



University of HUDDERSFIELD

University of Huddersfield Repository

Tran, Van Tung, Pham, Hong Thom and Yang, Bo-Suk

An improved hybrid of nonlinear autoregressive with exogenous input and autoregressive moving average for long-term machine state forecasting based on vibration signal

Original Citation

Tran, Van Tung, Pham, Hong Thom and Yang, Bo-Suk (2010) An improved hybrid of nonlinear autoregressive with exogenous input and autoregressive moving average for long-term machine state forecasting based on vibration signal. *Expert Systems With Applications*, 37 (4). pp. 3310-3317. ISSN 0957-4174

This version is available at <http://eprints.hud.ac.uk/id/eprint/16581/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

An improved hybrid of nonlinear autoregressive with exogenous input and autoregressive moving average for long-term machine state forecasting based on vibration signal

Hong Thom Pham, Van Tung Tran, Bo-Suk Yang*

*School of Mechanical Engineering, Pukyong National University,
San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea*

Abstract

This paper presents an improvement of hybrid of nonlinear autoregressive with exogenous input (NARX) and autoregressive moving average (ARMA) for long-term machine state forecasting based on vibration data. In this study, vibration data is considered as a combination of two components which are deterministic data and error. The deterministic component may describe the degradation index of machine, whilst the error component can depict the appearance of uncertain parts. An improved hybrid forecasting model, namely NARX-ARMA model, is carried out to obtain the forecasting results in which NARX network model which is suitable for nonlinear issue is used to forecast the deterministic component and ARMA model are used to predict the error component due to appropriate capability in linear prediction. The final forecasting results are the sum of the results obtained from these single models. The performance of the NARX-ARMA model is then evaluated by using the data of low methane compressor acquired from condition monitoring routine. In order to corroborate the advances of the proposed method, a comparative study of the forecasting results obtained from NARX-ARMA model and traditional models is also carried out. The comparative results show that NARX-ARMA model is outstanding and could be used as a potential tool to machine state forecasting.

Keywords: Autoregressive moving average (ARMA), Nonlinear autoregressive with exogenous input (NARX), Long-term prediction, Machine state forecasting

1. Introduction

Machine state forecasting gradually plays an important role in modern industry due to its ability to foretell the states of machine in the future. This provides the necessary information for system operators to implement the essential actions in order to avoid the catastrophic failures, which lead to a costly maintenance or even human casualties. Moreover, foretelling the states of machine enables maintenance action to be scheduled more effectively, avoids unplanned breakdown, assists maintainers in estimating the remaining useful life, provides alarms before a

fault reaches the critical levels to prevent machinery performance degradation and malfunction [1], etc. Consequently, machine state forecasting has been considerably attracted the attention of researchers in the recent time.

In order to predict the future states of machine, the forecasting model uses the available observations that are generated from measured data by using appropriate signal processing techniques. The measured data could be vibration, acoustic, oil analysis, temperature, pressure, moisture, etc. Among of them, vibration data is commonly used because of the easy-to-measure signals and analysis. Several forecasting models have been successfully proposed in literature in which model-based techniques and data-driven based techniques were commonly utilized. Model-based techniques are applicable to where the accurate mathematical models can be constructed based on the physical fundamentals of a system, whilst data-driven based techniques utilize and require large amount of historical failure data to build a forecasting models that learn the system behavior. Obviously, data-driven based techniques are inaccurate in comparison with model-based techniques in prediction capability. However, data-driven based techniques, which are frequently based on artificial intelligence, can flexibly generate the forecasting models regardless of the complexity of system. Therefore, these techniques that some of those have been proposed in references [1-5] are the first selection of researchers' investigations.

An alternative approach to ameliorate the predicting capability in time-series forecasting is the combination of model-based and data-driven based techniques. According to Zhang [6], the reasons for hybridizing these models are: (i) in practice, it is difficult to determine whether a time-series under study is generated from a linear or nonlinear underlying process or whether one particular method is more effective than the other in out-of-sample forecasting; (ii) data obtained from real-work is purely linear or nonlinear that neither model-based techniques nor data-driven based techniques can be adequate in modeling and forecasting. Model-based techniques can adequately capture the linear component of time series while data-driven based techniques are highly flexible in modeling the nonlinear components. Accordingly, numerous hybrid models have been depicted to provide the investors with more precise prediction. For instance, Zhang [6] combined autoregressive integrated moving average (ARIMA) model and neural network model to forecast three well-known time-series sets that were sunspot data, Canadian lynx data and the British pound/US dollar exchange rate data. Ince and Trafalis [7] proposed a hybrid model including parametric techniques (e.g. ARIMA, vector autoregressive) and nonparametric techniques (e.g. support vector regression, artificial neuron networks) for forecasting the exchange market. A hybrid of ARIMA and support vector machines was successfully presented by Pai et.al [8] for predicting stock prices problems. Other outstanding hybrid approaches could be found in references [9-11]. Most of these hybrid models were implemented as a following process: first, the model-based technique was used to predict the

linear relation, then the data-driven based technique was utilized to forecast the residuals between actual values and predicted results obtained from previous step. The final results were the sum of results gained each model. Furthermore, these hybrid approaches merely regarded as short-term prediction methodology.

In this study, an improved hybrid forecasting model is proposed for long-term prediction the operating states of machine. The prediction strategy used here is recursive which is one of the strategies mentioned in reference [12]. This forecasting model involving nonlinear autoregressive with exogenous input (NARX) [13] and autoregressive moving average (ARMA) [14] is novel in the following aspects: (1) vibration data indicating the state of machine is divided into deterministic component and error component that is the residual between the actual data and deterministic component. NARX and ARMA are simultaneously employed to forecast the former and the latter, respectively. The final forecasting results are the sum of results obtained from single model; (2) long-term forecasting, which is still a difficult and challenging task in time series prediction domain, is applied.

Additionally, the number of observations used as the input for forecasting model, so-called embedding dimension, is the problem often encountered in time series forecasting techniques. Embedding dimension could be estimated by using either Cao's method [15] or false nearest neighbor method (FNN) [16]. However, FNN method depends on the chosen parameters wherein different values lead to different results. Furthermore, FNN method also depends on the number of available observations and is sensitive to additional noise. Cao's method overcomes the shortcomings of the FNN approach and therefore, it is chosen in this study.

