

1. J. MACINTYRE, P. SMITH, T. HARRIS and A. BRASON 1994 *Engineering Systems Design and Analysis* **64**, 507–512. Neural networks for intelligent machinery diagnostics.
2. R. S. BEEBE 1991 *PWR* **15**, 25–31. Experience with plant condition monitoring at fossil power plants.
3. M. N. NARAKESARI, V. S. DORAI, J. A. ULLAI, P. K. KHASGIWALA, C. B. MEHER-HOMJI, G. MANI, F. J. MEHER-HOMJI, and J. A. WHITTEN 1993 *Joint ASME/IEEE Power Generation Conference, Kansas City, Kansas*, 9p. Design installation and experience with a comprehensive condition monitoring and diagnostic system for a 500MW thermal power plant.
4. L. MONOSTORI 1988 *Robotics and Computer-Integrated Manufacturing*, **4**, 457–464. New trends in machine tool monitoring and diagnostics.
5. J. M. WEISS, R. J. COLSHER, S. M. DIVAKARUNI, and T. NORDBO 1992 *Publication of EPRI*, 41–59—*Paper* #92-0602. Predictive controls for fossil generating stations.
6. S. HIGGINS 1996 *Ph.D. Thesis, University of the Witwatersrand, South Africa.*
7. I. E. ALGUINDIGUE, A. LOSKIEWICZ-BUCZAK, and R. E. UHRIG 1993 *IEEE Transactions on Industrial Electronics* **40**, 209–217. Monitoring and diagnosis of rolling element bearings using artificial neural networks.
8. J. MAO, and A. K. JAIN, 1995 *IEEE Transactions on Neural Networks* **6**, 296–317. Artificial neural networks for feature extraction and multivariate data projection.