

Using abductive machine learning for online vibration monitoring of turbo molecular pumps

R.E. Abdel-Aal* and M. Raashid

Center for Applied Physical Sciences, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Received 26 April 1999

Revised 13 September 1999

Turbo molecular vacuum pumps constitute a critical component in many accelerator installations, where failures can be costly in terms of both money and lost beam time. Catastrophic failures can be averted if prior warning is given through a continuous online monitoring scheme. This paper describes the use of modern machine learning techniques for online monitoring of the pump condition through the measurement and analysis of pump vibrations. Abductive machine learning is used for modeling the pump status as 'good' or 'bad' using both radial and axial vibration signals measured close to the pump bearing. Compared to other statistical methods and neural network techniques, this approach offers faster and highly automated model synthesis, requiring little or no user intervention. Normalized 50-channel spectra derived from the low frequency region (0–10 kHz) of the pump vibration spectra provided data inputs for model development. Models derived by training on only 10 observations predict the correct value of the logical pump status output with 100% accuracy for an evaluation population as large as 500 cases. Radial vibration signals lead to simpler models and smaller errors in the computed value of the status output. Performance is comparable with literature data on a similar diagnosis scheme for compressor valves using neural networks.

Keywords: Vibration monitoring and diagnostics, statistical vibration analysis, turbo molecular pumps, machine learning, abductive networks

*Corresponding author: R.E. Abdel-Aal, P.O. Box 1759, KFUPM, Dhahran 31261, Saudi Arabia. Fax: + 966 3 860 4281; E-mail: radwan@kfupm.edu.sa.

1. Introduction

The 350 kV light ion accelerator facility [4] at King Fahd University of Petroleum and Minerals (KFUPM) employs some 15 Balzers turbo molecular vacuum pumps of various capacities to achieve a minimum vacuum level of 1.33×10^{-4} Pa. Table 1 gives a summary of the specifications and operating conditions for a typical $0.5 \text{ m}^3/\text{s}$ pump, model TPU 510, and its electronic drive unit model TPC 300. Many of the pumps run continuously for extended periods, and operational experience has shown that bearing failures while the pump is running at full speed can completely destroy the pump. Such failures often occur without adequate warning signs that can be detected through routine manual inspection. Even in cases when there is a change in the pump noise, this may go unnoticed in the noisy environment of the accelerator vault or may occur after normal working hours when the facility is left unattended. A turbo pump is an expensive piece of equipment, and pump failures can also be costly in terms of lost beam time if there is a need to wait for in-house repair or for a replacement pump to be ordered from abroad. We have recently initiated work on the development of an online monitoring scheme for the accelerator pumps with the objective of automatically detecting abnormalities in the pump condition and warning the accelerator operator in advance to avert serious failures. The importance of continuous online monitoring for critical machinery is well established [21], since monthly or weekly manual measurements may not be frequent enough or consistent enough to detect developing problems.

Vibration analysis truth tables have been used for many years as a guide for diagnosing vibrations in rotating machinery, but conclusive results often require further evidence [24]. Recent advances in computers, instrumentation, and signal processing techniques have made online predictive vibration monitoring of machinery available and cost-effective approach in many

Table 1

Summary of technical data and operating conditions for the Balzers model TPU 510 turbo molecular pump and its electronic drive unit model TCP 300

Pump:		
Nominal diameter		
Inlet	0.160	m
Outlet	0.040	m
Volume flow rate for nitrogen	0.5	m ³ /s
Compression ratio for nitrogen	10 ⁸	
Theoretical ultimate pressure	10 ⁻¹⁴	bar
Speed		
Rated	1000	s ⁻¹
Standby	666	s ⁻¹
Rated blade passage frequency	40000	s ⁻¹
Run-up time to 90% of rated speed	300	s
Weight	30	kg
Cooling	Air	
Lubricating oil		
	Pfeiffer	
	Turbo oil	
	TL 011	
Permissible ambient temperature	0–35	°C
Electronic Drive Unit TCP 300:		
Input voltage (AC)	110–240	V
Maximum power input	300	VA
Maximum output voltage	40	V
Run-up current	7	A
Rated frequency (±2%)	1000	Hz

situations [21]. Techniques used include time domain and frequency domain analysis as well as combinations of both. Univariate time series analysis [29] and multivariate linear regression methods [19] have been employed to model normal vibration behavior in the time domain. Problems with the first approach include strong nonstationarity of the vibration time series, as in the case of reciprocating machinery. The second technique suffers from difficulties in determining suitable relevant time series that explain variations in the vibration data, as well as strong correlations between the various input time series. The two techniques require complex computations and considerable user intervention for each analysis performed, which makes them difficult to implement online using simple portable apparatus. Frequency domain techniques use the frequency spectrum of the vibration signal as a signature for the pump condition, e.g. [8].

A recent trend in many areas of applied sciences has been to resort to a machine learning approach when a rigorous algorithmic solution becomes too complex or

when the underlying relationships between inputs and outputs are not known. With this approach, a model is developed automatically through training on an adequate number of solved examples. Once the model is synthesized, it can be used to perform fast predictions of outputs corresponding to new cases. A major advantage of this approach in vibration analysis is that intensive computations are now required only once, i.e., during training for model synthesis, rather than being repeated for every vibration record analyzed during actual monitoring. Using the model to process new records becomes a simple and speedy operation that can be implemented in real time using compact and portable apparatus. With the shape of the vibration spectrum as input, the problem is reduced to that of automatic pattern recognition which has been applied in many disciplines, e.g., [28].

Numerous techniques currently exist for the development of machine learning systems [27]. These include statistical pattern recognition methods such as Bayesian classifiers and discriminant functions [11], artificial neural networks modeled roughly on how the human brain is believed to function [26], as well as methods for the induction of decision trees [9]. These techniques vary in their accuracy, complexity, computational requirements during training, and their ability to provide human-like explanations for their conclusions. Such variations have led to newer techniques combining good features from various methods. An example of such 'hybrids' is the AIM abductive network tool [20] which draws on statistical and multiple regression analysis methods as well as neural networks, resulting in a faster and more automated approach to model synthesis. The development of this approach for machine learning through self-organization has followed the track of the group method of data handling (GMDH) algorithm [12] and the closely related adaptive learning network (ALN) technique [7]. A mathematical description of the GMDH-ALN foundations for AIM is given in Section 2. Previous experience with this approach in modeling and forecasting daily minimum temperature [1] has indicated improved prediction accuracy compared to neural network models. Model synthesis is also more automated, requiring little or no user intervention and the resulting models can provide more insight into the phenomenon being modeled. AIM models take the form of simple equations that can be easily implemented on a portable apparatus.

Artificial neural network techniques have been proposed as a new approach to automate vibration analysis, primarily in the frequency domain [5,17,19,23].