2. Background knowledge

2.1. Nonlinear autoregressive model with exogenous inputs (NARX)

The NARX model is an important class of discrete-time nonlinear systems that can be mathematically represented as follows:

$$\begin{aligned} y(t+1) &= f[y(t), y(t-1), \dots, y(t-n_y+1); u(t), u(t-1), \dots, u(t-n_u+1), \mathbf{W}] \\ &= f[\mathbf{y}(n); \mathbf{u}(n); \mathbf{W}] \end{aligned} \quad (1)$$

where $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$ respectively represent the input and output of the model at time t , $n_y \geq 1$ and $n_u \geq 1$ ($n_y \geq n_u$) are the input-memory and output-memory orders, \mathbf{W} is a weights matrix, f is the nonlinear function which should be approximated by using multilayer perceptron. The structure of an NARX network is depicted in Fig.1.

Fig. 1 Structure of an NARX network

Basically, NARX network is trained under one out of two models:

Parallel (P) mode: the output is fed back to the input of the feed-forward neural network as part of the standard NARX architecture:

$$\begin{aligned}\hat{y}(t+1) &= \hat{f}[\mathbf{y}_p(n); \mathbf{u}(n); \mathbf{W}] \\ &= \hat{f}[\hat{y}(t), \hat{y}(t-1), \dots, \hat{y}(t-n_y+1); u(t), u(t+1), \dots, u(t-n_u+1), \mathbf{W}]\end{aligned}\quad (2)$$

Series-Parallel (SP) mode: the output's regressor is formed only by actual values of the system's output:

$$\begin{aligned}\hat{y}(t+1) &= \hat{f}[y_{sp}(t); u(n); \mathbf{W}] \\ &= \hat{f}[y(t), y(t-1), \dots, y(t-n_y+1); u(t), u(t-1), \dots, u(t-n_u+1), \mathbf{W}]\end{aligned}\quad (3)$$

As mentioned above, NARX network inputs include the regressors of inputs and outputs of system while a time series is one or more measured output channels with no measured input. Hence, the forecasting abilities of the NARX network may be limited when applying for time series data without regressor of inputs. In this kind of application, the tapped-delay line over the input signal is eliminated, thus the NARX is reduced to the plain focused time-delay neural network architecture [17]:

$$\hat{y}(t+1) = f[y(t), y(t-1), y(t-n_y+1)] \quad (4)$$

According to [18], a simple strategy based on Takens' embedding theorem was proposed for solving this problem. This strategy allows the computational abilities of the original NARX network to be fully exploited in nonlinear time series prediction tasks and is described as following processes:

Firstly, the input signal regressor, denoted by $u(t)$, is defined by the delay embedding coordinates:

$$u(t) = [x(t), x(t-\tau), x(t-(d_E-1)\tau)] \quad (5)$$

where $d_E = n_u$ is embedding dimension and τ is embedding delay.

Secondly, since the NARX network can be trained in two different modes, the output signal regressor $y(t)$ can be written as follows:

$$y_{sp}(t) = [x(t), x(t-1), \dots, x(t-n_y+1)] \quad (6)$$

$$y_p(t) = [\hat{x}(t), \hat{x}(t-1), \hat{x}(t-n_y+1)]; \quad (7)$$

where the output regressor $y(t)$ for the SP mode in Eq. (6) contains n_y past values of the actual time series, while the output regressor $y(t)$ for the P mode in Eq. (7) contains n_y past values of the estimated time series.

For a suitably trained network, these outputs are estimates of previous values of $x(t+1)$, and should obey the following predictive relationships implemented by the NARX network:

$$\hat{x}(t+1) = f[\mathbf{y}_{sp}(t); \mathbf{u}(t); \mathbf{W}] \quad (8)$$

$$\hat{x}(t+1) = f[\mathbf{y}_p(t); \mathbf{u}(t); \mathbf{W}] \quad (9)$$

The NARX networks trained according to Eqs. (8) and (9) are denoted onwards by NARX-SP and NARX-P networks, respectively.

2.2. Autoregressive Moving Average (ARMA)

ARMA (p, q) prediction model for time series y_t is given as follows:

$$y_t = c + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \phi_j \varepsilon_{t-j} + \varepsilon_t \quad (10)$$

where c is a constant, p is the number of autoregressive orders, q is the number of moving average orders, φ_i is autoregressive coefficients, ϕ_j is moving average coefficients and ε_t is a normal white noise process with zero mean and variance σ^2 .

Box and Jenkins [19] proposed three iterative steps to build ARMA models for time series: model identification, parameter estimation and diagnostic checking. The elaborate information of each step could be found in reference [6]. In order to determine the orders of ARMA model, autocorrelation function (ACF) and partial autocorrelation function (PACF) are used in conjunction with the Akaike information criterion. Other selection technique in associated with ACF and PACF for estimating the orders of ARMA model is maximum likelihood estimation (MLE) [20] which is used in this study.

For a weak stationary stochastic process, the first and second moments exist and do not depend on time:

$$\begin{aligned} E(y_1) &= E(y_2) = \dots = E(y_t) = \mu \\ V(y_1) &= V(y_2) = \dots = V(y_t) = \sigma^2 \\ Cov(y_t, y_{t-k}) &= Cov(y_{t+1}, y_{t-k+1}) = \gamma_k \end{aligned} \quad (11)$$

From the conditions in the Eq. (11), the covariances are functional only of the lag k . These are usually called autocovariances. The autocorrelations, denoted as ρ_k , can be derived only depend on the lag.

$$\begin{aligned} \rho_k &= \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t+k})}} = \frac{E[(Y_t - \mu)(Y_{t+k} - \mu)]}{\sigma_y^k} \\ &= \frac{\gamma_k}{\gamma_0} \end{aligned} \quad (12)$$

The autocorrelations considered as a function of k are referred to as the ACF. Note that since:

$$\begin{aligned}\gamma_k &= Cov(Y_t, Y_{t-k}) = Cov(Y_{t-k}, Y_t) \\ &= Cov(Y_t, Y_{t+k}) = \gamma_{-k}\end{aligned}\quad (13)$$

it follows that $\gamma_k = \gamma_{-k}$, and only the positive half of the ACF is usually given.

