Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Model-based monitoring and fault diagnosis of fossil power plant process units using Group Method of Data Handling

Fan Li[a], Belle R. Upadhyaya[a,*,1], Lonnie A. Coffey[b]

[a] *Department of Nuclear Engineering, University of Tennessee, Knoxville, TN 37996-2300, United States*

[b] *Combustion Consultants, 6505 Westland Drive, Knoxville, TN 37919, United States*

ARTICLE INFO

Article history:
Received 23 June 2008
Received in revised form
15 September 2008
Accepted 28 October 2008
Available online 11 December 2008

Keywords:
Fault diagnosis
Group method of data handling
Feedwater heaters

ABSTRACT

This paper presents an incipient fault diagnosis approach based on the Group Method of Data Handling (GMDH) technique. The GMDH algorithm provides a generic framework for characterizing the interrelationships among a set of process variables of fossil power plant sub-systems and is employed to generate estimates of important variables in a data-driven fashion. In this paper, ridge regression techniques are incorporated into the ordinary least squares (OLS) estimator to solve regression coefficients at each layer of the GMDH network. The fault diagnosis method is applied to feedwater heater leak detection with data from an operating coal-fired plant. The results demonstrate the proposed method is capable of providing an early warning to operators when a process fault or an equipment fault occurs in a fossil power plant.

1. Introduction

The on-line monitoring and fault diagnosis of a large industrial process, such as a fossil power plant, is very important in process operation and performance monitoring. Increasing the safety, reliability and availability of the different units involved in the process scheme is a major concern for plant operation and equipment maintenance. In recent years, fossil power plants have developed a capability of acquiring historic data, due to the incorporation of distributed control system (DCS) technology. Process variables are continuously stored in databases at discrete time intervals. This explosive growth in data and databases at power plants has generated an urgent need for new techniques and tools that can intelligently and automatically transform the process data into useful equipment health information and maintenance knowledge.

Many modern fault diagnostic approaches are based on analytical redundancy. Functional relationships among process variables governed by fundamental conservation laws such as mass, momentum, and energy balance, can replace hardware redundancy for plant measurements [1]. However, despite the fact that a vast variety of analytical techniques for fault diagnosis of nonlinear systems can be found in the literature [2–4], they usually lack an appropriate mathematical description of the system of interest. System identification techniques are applied in many fields in order to model and predict behaviors of complex systems based on given input–output data [5]. The so-called soft computing methods, such as neural networks [6], principal component analysis [7], fuzzy logic [8], and evolutionary algorithms [9], and other data-based techniques [10] have shown great capabilities of solving complex nonlinear system identification and control problems.

In this paper, robust fault detection using nonlinear models that are designed with an evolutionary technique is proposed for process units of a fossil power plant, such as feedwater heaters, pumps, pulverizers, and others. The nonlinear model considered here is developed by a group method of data handling (GMDH) learning algorithm. The GMDH was first developed by Ivakhnenko [11] as a multivariate analysis method for complex system modeling and identification. It constructs high-order regression type models beginning with a few basic quadratic equations and generates gradually more complicated models based on the evaluation of model performances on a set of multi-input, single-output data pairs. As this algorithm continues for several generations, a set of empirical models is developed that behaves more like the actual system of interest [12]. The GMDH algorithm has the ability of circumventing the difficulty of having priori knowledge of the mathematical model of the process being considered. The functional relationships between the response and predictor variables are learned directly from a self-organization of the data. The algorithm has a high level of flexibility as different number of nonlinear input variables can be chosen at each layer of the GMDH network. In comparison

* Corresponding author. Tel.: +1 865 974 7576; fax: +1 865 974 0668.
E-mail address: bupadhya@utk.edu (B.R. Upadhyaya).

[1] Senior Member, ISA.

0019-0578/$ – see front matter © 2008 ISA. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.isatra.2008.10.014

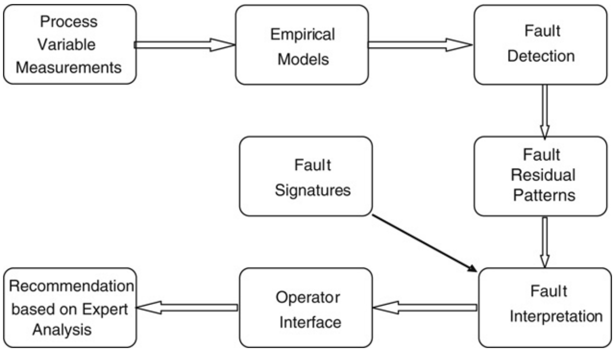

Fig. 1. Schematic of fault diagnosis system.