One application diagnoses a compressor valve as good or bad by training a back propagation network using a set of features extracted from the acoustic spectra [5]. Kotani et al. [17] reports on the use of an acoustic feature extraction auto-associative neural network followed by a fault discriminating network to diagnose eight types of compressor faults, also using acoustic spectra. Time series of the vibration levels for the rotor of a 500 MW generator have been modeled using radial basis function neural networks [19]. This paper describes the use of GMDH-based abductive machine learning to diagnose a turbo molecular pump as good or bad based on the vibration spectra measured using an accelerometer. Following an overview of the machine learning approach used, the experimental set up is described and models are derived for predicting the pump status through training on a number of known cases. Performance of these models in predicting the pump status from new spectra is described. We compare results obtained using both radial and axial vibration signals from the pump. Results are also related to data in the literature on the performance of similar neural network approaches.

2. GMDH-type modeling techniques

2.1. The classical GMDH approach

In quest for optimal objective models that look only at collected data representing system behavior, Ivakhnenko has proposed the GMDH algorithm in 1966 [15]. He observed that physical models often require information that is not readily available and are therefore subject to many assumptions and simplifications which degrade the quality of the resulting model. Other techniques for quantitative modeling include time-series analysis using various linear statistical methods and multivariate regression [18]. These techniques have difficulties in handling nonlinearities in the modeled phenomena and in dealing with small data sets, which is the case in many environmental, ecological, and social applications [14]. Attempts to incorporate nonlinear relationships in such models require the nonlinearity forms to be presumed a priori, rather than being naturally and automatically derived from the data. Inclusion of postulated nonlinearities in this way also increases the possibility of the model curve-fitting noise in the data [25]. GMDH-type algorithms solve these problems since they automatically determine the inherent structure of complex and highly

nonlinear systems and can synthesize adequate models with relatively few data points. The automation of model synthesis not only lessens the burden on the analyst but also safeguards the model generated from being influenced by human biases and misjudgements.

The GMDH approach is a formalized paradigm for iterated (multi-phase) polynomial regression capable of producing a high-degree polynomial model in effective predictors. The process is ‘evolutionary’ in nature, using initially simple (myopic) regression relationships to derive more accurate representations in the next iteration. To prevent exponential growth and limit model complexity, the algorithm selects relationships which have good predicting powers and discards all the others within each phase. Iteration is stopped when the new generation regression equations start to have poorer prediction performance than those of the previous generation, at which point the model starts to become overspecialized and therefore unlikely to perform well with new data. It is seen that the algorithm has three main elements: representation, selection, and stopping. The algorithm applies abduction heuristics for making decisions concerning some or all of these three aspects. A detailed description of the steps of the classical GMDH algorithm can be found in [12].

To illustrate these steps for the classical GMDH approach, consider an estimation data base of n_e observations (rows) and $m + 1$ columns for m independent variables (x_1, x_2, \dots, x_m) and one dependent variable y . In the first iteration we assume that our predictors are the actual input variables. The initial rough prediction equations are derived by taking each pair of input variables (x_i, x_j ; $i, j = 1, 2, \dots, m$) together with the input variable y and computing the quadratic regression polynomial

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_ix_j. \quad (1)$$

Each of the resulting $m(m - 1)/2$ polynomials is evaluated using data for the pair of x variables used to generate it, thus producing new estimation variables ($z_1, z_2, \dots, z_{m(m-1)/2}$) which would be expected to describe y better than the original variables. The resulting z variables are screened according to some selection criterion and only those having good predicting power are kept. The original GMDH algorithm employs an additional and independent ‘selection set’ of n_s observations for this purpose and uses the regularity selection criterion based on the root mean squared error r_k over the selection data set, where

$$r_k^2 = \frac{\sum_{l=1}^{n_s} (y_l - z_{kl})^2}{\sum_{l=1}^{n_s} y_l^2},$$

$$k = 1, 2, \dots, m(m-1)/2. \quad (2)$$

Only those polynomials (and associated z variables) that have r_k below a prescribed limit are kept and the minimum value, r_{\min} , obtained for r_k is also saved. The selected z variables are more effective in describing and predicting y than the original input variables, and the corresponding columns represent a new data base for repeating the estimation and selection steps in the next iteration to derive a set of higher-level variables. At each iteration r_{\min} is compared with its previous value and the process is continued as long as r_{\min} decreases or until a given complexity is reached. An increasing r_{\min} is an indication of the model becoming overly complex, thus over-fitting the estimation data and performing poorly in predicting the new selection data. Keeping model complexity checked is an important aspect of GMDH algorithms which keep an eye on the final objective of constructing the model, i.e., using it with new data previously unseen during training. The best model for this purpose is that providing the shortest description for the data available [6]. Computationally, the resulting GMDH model can be seen as a layered network of partial quadratic descriptor polynomials, each layer representing the results of an algorithm iteration. The single polynomial in the final layer predicts the independent variable y in two variables, which are themselves quadratics in two lower-level variables, etc. The lowest-level polynomials in the first layer operate directly on the m independent input x variables. Making necessary substitutions for the complete model, we can reach a highly complex polynomial (known as the Ivakhnenko polynomial) of the form:

$$y = a + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} x_i x_j x_k + \dots \quad (3)$$

It is worth pointing out a number of advantages for the GMDH approach compared to conventional regression analysis, particularly for modeling large and complex systems. For the modest case of $m = 10$ and order $p = 8$, obtaining the coefficients of the regres-

sion polynomials directly requires solving 43758 ill-conditioned linear equations simultaneously, while a 3-layer GMDH model obtains the equivalent Ivakhnenko polynomial by repetitively solving regression equations in 6 variables only (Eq. (1)). This reduces ill-conditioning effects (smaller matrices) and allows GMDH to model complex relationships using only a small number of data points, while with conventional regression we need at least as many observations as coefficients. GMDH also keeps generating new variables by intermixing lower-level variables, thus reducing linear dependence [12].

2.2. GMDH variations and the ALN approach

The original GMDH algorithm has been subject to many variations in the methods used for estimating the partial descriptor functions and in the choice of decision rules for descriptor selection and criteria for stopping the iterations. Optimized polynomials as well as different descriptor functions have been proposed. The external regularity criterion used for selection and stopping required the splitting of the training data available into two sets which reduces the amount of data available for estimation. Moreover, results depend on the way the data was divided and some splitting heuristics and cluster analysis were often required [7]. A number of methods have been proposed which operate on the whole training data set for estimation, selection, and stopping, thus avoiding these limitations. GMDH-related activities in the U.S.A. has been generally identified as the adaptive learning network (ALN) approach (AIM being an example), which emphasized the use of the predicted squared error (PSE) criterion for selection and stopping to prevent model overfitting. The PSE criterion minimizes the expected squared error that would be obtained when the network is used for predicting new data which is different from that used during training [7]. For example, AIM expresses this PSE error as

$$\text{PSE} = \text{FSE} + \text{CPM} \left(\frac{2k}{n} \right) \sigma_p^2, \quad (4)$$

where FSE is the fitting squared error for the model on the training data, CPM is a complexity penalty multiplier selected by the user, k is the number of model coefficients, n is the number of samples in the training data set, and σ_p^2 is a prior estimate for the variance for the error obtained with the unknown model. This estimate does not depend on the model being evaluated

and is usually taken as half the variance of the dependent variable y [6]. It is noted that as the model becomes more complex, relative to the size of the training set, the second term increases linearly while the first term decreases. Therefore, the PSE goes through a minimum at the optimum model size which strikes a balance between accuracy and simplicity or exactness and generality. CPM has a default value of 1 in AIM. Lower values allow more complex models while higher values allow simpler ones.