In practice, due to a finite time series with N observations, the estimated autocorrelation can be only obtained. If r_k denotes the estimated autocorrelation coefficient, the formula to obtain these parameters is

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \mu)(Y_{t+k} - \mu)}{\sum_{t=1}^n (Y_t - \mu)^2}\quad (14)$$

then the partial ACF can be attained as

$$w_t = (Y_t - \mu) = \Phi_{1k} w_{t-1} + \Phi_{2k} w_{t-2} + \dots + \Phi_{kk} w_{t-k} + \varepsilon_t\quad (15)$$

3. Improved hybrid model for long-term forecasting

Vibration data which is used to indicate the state of machine is not easy to be captured due to its complexity. Hence, none of ARMA and NARX is a suitable model for forecasting this kind of data. By using NARX network, the high noise of this data leads to difficult convergence if the number of neurons is small or over-fitting if the number of neurons is large. On the other hand, ARMA model is not able to apply for the data which is inadequate the stationary condition.

In this paper, an improved hybrid model is proposed in which the vibration data is divided into two components: deterministic and error. The deterministic component $x = [x_1, x_2, \dots, x_k, \dots, x_{t-1}]$ is obtained from a time series data $y = [y_1, y_2, \dots, y_k, \dots, y_t]$ by using filtering technique, where x_k is described as:

$$x_k = \frac{y_{k-1} + y_k + y_{k+1}}{3}, \quad k = 1, 2, 3, \dots, t-1\quad (16)$$

The error component $e = [e_1, e_2, \dots, e_k, \dots, e_{t-1}]$ is the residual between y and x , where $e_k = y_k - x_k$, $k = 1, 2, 3, \dots, t-1$. The deterministic component is degradation indicator which describes clearly the machine's health. This component is suitably captured by NARX network. The error component which is suitable for ARMA model due to being stationary describes the appearance of uncertain parts. The process of m step-ahead prediction using this proposal is shown in [Fig. 2](#)

Fig. 2 The forecasting process of NARX-ARMA model

4. Proposed forecasting system

In order to forecast the future states of machine, the proposed system comprises four

procedures sequentially as shown in Fig. 3, namely, data acquisition, building model, validating model, and predicting. The role of each procedure is explained as follows:

Step 1 Data acquisition: this procedure is used to obtain the vibration data from machine condition. This data is then split into two parts: training set and testing set. Different data is used for different purposes in the prognosis system. Training set is used for creating the prediction models whilst testing set is utilized to test the trained models.

Step 2 Building model: Training data is separated into two components: deterministic component and error component. They are used to build NARX-ARMA model as the process mentioned in the previous section.

Step 3 Validating model: this procedure is used for measuring the performance capability.

Step 4 Forecasting: long-term prediction method is used to forecast the future states of machine. The predicted results are measured by the error between predicted values and actual values in the testing set.

Fig. 3 Proposed forecasting system

5. Experiments and results

5.1 Experiments

The proposed method is applied to a real system to predict the trending data of a low methane compressor of a petrochemical plant. The compressor shown in Fig. 4 is driven by a 440 kW motor, 6600 volt, 2 poles and operating at a speed of 3565 rpm. Other information of the system is summarized in Table 1

Fig. 4 Low methane compressor

Table 1 Information of the system

The condition monitoring system of this compressor consists of two types, namely off-line and on-line. In the off-line system, accelerometers were installed along axial, vertical, and horizontal directions at various locations of drive-end motor, non drive-end motor, male rotor compressor and suction part of compressor. In the on-line system, accelerometers were located at the same positions as in the off-line system but only in the horizontal direction.

The trending data was recorded from August 2005 to November 2005 which included peak acceleration and envelope acceleration data. The average recording duration was 6 hours during the data acquisition process. Each data record consisted of approximately 1200 data points as shown in Figs 5 and 6, and contained information of machine history with respect to time

sequence (vibration amplitude). Consequently, it can be classified as time-series data.

Fig. 5 The entire peak acceleration data of low methane compressor

Fig. 6 The entire envelope acceleration data of low methane compressor

5.2. Results

In order to build the forecasting model, 719 points of peak acceleration data and 749 points of envelope acceleration data are used. The remaining points of each data are then utilized to test the forecasting model. Additionally, the root-mean square error (RMSE) given in Eq. (17) is employed to evaluate forecasting capability

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (17)$$

where N represents the number of data points, y_i is actual value and \hat{y}_i represents the predicted value.

The deterministic component x_i is obtained from vibration data by using filtering technique. Figs. 7 and 8 show the deterministic data after filtering of the envelope and peak acceleration data. The NARX forecasting model is then generated by using these data. To build NARX model, the embedding dimension d_E must be firstly determined. As mentioned in the introduction section, FNN as well as Cao's method could be possibly used to estimate d_E . Cao's method can settle on a suitable embedding dimension of time series and distinguish deterministic signals and stochastic signals clearly. That is a reason why Cao's method is chosen in this paper. According to [15], there are two important values that are $E_1(d)$ and $E_2(d)$ needed to be calculated. $E_1(d)$ is used to choose the minimum embedding dimension d_E when it reaches the saturation. $E_2(d)$ is used for the problem in practical computations where $E_1(d)$ is slowly increasing or has stopped changing if embedding dimension d is sufficiently large. For random data, the future values are independent of the past values, hence, $E_2(d) \approx 1$ for any value of d . However, for deterministic data, $E_2(d)$ is certainly related to as a result, it cannot be a constant for all d . Figs. 9 and 10 depict the embedding dimension applied for deterministic component of envelope and peak acceleration data, respectively. In these figures, $E_1(d)$ obviously reaches its saturation at $d = 4$. Consequently, the minimum embedding dimension d_E is chosen as 4 for envelope acceleration data and 5 for peak acceleration data to build NARX model.

Fig. 7 The deterministic component of envelope acceleration data

Fig. 8 The deterministic component of peak acceleration data

Fig. 9 The value E_1 and E_2 of deterministic component of envelope data

Fig. 10 The value E_1 and E_2 of deterministic component of peak acceleration data

Other parameters to be considered are the number of neurons in two hidden layers of NARX model. The number of neurons, N_1 and N_2 , in the first and second hidden layers is chosen according to the following heuristics:

$$N_1 = 2d_E + 1, \quad N_2 = \sqrt{N_1} \quad (18)$$

where N_2 is rounded up toward the next integer number. By substituting $d_E = 4$ and $d_E = 5$ into Eq. (18), the number of neurons can be found as $N_1 = 9$, $N_2 = 3$ for envelope acceleration data and $N_1 = 11$, $N_2 = 3$ for peak acceleration data, respectively.