3. AIM abductive machine learning

AIM is a supervised inductive machine-learning tool for automatically synthesizing abductive network models from a data base of input and output values which represent a training set of example situations. Once synthesized by training on a training data set, the network can be queried with new input data to provide the corresponding predicted output. Abductive networks [20] combine the advantages of the neural network approach with those of advanced statistical methods. While the processing elements in neural networks are restricted by the neuron analogy, abductive networks consist of various types of more powerful numerical functional elements based on prediction performance. The network size, element types, connectivity, and coefficients for the optimum model are automatically determined using well-proven optimization criteria, thus reducing the need for user intervention. With neural networks, the user has to experiment with various architectures and there are no hard and fast design rules to determine optimum values for the number of hidden layers, number of neu-

rons in each layer, and the training parameters, and often a number of combinations need to be tried in search of the best solution. With the commonly used standard back propagation algorithm, training times can be huge and there are many training parameters to adjust which may have a major effect on the results [27]. The algorithm is not guaranteed to converge to a good solution [10], and because the method may be unstable and oscillate between solutions, it may not be clear when to stop [27]. This makes abductive networks easier to use and considerably reduces the learning/development time and effort. AIM advantages over back-propagation neural networks in forecasting the daily minimum temperature are demonstrated in [1] while large improvements in training speed are reported in [20]. It should be mentioned, however, that improvements are becoming available which attempt to alleviate some of the problems associated with older standard neural network paradigms. For example, using a stiff ordinary differential equation solver with the back propagation algorithm reduces the number of parameters to be adjusted, cuts on training time, and improves accuracy [27]. Using conjugate gradient methods for unconstrained optimization has been suggested for ensuring guaranteed and faster convergence during training [10]. A step-wise construction procedure has also been proposed for building and training single-layer neural networks without the requirement for the number of neurons in the hidden layer to be fixed a priori by educated guesses [16].

The work reported here used AIM version 1.0 for the Macintosh computer. AIM models take the form of layered feed-forward abductive networks of functional elements (nodes) [3], see Fig. 1. Elements in the first

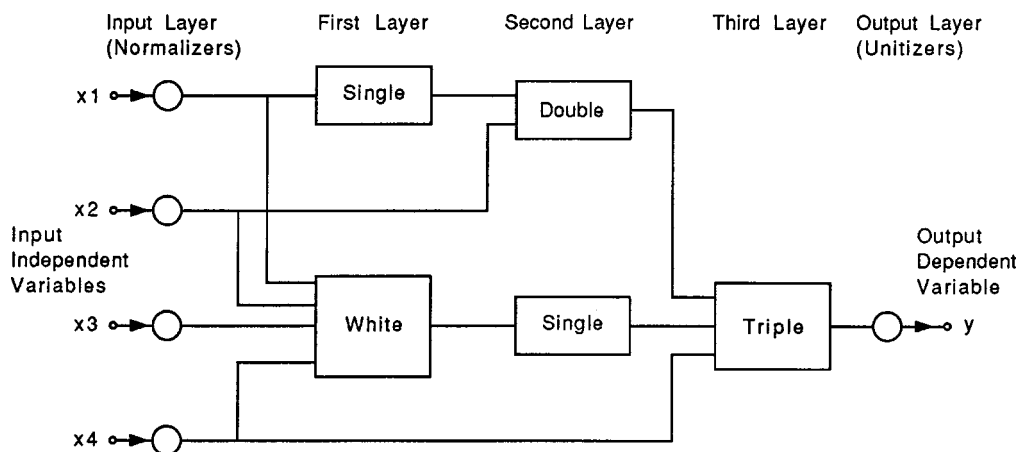


Fig. 1. A typical AIM network structure showing various types of functional elements.

layer operate on various combinations of the independent input variables (x 's) and the single element in the final layer produces the predicted output for the dependent variable y . In addition to the functional elements in the main layers of the network, an input layer of normalizers converts the input variables into an internal representation as Z scores with zero mean and unity variance[3], and an output layer of unitizers restores the results to the original problem space. Both the element type and the combination of inputs to it from all the previous layers are selected automatically for best prediction performance according to the predicted squared error (PSE) criterion[6]. The following main functional elements are proposed:

- (i) A white element which consists of a constant plus the linear weighted sum of all outputs of the previous layer, i.e.,

$$\text{Output}_{\text{White}} = W_0 + W_1 X_1 + W_2 X_2 + W_3 X_3 + \dots + W_n X_n, \quad (5)$$

where X_1, X_2, \dots, X_n are the inputs to the element and W_0, W_1, \dots, W_n are the element weights.

- (ii) Single, double, and triple elements which implement a third-degree polynomial expression with all possible cross-terms for one, two, and three inputs, respectively; for example,

$$\text{Output}_{\text{Double}} = W_0 + W_1 X_1 + W_2 X_2 + W_3 X_1^2 + W_4 X_2^2 + W_5 X_1 X_2 + W_6 X_1^3 + W_7 X_2^3. \quad (6)$$

The first step in solving a problem is preparing a data base of input-output training examples which AIM uses to derive the abductive network model. This model network is synthesized layer by layer until no further improvement in performance is possible or a preset limit on the number of layers is reached. Within each layer, every element is computed and its performance scored for all combinations of allowed inputs. The best network structure, element types and coefficients, and connectivity are all determined automatically by minimizing the PSE criterion. This selects the most accurate model that does not overfit the training data, and therefore strikes a balance between the accuracy of the model in representing the training data and its generality which allows it to fit yet unseen future data. In this way the model is optimized for

the actual use for which it is developed, rather than simply fitting the training data. The user may optionally control this trade-off between accuracy and generality using the complexity penalty multiplier (CPM) parameter[6]. Larger values than the default value of 1 lead to simpler models which are less accurate but are more likely to generalize well with unseen data, while lower values produce more complex networks which overfit the training data and may therefore degrade prediction performance with noise. An 'Evaluate' utility allows evaluation of the resulting abductive network on an independent set of data and generates a report of the results. To obtain good AIM models, the training set should be a good representation of the problem space. The learning task is also simplified by breaking the problem into smaller and more manageable assignments, and by utilizing any human knowledge on parameters relevant to the model in the choice of input variables to be included in the training data base.

4. Experimental setup

Figure 2 shows the experimental setup used for data acquisition and analysis of pump vibration. Vibration signals are detected using a ceramic shear mode accelerometer, PCB Piezoelectronics Inc., model 352B22 [22]. Table 2 gives a summary of the dynamic performance of this transducer. Vibration perpendicular to the flat mounting surface of the sensor are internally converted into shear forces. Sensitivity is low at very low and very high frequencies, with a reasonably flat response (± 3 dB variations) over the frequency range 1 Hz to 16 kHz. Sensitivity increases steadily with higher frequencies up to a mounted resonant frequency exceeding 32 kHz. The accelerometer response drops at frequencies above that resonant frequency. The sensor was adhesively mounted on the surface of the pump being monitored through a thin layer of petro wax. Two positions have been investigated for mounting the acceleration sensor as shown in Fig. 2. In position R the sensor is sensitive to radial vibrations, while in position A the sensor registers axial vibrations (parallel to the pump axis). One of the objectives of this work has been to compare vibration monitoring performance for these two types of vibration signals. AC-coupled signals from the built-in preamplifier of the accelerometer were further magnified in two timing filter amplifiers type ORTEC 474 with no CR or RC timing networks inserted (for a flat frequency response).

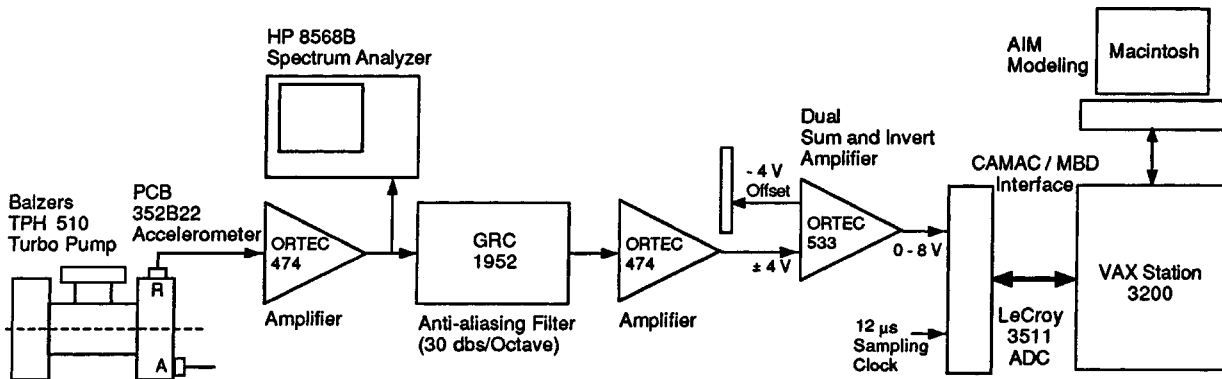


Fig. 2. Experimental setup for the acquisition and analysis of pump vibration signals.

The frequency spectrum of the unfiltered vibration signal was observed on a Hewlett Packard spectrum analyzer model HP 8568B which showed frequency components as high as 75 kHz with the pump running at the full speed of 1000 s^{-1} . It appears that the high frequency response of the accelerometer extends beyond the mounted resonant frequency with adequate sensitivity to allow the detection of vibration signals having such higher frequency components. An anti-aliasing low pass filter with 30 dBs/octave attenuation in the stop band was inserted between the two amplifier stages. The filtered signal had a peak-to-peak amplitude of about 4 V, and a DC offset of -4 V was applied to this signal using an ORTEC dual sum-and-invert amplifier model ORTEC 553. This converts the bipolar signal into a unipolar 0–8 V signal which was sampled by a LeCroy 3511 CAMAC ADC at a sampling frequency of about 84 kHz (sampling interval = $12 \mu\text{s}$). This sampling frequency is about five times the high end of the frequency range where the transducer response is flat within $\pm 3 \text{ dB}$, see Table 2. Cut-off frequency on the anti-aliasing filter was selected to prevent aliasing effects due to the higher frequency components of the vibration signal. Vibration waveform records, each being 4K-samples in length, were acquired by a VAX 3200 workstation running the XSYS data acquisition and analysis package [2,13]. Software in the VAX computed the magnitude of the fast Fourier transform (FFT) of the sampled record and generated data bases for training and evaluating the AIM models. These data were transferred to a Macintosh computer, connected to the VAX as a terminal, for use in the development and evaluation of AIM models for vibration monitoring.

Table 2

Summary of the dynamic performance data for the PCB 352B22 accelerometer

Voltage sensitivity ($\pm 15\%$)	1.02	mV/ms^{-2}
Frequency range ($\pm 5\%$ variations)	5–8000	Hz
Frequency range ($\pm 10\%$ variations)	2.5–9000	Hz
Frequency range ($\pm 3 \text{ dB}$ variations)	1–16000	Hz
Mounted resonant frequency	> 32	kHz
Measurement range	± 3924	ms^{-2} Peak
Resolution (broadband)	0.015	ms^{-2} Peak
Amplitude linearity	± 1	%
Transverse linearity	≤ 5	%

5. Data analysis

Figure 3 shows 500-sample sections of the vibration waveform records from a TPU 510 pump. The records were measured after approximately 5000 hours of operation (when bearing change becomes due) as well as shortly following bearing change. The former condition was used to represent a 'bad' pump, and the latter a 'good' pump. Shown in the figure are waveforms measured in both the radial and axial positions of the accelerometer (positions R and A in Fig. 2, respectively). It is noted that the good pump is generally much quieter, where the vibration signals are lower in amplitude and have predominantly lower frequency components. It is also noted that signals at the radial position of the sensor are relatively richer in frequency components compared to the axial position. Figure 4 shows the complete 4K-point FFT spectrum of the 4K-sample waveform record of the radial vibration signal for the bad pump. Each channel increment represents a frequency interval of $1/(4095 \times 12 \times 10^{-6}) \approx 20.35 \text{ Hz}$. The figure indicates that the majority of the frequency content of the sampled vibration signal lies below 10 kHz. Frequency analysis in this re-