The order of the output regressor n_y in NARX-P and NARXSP models is calculated by product of time delay τ and minimum embedding dimension d_E . The time delay τ is chosen as 1 because long-term recursive forecasting methodology is used in this paper. Thus, the order of the output regressor n_y is respectively set to $n_y = \tau \times d_E = 4$ and $n_y = 5$ for enveloped and peak acceleration data. Both the NARX-P and NARX-SP models are employed to select the proper NARX model. The standard back propagation algorithm in which 500 epochs and learning rate equals to 0.01 is used to train the networks. The forecasting capability of these models is evaluated by RMSE values that show in Table 2. From this table, NARX-SP model is superior to NARX-P model in showing more the accuracy. Hence, NARX-SP is chosen to hybridize with ARMA model.

Table 2 The RMSE values of NARX-SP and NARX-P

In the next step in forecasting process, the error e mentioned in section 3 is forecasted by ARMA model. In order to create this model, the model identification procedure is initially implemented to check the stationary condition. In case of the inadequate stationary condition, it is considered that how many orders of differencing need to stationalize the data. Time series is considered to be stationary if its autocorrelation structure is constant over time or the lag-1 autocorrelation is zero or negative. Figs. 11 and 12 which depict the ACF of envelope and peak acceleration data show that the error component is satisfied the requirement of stationary condition. Thus, it can be directly applied to generate ARMA model without necessitating higher order of differencing. Basing on ACF, PACF and experimental results, ARMA (3, 4) model for envelope acceleration data and ARMA (3, 3) for peak acceleration data are chosen in this study. Furthermore, MLE is used to estimate the model parameters φ_i , ϕ_i .

Fig. 11 ACF of error component of envelope acceleration data

Fig. 12 ACF of error component of peak acceleration data

Finally, the final forecasting values of the hybrid model are the sum of the results obtained

from NARX-SP and ARMA models. Table 3 shows a summary of the RMSE values of three models applied for peak and envelope acceleration data. In this table, all the RMSEs of the NARX-ARMA model are vastly superior to the other traditional models in that it is more accurate in both cases of peak and envelope acceleration data. The example of forecasting capability using one-step ahead shows in Figs. 13 and 14 of peak acceleration and envelope acceleration data, respectively. They indicate that the NARX-ARMA hybrid model can effectively capture and track the system behavior.

Table 3 The RMSE values of ARMA, NARX and NARX-ARMA

Fig. 13 The forecasted results of NARX-ARMA model for envelope acceleration data using one-step-ahead

Fig. 14 The forecasted results of NARX-ARMA model for peak acceleration data using one-step-ahead

6. Conclusions

Machine state forecasting gradually plays an important role in modern industry due to its ability to foretell the operating condition of machine in the future. Hence, finding out the precise and reliable forecasting model is an important and challenging task. In this paper, an improvement of hybrid model consisted of NARX and ARMA is investigated for long-term forecasting. Peak acceleration and envelope acceleration trending data of a low methane compressor are used to demonstrate the predictive ability of proposed method. From the results of a comparative study, the improved hybrid model (NARX-ARMA) has a higher forecasting accuracy the other traditional models. This demonstrates that the NARX-ARMA model is a reliable and accurate tool for forecasting the machine state.

References

- [1] J. Liu, W. Wang, F. Golnaraghi, A multi-step predictor with a variable input pattern for system state forecasting, *Mechanical Systems and Signal Processing*, in press.
- [2] G. Vachtsevanos, P. Wang, Fault prognosis using dynamic wavelet neural networks. *AUTOTESTCON Proceedings, IEEE Systems Readiness Technology Conference, 2001*, pp. 857-870.
- [3] W.Q. Wang, M.F. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy system. *Mechanical System and Signal Processing* 18 (2004) 813-831.
- [4] Wilson Wang, An adaptive predictor for dynamic system forecasting, *Mechanical Systems and Signal Processing* 21 (2007) 809-823.
- [5] V.T. Tran, B.S. Yang, M.S. Oh, A.C.C Tan, Machine condition prognosis based on

- regression trees and one-step-ahead prediction, *Mechanical Systems and Signal Processing*, 22 (5) (2008), 1179–1193.
- [6] G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing* 50 (2003) 159-175.
- [7] H. Inde, T.B. Trafalis, A hybrid model for exchange rate prediction, *Decision Support Systems* 42 (2006) 1054-1062.
- [8] P.-F. Pai, C.-S. Lin, A hybrid ARIMA and support vector machines model in stock price forecastin, *Omega* 33 (2005) 497-505.
- [9] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L.J. Herrera, A. Guillen, L. Marquez, M. Pasadas, Hybridization of intelligent techniques and ARIMA models for time series prediction, *Fuzzy Sets and Systems* 159 (2008) 821-845.
- [10] I. Rojas, O. Valenzuela, F. Rojas, A. Guillen, L.J. Herrera, H. Pomares, L. Marquez, M. Pasadas, Soft-computing techniques and ARMA model for time series prediction *Neurocomputing* 71 (2008) 519-537.
- [11] F.-M. Tseng, H.-C. Yu, G.-H. Tzeng, Combining neural network model with seasonal time series ARIMA model *Technological Forecasting & Social Change* 69 (2002) 71-87.
- [12] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, A. Lendasse, Methodology for long-term prediction of time series, *Neurocomputing* 70 (2007) 2861-2869.
- [13] I.J. Leontaritis, S.A. Billings, Input-output parametric models for non-linear systems, *International Journal of Control* 41 (1985) 303-344.
- [14] G.E.P. Box, G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA, 1970
- [15] L.Y. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, *Physical D* 110 (1997) 43-50.
- [16] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review A* 45 (1992) 3403-3411.
- [17] T. Lin, B. G. Horne, P. Tino, C. L. Giles. A delay damage model selection algorithm for NARX neural networks, *IEEE Transactions on Signal Processing* 45(11) (1997) 2719-2730.
- [18] J.M.P. Menezes, G.A. Barreto, A new look at nonlinear time series prediction with NARX recurrent neural network, *Proceeding of the Ninth Brazilian Symposium on Neural Networks*, 2006, pp. 160-165.
- [19] G.E.P. Box, G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA, 1970.
- [20] L. Ljung, G.E.P. Box, *System Identification Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

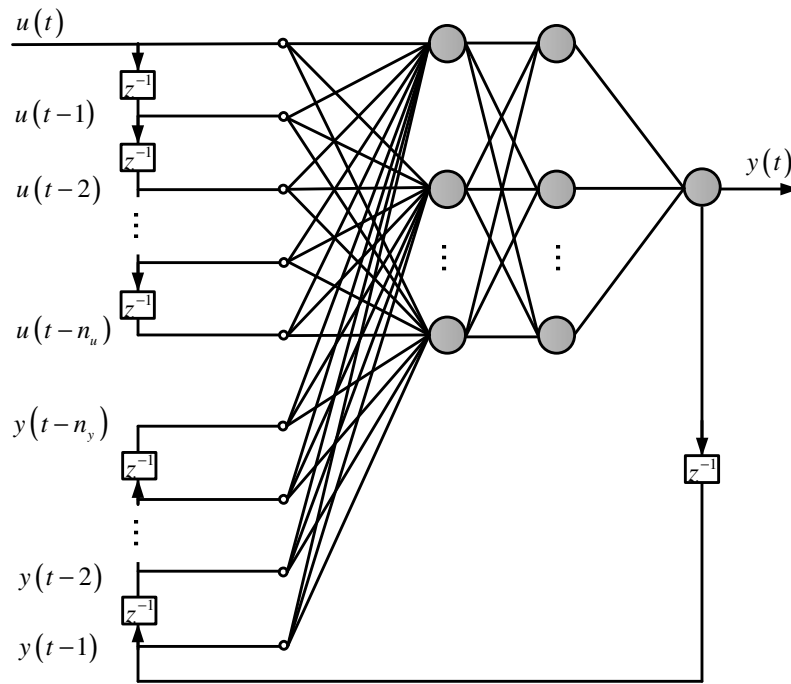


Fig. 1 The structure of NARX with n_u inputs and n_y output delays

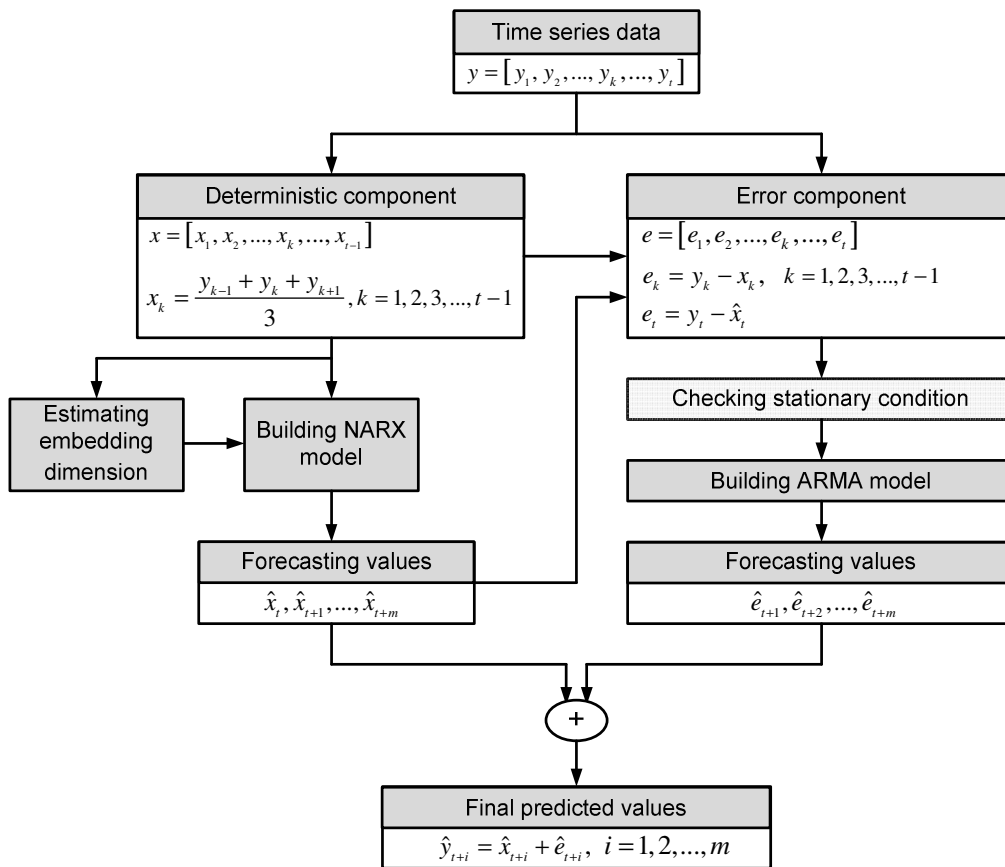


Fig. 2 The forecasting process of NARX-ARMA model

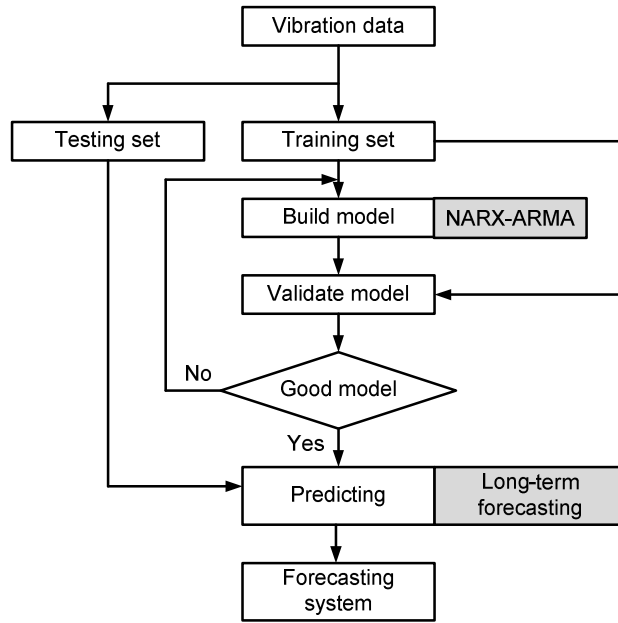


Fig. 3 Proposed forecasting system

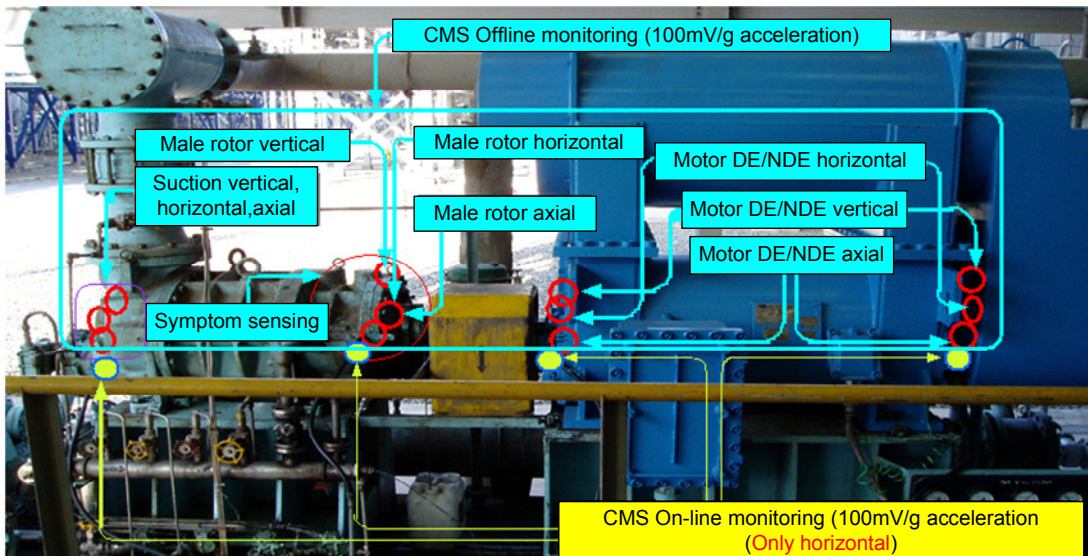


Fig. 4 Low methane compressor: wet screw type.