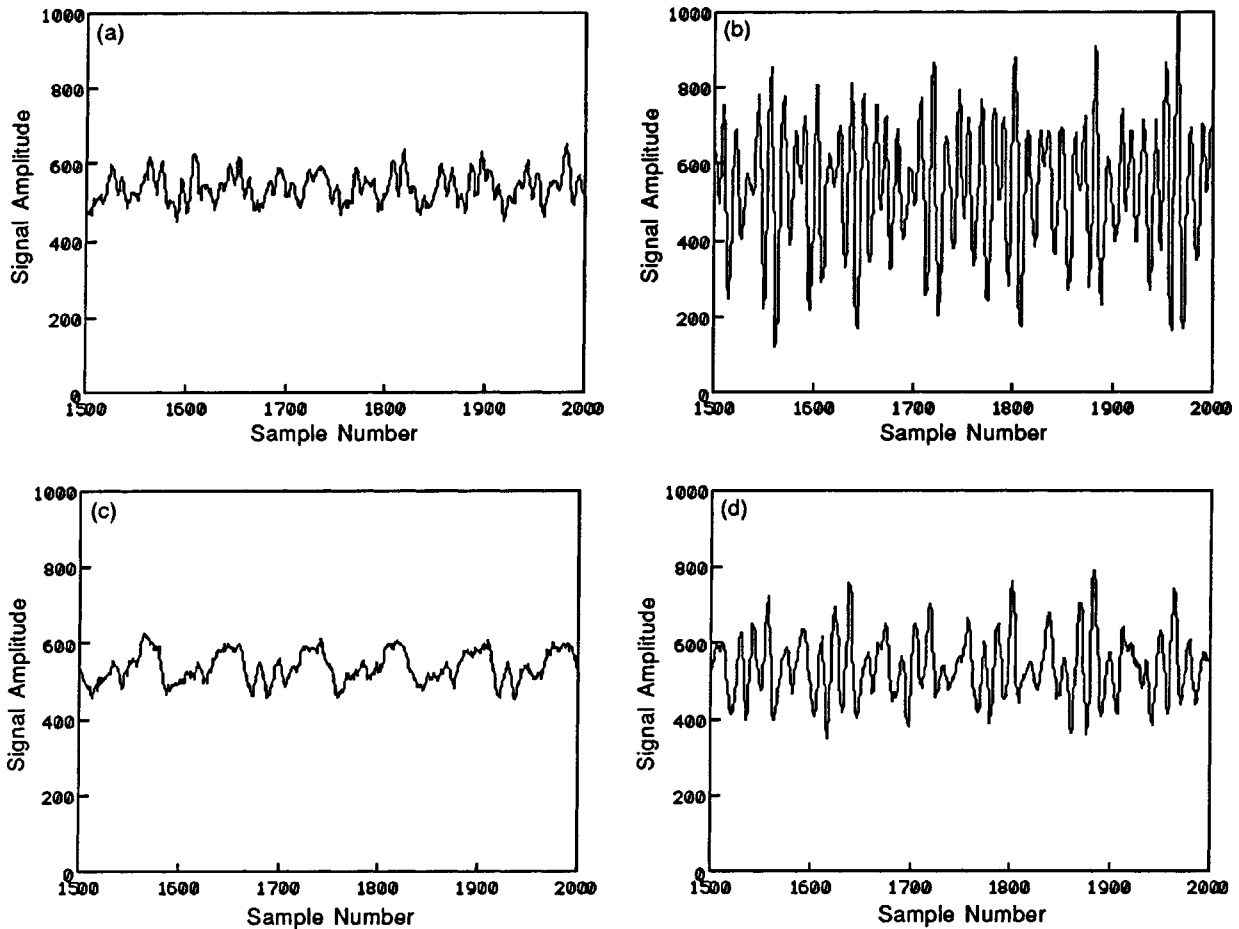


Fig. 3. 500-sample waveform records of the pump vibration signals: (a), (b) at the radial (R) position for the 'good' and 'bad' pump conditions, respectively; (c), (d) at the axial (A) position for the 'good' and 'bad' pump conditions, respectively. Sampling interval = $12 \mu\text{s}$.

gion is acknowledged as the most effective method for detecting imbalance, misalignment, mechanical resonances, and looseness in rotating machinery [21]. This region (channels 1 to 500 of the FFT record) was used throughout as the vibration signature of the pump condition. Figure 5 shows plots of this region of the vibration spectra corresponding to the cases of the four waveform records in Fig. 3. The fundamental component in all cases is about 1 kHz, which is the frequency corresponding to the rated pump speed of 1000 s^{-1} . The plots confirm observations made above on the time records regarding relative density of frequency components for good vs bad pumps as well as for radial vs axial sensor positions.

Data bases for training and evaluation were derived from the 500-channel frequency spectra shown in Fig. 5 through a data reduction procedure, since the version of AIM used allows a maximum of only 50 input parameters for use in model synthesis and evalua-

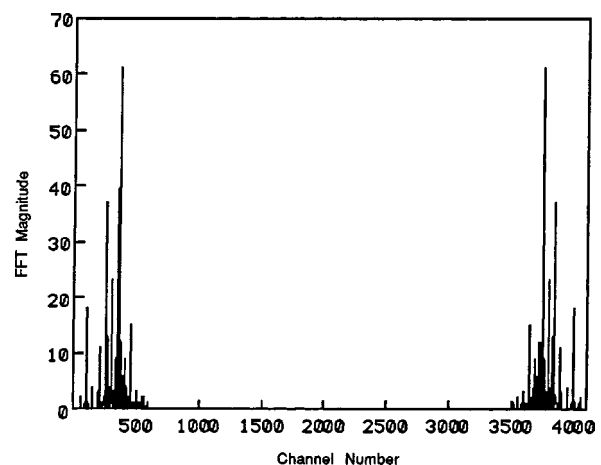


Fig. 4. FFT spectrum for a 4096-sample waveform record for the vibration signal at the radial (R) position and the 'bad' pump condition. 1 channel = 20.35 Hz .