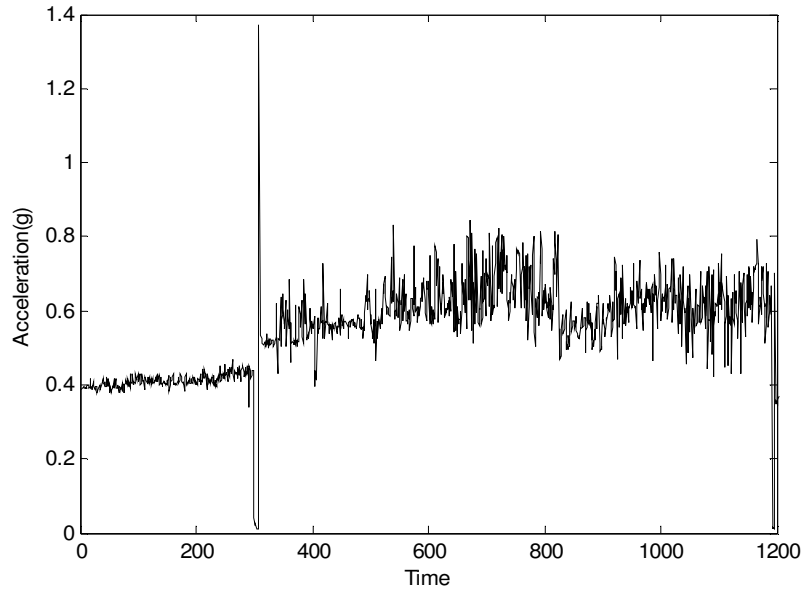


Fig. 5 The entire peak acceleration data of low methane compressor

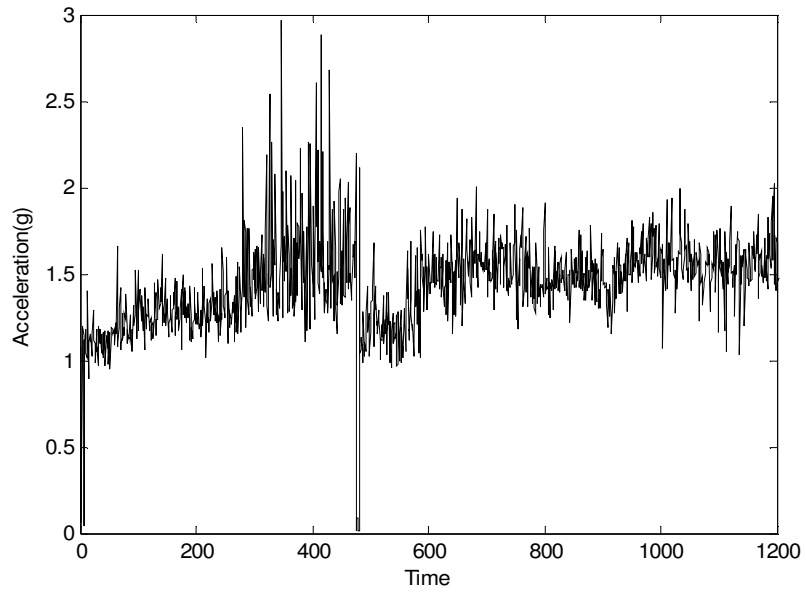


Fig. 6 The entire envelope acceleration data of low methane compressor

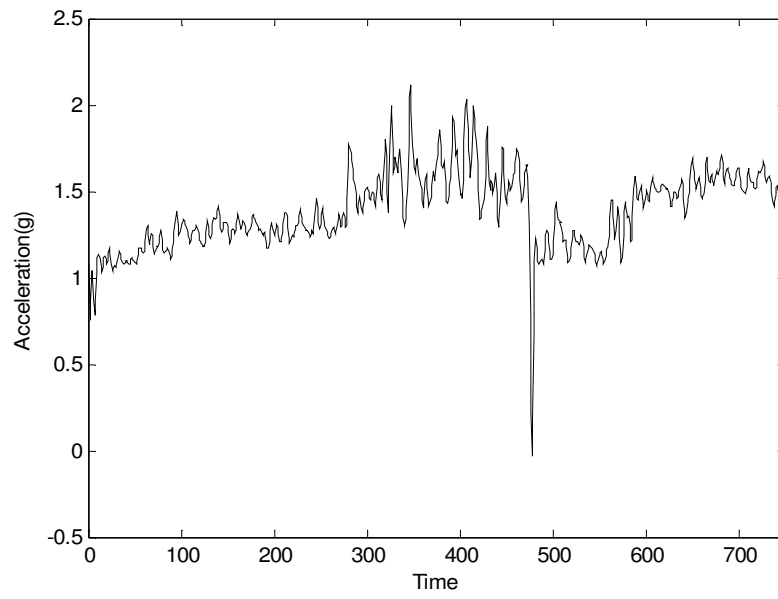


Fig. 7 The filtered envelope acceleration data of low methane compressor