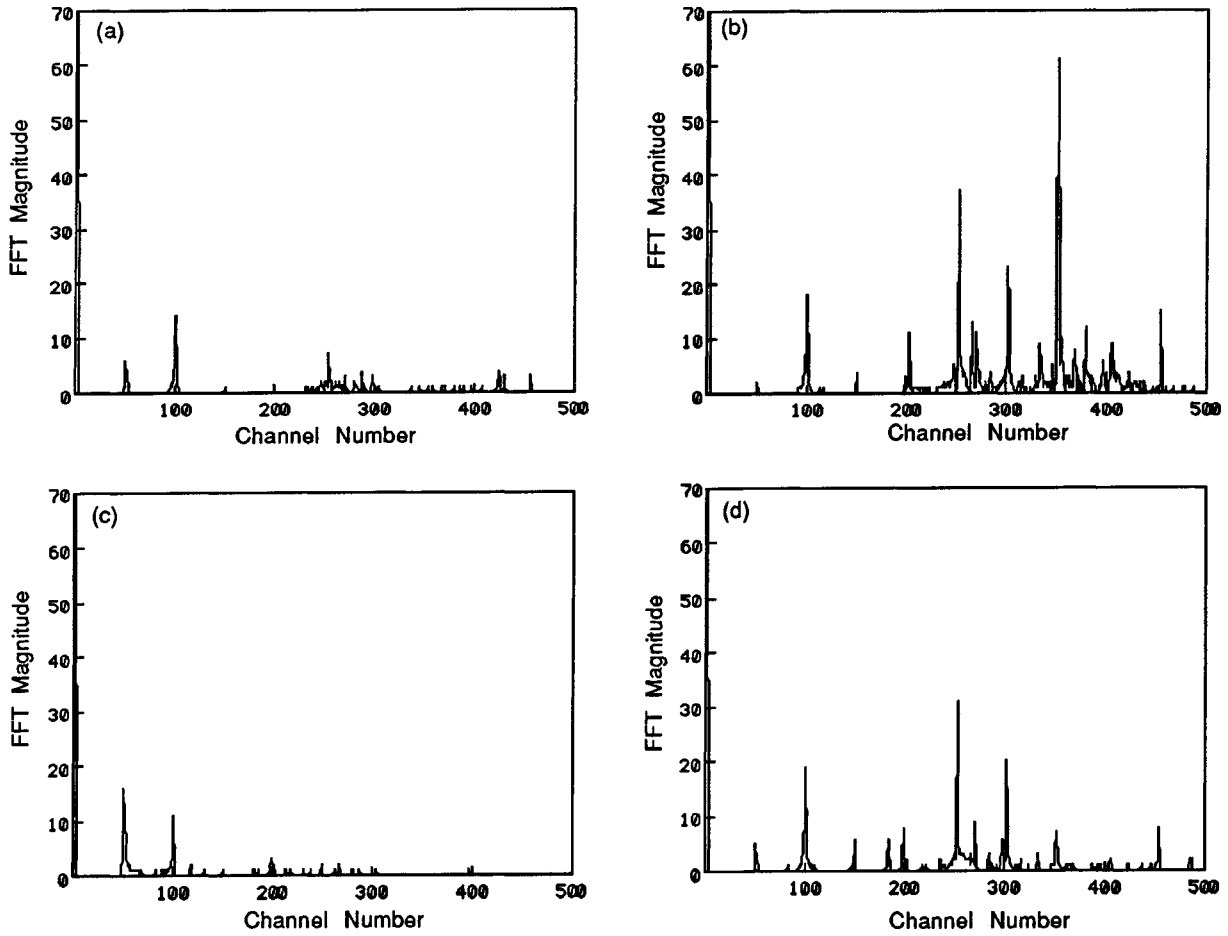


Fig. 5. FFT spectra in the low-frequency region (0–10 kHz), 1 channel = 20.35 Hz: (a), (b) at the radial (R) position for the ‘good’ and ‘bad’ pump conditions, respectively; (c), (d) at the axial (A) position for the ‘good’ and ‘bad’ pump conditions, respectively.

tion. We employed a method adopted by [5] to reduce the 500-channel spectra to 50-channel spectra. With this method, the original 500-channel spectrum S is divided into 50 segments; S_1, S_2, \dots, S_{50} , each consisting of 10 channels. The 50-channel percentage area ratio spectrum R used for AIM training is derived such that its j th channel R_j contains the percentage ratio of the sum of the contents of all channels within segment S_j to the total sum of the contents of all channels in all the 50 segments, which is the total area of the original spectrum S , i.e.,

$$R_j = 100 \times \frac{\sum_{i=1}^{10} S_{ji}}{\sum_{k=1}^{50} \sum_{i=1}^{10} S_{ki}}. \quad (7)$$

In addition to satisfying the data reduction requirement by AIM, this averaging approach is useful in reduc-

ing sensitivity of the resulting models to changes in individual spectrum components which may result, for example, from slight changes in sensor location if the spectrum data were to be applied directly to the AIM network. The method also produces normalized training/evaluation data such that the models are derived and used with input data always in the same number space (range 0–100). Using ratios of the spectral component levels, rather than amplitudes of the spectral components themselves, makes the resulting models more robust against variations in vibration signal levels which are reflected directly in the spectrum amplitude. Therefore, the normalized area ratio spectrum used to derive the training/evaluation data bases provides a spectral signature that contains information on the relative strength of frequency components in the various regions of the spectrum, while being fairly immune to signal amplitude variations including those caused by gain and offset changes.

6. AIM modeling

As a first exercise, we considered the development of abductive network models that continuously perform go-no go checks on the pump condition by classifying the frequency spectrum as representing a good or a bad pump. Figure 6 shows typical 50-channel percentage area ratio spectra for good and bad pumps using signals from both the radial and axial sensor positions. In each position, spectra for both good and bad pumps are shown superimposed on the same plot to indicate the relative ease of separating the two patterns. It is seen that discrimination appears easier with signals at the radial sensor position, and therefore this task should be achieved using simpler models. A typical data base record for modeling the pump status in terms of the percentage area ratio spectrum is given below:

Inputs:	Output:
Channel contents of the percentage area ratio spectrum $Ch_1, Ch_2, Ch_3, \dots, Ch_{50}$	Logical variable Pump status (1 = 'Good', 0 = 'Bad')

Throughout this paper, the numbers of training and evaluation observations will be designated NT and NE, respectively. All models were developed using the default CPM value of 1. Models were synthesized using a training data base of 1000 records (NT = 1000) with both good and bad pump status equally represented (500 records each). Structure of the resulting networks for both the radial and axial sensor positions is shown in Fig. 7. The network for the radial case is simpler, since the 'good' and 'bad' spectrum signatures are easier to differentiate as noted above. The radial network is a single-layer network that uses only the contents of channel 5 to achieve the required discrimination, while discrimination in the axial case requires a 2-layer network operating on two variables (Ch_5 and Ch_{15}). Substituting the equations shown in Fig. 7(a) for the various functional elements produces the following model relationship for the pump status at the radial sensor position:

$$\text{Status}_{\text{Radial, NT}=1000} = -0.081056 + 0.21675(Ch_5) - 0.01076(Ch_5)^2. \quad (8)$$

The corresponding polynomial expression for the axial configuration is much more complex, consisting of 28 terms with powers as high as 9 and 3 for Ch_5

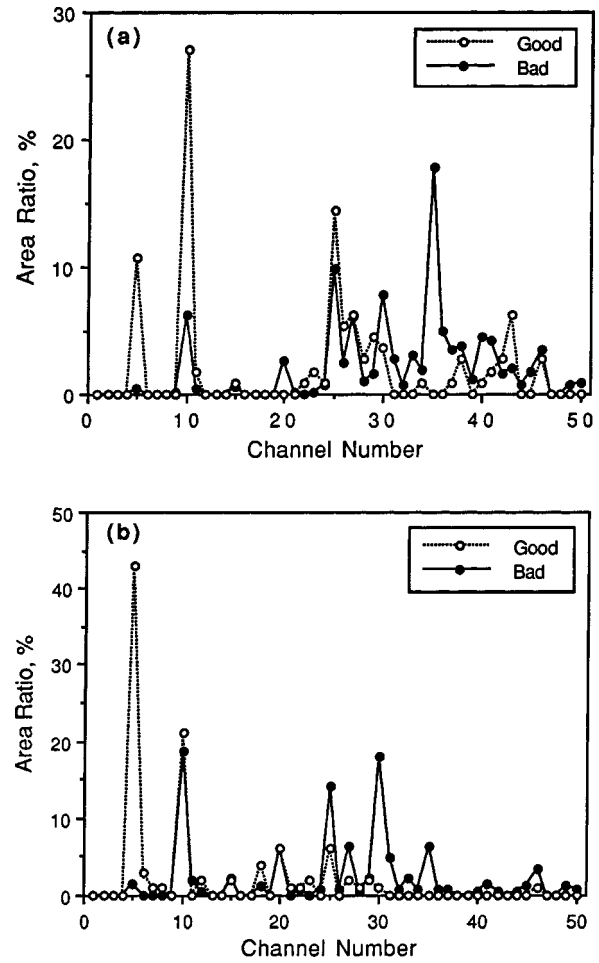


Fig. 6. Percentage area ratio spectra for the 'good' and 'bad' pump conditions used for developing the AIM models: (a) at the radial (R) position, (b) at the axial (A) position.