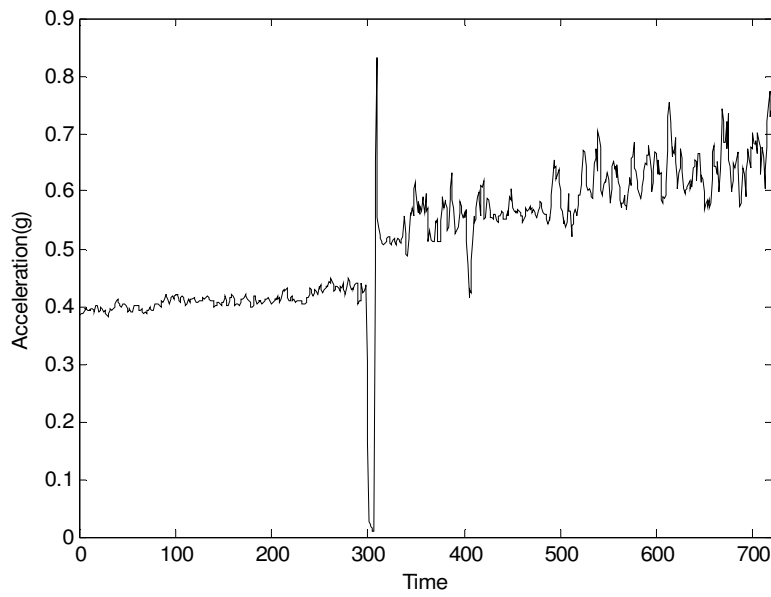


Fig. 8 The deterministic component of peak acceleration data

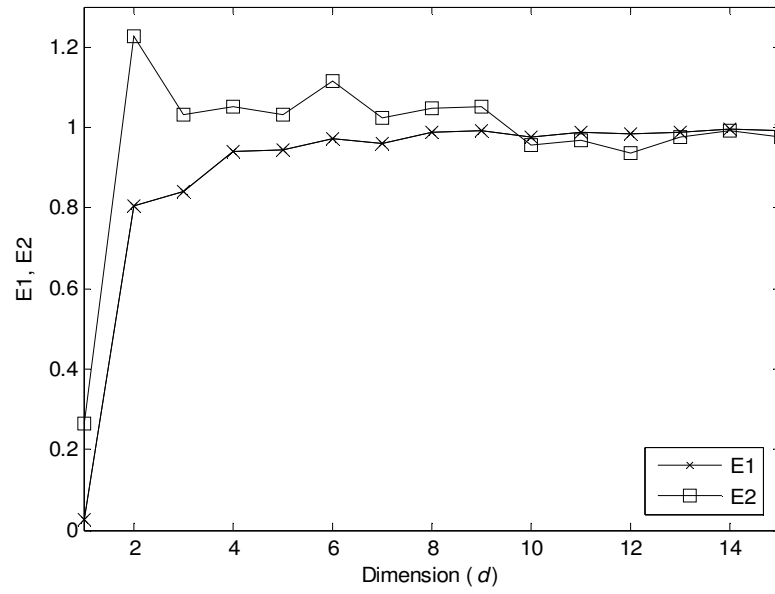


Fig. 9 The value E_1 and E_2 of deterministic component of envelope data

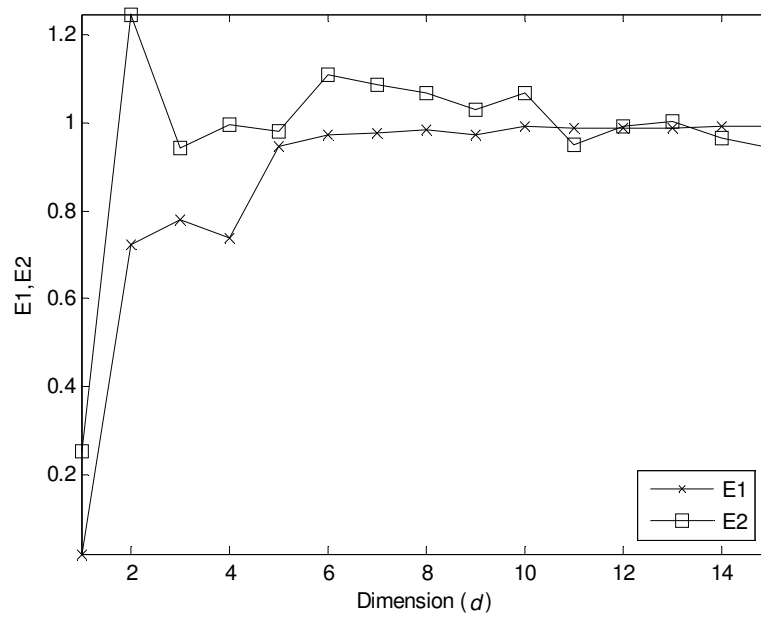


Fig. 10 The value E_1 and E_2 of deterministic component of peak acceleration data

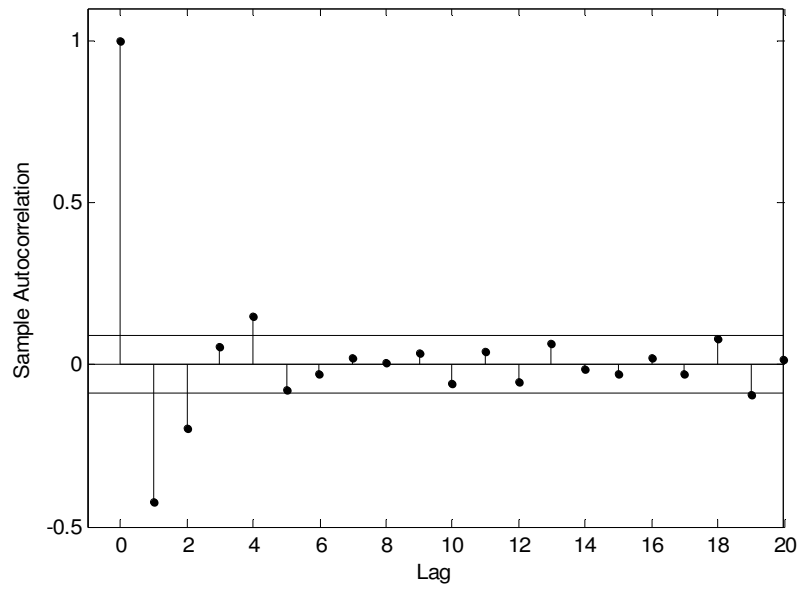


Fig. 11 ACF of error component of envelope acceleration data

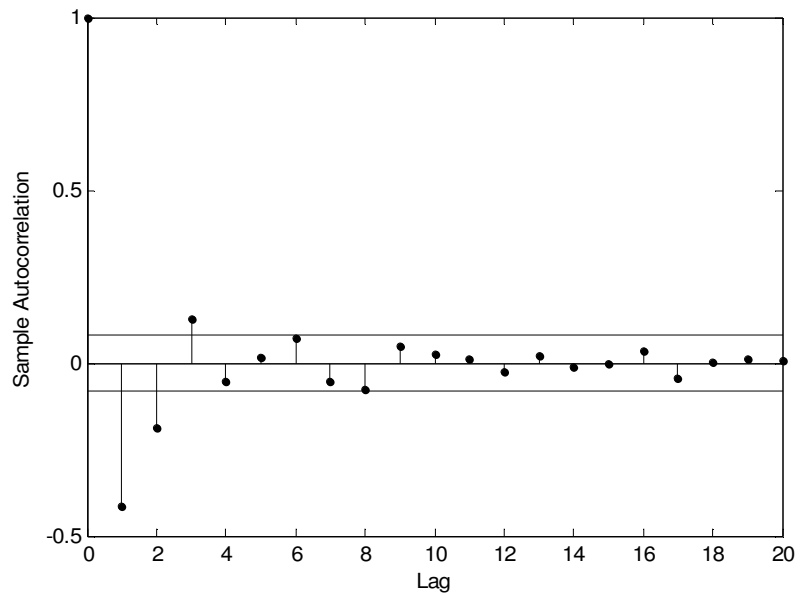


Fig. 12 ACF of error component of peak acceleration data

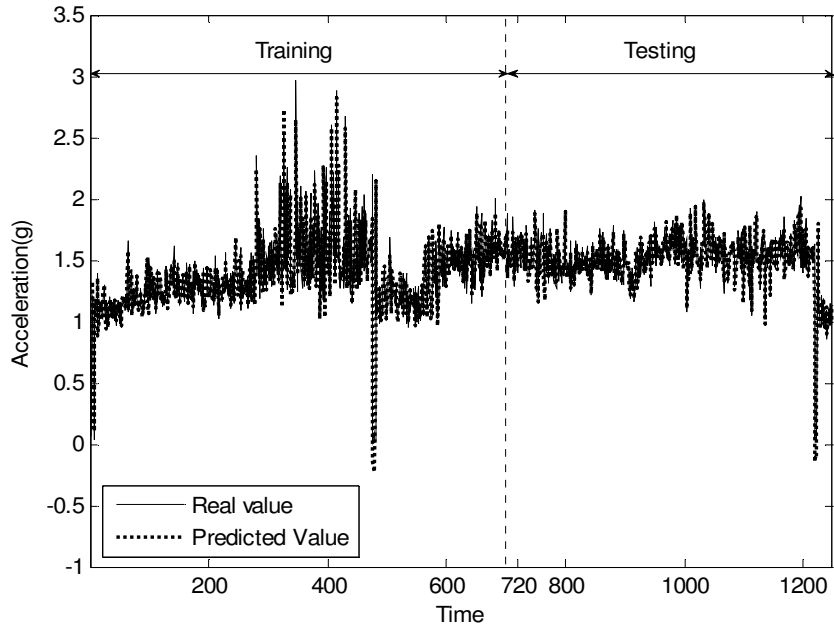


Fig. 13 The forecasting results of NARX-ARMA model for envelope acceleration data using one-step-ahead

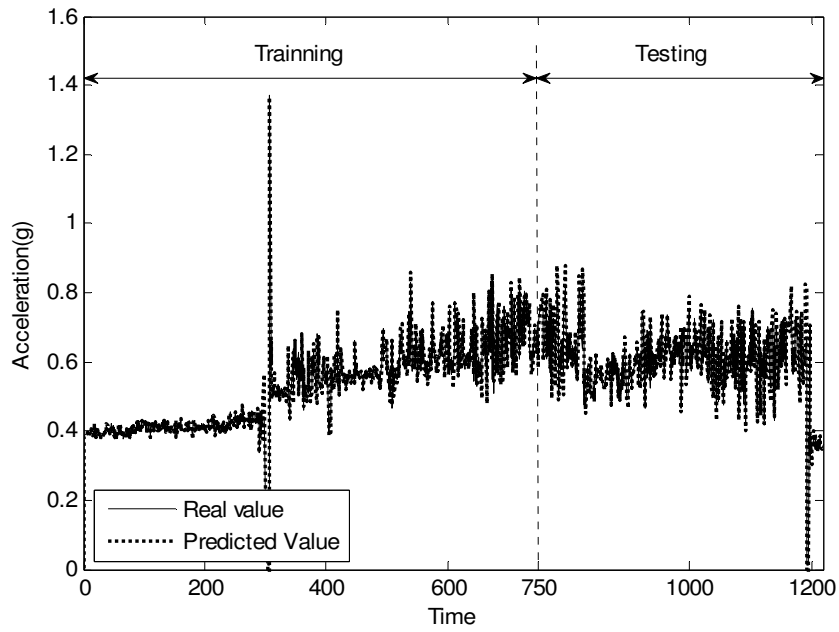


Fig. 14 The forecasting results of NARX-ARMA model for peak acceleration data using one-step-ahead

Table 1 Description of system

Electric motor		Compressor	
Voltage	6600 V	Type	Wet screw
Power	440 kW	Lobe	Male rotor (4 lobes)
Pole	2 Pole		Female rotor (6 lobes)
Bearing	NDE:#6216, DE:#6216	Bearing	Thrust: 7321 BDB
RPM	3565 rpm		Radial: Sleeve type

Table 2 The RMSE values of NARX-SP and NARX-P

Vibration data	NARX-SP		NARX-P	
	Training	Predicting	Training	Predicting
Peak	0.0232	0.0260	0.4170	0.0450
Envelope	0.0241	0.0188	0.0597	0.0363

Table 3 The RMSE values of forecasting results of ARMA, NARX and NARX-ARMA

Number of step ahead	Peak acceleration data			Envelope acceleration data		
	ARMA	NARX	NARX-ARMA	ARMA	NARX	NARX-ARMA
1	0.2647	0.0703	0.0363	0.2647	0.2090	0.0693
2	0.2821	0.2221	0.0595	0.2821	0.2581	0.1178
3	0.2921	0.2421	0.0671	0.2826	0.2664	0.1232
4	0.3026	0.2826	0.0896	0.2935	0.2867	0.1276
5	0.3226	0.3026	0.0956	0.2950	0.3018	0.1310
6	0.3950	0.3250	0.1764	0.3125	0.3271	0.1331