and Ch_{15} , respectively. Both networks were evaluated on a mixture of 250 'good' and 250 'bad' new cases (NE = 500). The resulting logical value for the pump status (ideally 1 for 'Good' and 0 for 'Bad') was derived from the real number predicted by the network for the status output through simple thresholding at 0.5. Table 3 lists the maximum, average, and standard deviation of the error in the predicted pump status output as well as the good/bad classification accuracy for both sensor positions. The table indicates 100% accuracy for the AIM model in predicting the pump status for the 500 evaluation cases, since the error in the value of the computed status output never exceeds 0.5. It is noted that the maximum error is higher for the axial sensor configuration, although a lower average error is obtained.

Table 3

Summary of errors in the predicted pump status output for two sizes of the training data base, NT = 1000 and NT = 10. In both cases the evaluation data base has 500 cases. Classification accuracy is for the logical status (0, 1) obtained from computed status output by simple thresholding at 0.5

	NT = 1000, NE = 500		NT = 10, NE = 500	
	Radial position	Axial position	Radial position	Axial position
Average absolute error	0.0077	0.0031	0.0602	0.0973
Standard deviation of absolute error	0.0116	0.0082	0.0804	0.1122
Maximum absolute error	0.0532	0.1072	0.3064	0.3583
Classification accuracy, %	100	100	100	100

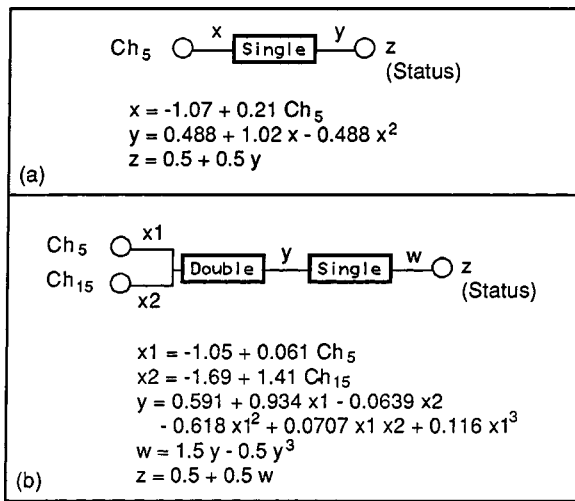


Fig. 7. Structure of the AIM abductive network models obtained for the pump status: (a) using radial vibration signals, (b) using axial vibration signals.

The 100% ‘good’ vs ‘bad’ classification accuracy for the above models suggests that adequate results may still be obtained with much smaller sizes for the training data base, i.e., lower values for NT. It was found that the 100% classification accuracy with the same 500 evaluation cases used above was maintained with NT as small as 10 training cases (5 ‘good’ and 5 ‘bad’). Networks obtained in this case are much simpler, leading to the following linear relationships for the pump status:

$$\text{Status}_{\text{Radial,NT=10}} = -0.03 + 0.095(\text{Ch}_5), \quad (9)$$

$$\text{Status}_{\text{Axial,NT=10}} = -0.05 + 0.0274(\text{Ch}_5). \quad (10)$$

Both models use only the contents of channel 5 to determine the pump status. The simpler models obtained with NT = 10 are attributed to the fact that they need to reconcile much less statistical variations in the input parameters (channels contents of the percentage area

ratio spectrum) during training with NT = 10, as compared with NT = 1000. However, it is expected that such models would be less robust with changes in the training sets than those obtained with larger values of NT. In other words, a different model may be obtained by training on a different set of 10 training spectra. Table 3 lists data on the error in the predicted real value for the pump status output at both sensor positions. The results indicate that, for the same network complexity, axial vibrations are associated with larger errors in the status output.

7. Discussion

GMDH-based abductive machine learning has been used for diagnosing a turbo molecular pump as good or bad as judged by low frequency vibration spectra collected in both the radial and axial directions. 100% diagnosis accuracy for a 500-case evaluation population is maintained for training data bases as small as 10 cases. Similar accuracy has been reported in the literature with neural networks diagnosing a compressor valve [5], but with a larger training data base (NT = 20) and a much smaller evaluation data set (NE = 21). Results indicate that radial vibrations produce simpler and more accurate models in general, due to more distinct signatures for the good and bad cases. AIM provides a fast, convenient, and accurate approach to modeling and classifying the vibration spectra. Adequate models were synthesized automatically with the default value of the CPM parameter without the user having to experiment with various architectures as in the case with presently available neural network tools for which no design methodologies exist. Network training is also expected to be faster with AIM, which considerably reduces development time and effort. Resulting models in the form of a few polynomial equations readily reveal input variables that influence the classi-

fication and allow fast and efficient online monitoring using simple portable apparatus.

The way 'good' and 'bad' pump conditions were defined here may be somewhat idealistic, since 'good' was represented by a pump having a brand new bearing and 'bad' with the bearing approaching the end of its useful life. In practice, a pump would normally be considered satisfactory for a considerable interval following bearing change. This can be taken into account by extending the spectrum measurements for the 'good' pump over such an interval to improve the representation of natural variability within the 'good' class. A similar approach can be applied on the other extreme for the representation of the 'bad' condition. However, representing diverse and truly bad conditions in a unique way would be more difficult. Mayes [19] argues that it is not possible to model the abnormal behavior since it is unknown, and therefore only the normal behavior should be modeled. Diagnosis is then made by looking for clear departures from this model which go beyond acceptable normal variability, thus indicating genuine changes. This approach is possible with time series modeling using either ARIMA [29], regression analysis [19], or neural networks [19]. The first two techniques would be particularly easier to implement, since deviations would show simply as changes in a few model coefficients. However, the categorical classification approach adopted here and in [5] requires that the class type ('good' or 'bad') is known for each training or evaluation observation, which implies modeling abnormal behavior. In situations where a finite number of fault modes can be identified and modeled, the technique can be useful in diagnosing a good unit and classifying the fault type [17]. It is also possible to extend the 2-state classification method described here to model a finite number of pump operating conditions spanning the bearing lifetime as represented by their vibration frequency spectra.

Work described here has utilized existing VAX-based hardware and software designed originally for the acquisition and analysis of nuclear physics experimental data. This has limited the frequency range analyzed due to the limited sampling frequency of 84 kHz. Higher frequency parts of the vibration spectrum, presently excluded by the anti-aliasing filter, may prove useful in identifying problems such as pitting and cracking in bearings, insufficient lubrication, shaft rubbing and pump cavitation [21]. To include such higher frequency components requires sampling at a higher frequency and using an accelerometer with a broader bandwidth. The present sensor may still be

used if proper equalization is included to account for variations in the frequency response over the wider band of interest. Future work would consider faster and more efficient modern PC-based hardware and software platforms for sampling and analyzing the vibration signals. Moreover, more up-to-date versions of the AIM software are now becoming available on the PC. These factors will allow an integrated approach to the acquisition, analysis, and monitoring of the vibration data as well as the modeling of the vibration status for all the pumps of interest. It is worth noting that the newer AIM versions support a larger number of the input parameters, which allows greater resolution for the vibration spectra that can be modeled.

Acknowledgement

This work is part of KFUPM/RI Center for Applied Physical Sciences project supported by the Research Institute of King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

References

- [1] R.E. Abdel-Aal and M.A. Elhadidy, A machine-learning approach to modeling and forecasting the minimum temperature at Dhahran, Saudi Arabia, *Energy – The International Journal* **19** (1994), 739–749.
- [2] R.E. Abdel-Aal, H.A. Al-Juwair and M.A. Al-Sunaidi, A computerized system for beam measurements and control at the KFUPM 350 keV ion accelerator, *Nuclear Instruments and Methods in Physics Research* **A301** (1991), 489–496.
- [3] AIM User's Manual, AbTech Corporation, Charlottesville, VA, USA, 1990.
- [4] H.A. Al-Juwair, G. Blume, R.J. Jaarsma, C.R. Meitzler and K.H. Purser, 350 keV accelerator facility for the University of Petroleum and Minerals, *Nuclear Instruments and Methods in Physics Research* **B24/25** (1987), 810–812.
- [5] K. Arai, H. Shimodaira, Y. Sakaguchi and K. Nakano, Application of neural computation to sound analysis for valve diagnosis, in: *Proc. IEEE Int. Joint Conf. on Neural Networks*, Seattle, WA, July 8–12, 1991, I: pp. 177–182.
- [6] A.R. Barron, Predicted squared error – a criterion for automatic model selection, in: *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S.J. Farlow, ed., Dekker, New York, 1984, pp. 87–103.
- [7] R.L. Barron, A.N. Mucciardi, F.J. Cook, J.N. Craig and A.R. Barron, Adaptive learning networks: developments and applications in the United States related to GMDH, in: *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S.J. Farlow, ed., Dekker, New York, 1984, pp. 25–65.

- [8] J.D. Boyes, Noise and vibration analysis of reciprocating machinery, *Noise and Vibration Monitoring Worldwide* (April/March 1981), 90–92.
- [9] L. Breiman, J. Friedman, R. Olshen and C. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1984.
- [10] C. Charalambous, Conjugate gradient algorithm for efficient training of artificial neural networks, *IEE Proceedings-G* **139** (1992), 301–310.
- [11] R. Duda and P. Hart, *Pattern Recognition and Scene Analysis*, Wiley, New York, 1973.
- [12] S.J. Farlow, The GMDH Algorithm, in: *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S.J. Farlow, ed., Dekker, New York, 1984, pp. 1–24.
- [13] C.R. Gould, L.G. Holzweig, S.E. King, Y.C. Lau, R.V. Poore and N.R. Roberson, The XSYS data acquisition system at the Triangle Universities nuclear laboratory, *IEEE Trans. Nuclear Science* **NS-28** (1981), 3708–3714.
- [14] S. Ikeda, Nonlinear prediction models for river flows and typhoon precipitation by self-organizing methods, in: *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S.J. Farlow, ed., Dekker, New York, 1984, pp. 149–167.
- [15] A.G. Ivakhnenko, Polynomial theory of complex systems, *IEEE Trans. Systems, Man, and Cybernetics* **SMC-1** (1971), 364.
- [16] S. Knerr, L. Personnaz and G. Dreyfus, Single-layer learning revisited: A stepwise procedure for building and training a neural network, in: *Neurocomputing: Algorithms, Architectures and Applications*, F.F. Soulie and J. Hérault, eds, Springer-Verlag, Berlin, 1990, pp. 41–50.
- [17] M. Kotani, H. Matsumoto and T. Kanagawa, Acoustic diagnosis for compressor with hybrid neural network, in: *Proc. IEEE Int. Joint Conf. on Neural Networks*, Seattle, WA, July 8–12, 1991, I: pp. 251–256.
- [18] J.M. Malone II, Regression without models: Directions in search for structure, in: *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S.J. Farlow, ed., Dekker, New York, 1984, pp. 67–86.
- [19] I.W. Mayes, Use of neural networks for on-line vibration monitoring, *Proc. of the Institution of Mechanical Engineers* **208** (1994), 267–274.
- [20] G.J. Montgomery and K.C. Drake, Abductive networks, in: *Proc. SPIE Applications of Artificial Neural Networks Conf.*, Orlando, Florida, 1990, pp. 56–64.
- [21] E.A. Page and C. Berggren, Improve predictive maintenance with HFE monitoring, *Hydrocarbon processing*, January 1991, pp. 69–72.
- [22] PCB Piezoelectronics, Inc. 3425 Walden Avenue, Depew, NY 14043-2495.
- [23] J.P. Peck and J. Burrows, On-line condition monitoring of rotating equipment using neural networks, *ISA Transactions* **33** (1994), 159–164.
- [24] S.J. Richards, Vibration analysis of rotating equipment in mill environment, *Iron and Steel Engineer*, June 1992, pp. 33–38.
- [25] D.E. Scott and C.E. Hutchinson, An application of the GMDH algorithm to economic modeling, in: *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, S.J. Farlow, ed., Dekker, 1984, pp. 67–86.
- [26] P.D. Wasserman, *Neural Computing: Theory and Practice*, Van Nostrand–Reinhold, New York, 1989.
- [27] S.M. Weiss and C.A. Kulikowski, *Computer Systems That Learn.*, Kaufmann, Los Altos, CA, 1991.
- [28] R.A. Vaughan (ed.), *Pattern Recognition and Image Processing in Physics*, SUSSP Publications, Edinburgh, and Adam Hilger, Bristol, 1991.
- [29] Q. Zhuge, Y. Lu and S. Yang, Signature analysis for reciprocating machinery with adaptive signal processing, *Mechanical Systems and Signal Processing* **4** (1990), 355–365.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