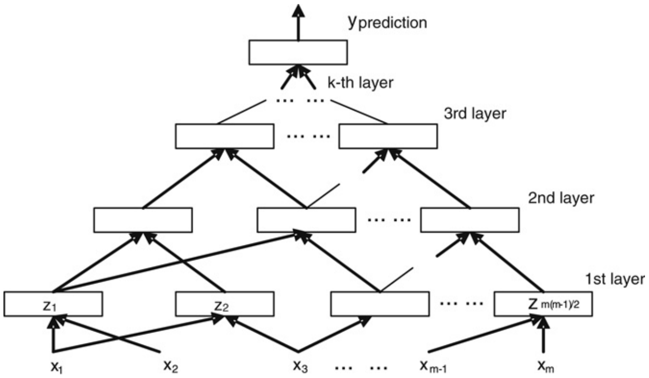

Fig. 2. GMDH model structure (reproduced from Ref. [16]).

with conventional multilayered neural networks whose topologies are usually predefined prior to training, the GMDH architecture is not fixed in advance but fully optimized both structurally and parametrically during training. Particularly, the different number of iterative layers of the GMDH network can be selected according to the requirement of computational accuracy. Moreover, the model structure is fully known, unlike that of a neural network model. Recently, the use of the GMDH algorithm has led to successful application in a broad range of areas in engineering, science, and economics [13–15].

A schematic of the proposed fault diagnostic system is given in Fig. 1, where it is applied to a process unit of a large-scale industrial system. The goal is to use GMDH models to characterize the interrelationship among a multitude of process variables. The mathematical difference between an observed value and the corresponding model estimate for that observation, usually referred as *residual*, is calculated for each observation. The residual values and their variations from allowable deviations are indicative of possible faults. Also, fault interpretation relies on the fact that different failure modes may be distinguished based on residual deviation patterns of each process variable being monitored. Each of these residual patterns is unique to a given failure mode, and hence is called a *fault signature*. There are numerous approaches by which fault signatures can be developed for fault interpretation task. Expert knowledge, their engineering judgment, and data analysis of actual faulty equipment are among those most commonly used.

This paper has three major sections. Section 2 describes the GMDH algorithm and the steps involved in its implementation, along with the discussion regarding ridge regression techniques. The application of the proposed fault diagnosis method to a high-pressure feedwater heater leak detection using operational data is presented in Section 3. Finally, some concluding remarks are given in Section 4.

2. Group Method of Data Handling (GMDH) algorithm

One of the advantages of a data-based modeling method is the simplicity of its implementation due to the automated searching process in model construction. The GMDH is developed heuristically to identify mathematical description of a complex system by using only the data and the choice of input and output signals. The technique is described briefly in this section.

2.1. Principle of GMDH algorithm

The GMDH is an evolutionary approach for constructing system models in a self-organizing fashion. For a given input vector $X = (x_1, x_2, \dots, x_M)$, the idea of the identification problem is to find a function $\hat{f}$ that can be approximately used instead of the actual function f, in order to predict output $\hat{y}$ as closely as possible to its actual output y.

The algorithm constructs a high-order polynomial of the Kolmogorov–Gabor form,

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

By a composition of lower-order polynomials via a generative, self-organizing algorithm, this full form of mathematical description can be represented by a set of quadratic polynomials consisting of only two variables in the form

$$\hat{y} = A + Bx_1 + Cx_2 + Dx_1^2 + Ex_2^2 + Fx_1x_2. \quad (2)$$

Thus, the GMDH algorithm combines second-order regression-type polynomials at each generation to arrive at the next generation of approximations. Fig. 2 illustrates the structure adopted in the GMDH network, which begins with the original input variables and becomes more complex in its model structure as the number of layers increases.

The input data set, represented by the input variable matrix $X = \{x_{ij}\}$ is generally divided into two subsets: the training set and the testing set, as well as the output variable vector $Y = \{y_i\}$, $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$, where N is the total number of observations and M is the total number of input variables. The purpose is to check for over-fitting during the training phase and to achieve parsimony in the functional approximation.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1M} \\ x_{21} & x_{22} & \dots & x_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tM} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NM} \end{bmatrix}, \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_N \end{bmatrix}.$$

The data from 1 to t are used for model training and the data from $t+1$ to N are used for model validation and for an overfitting check.

The main procedures for GMDH algorithm implementation used for a given set of N observations of M independent variables is described below [11,12].

Step 1: Calculate the regression polynomials of the form given in Eq. (2) that best fits the dependent observations y^i (i_{th} observation of the training output vector) for each pair of input variables (columns of X matrix) in the training data set. All possible combinations give a total number of $M(M-1)/2$ regression polynomials that are computed from the observations. For each pair of independent variables, a set of coefficients of the regression will be estimated using the least squares method so that the difference between the actual output and the estimated output is minimized.

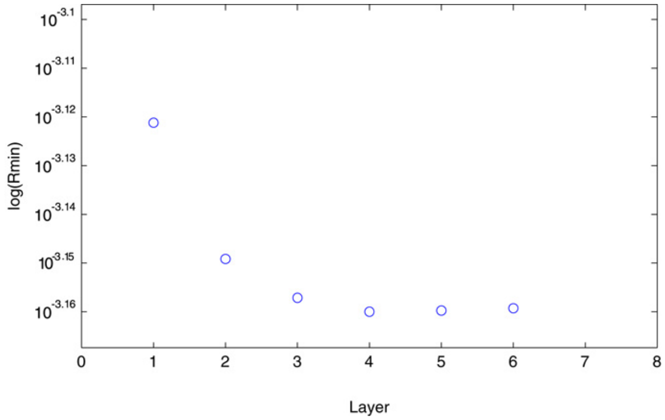

Fig. 3. GMDH training termination criterion.

Step 2: At the completion of step 1, the algorithm has constructed $M(M-1)/2$ new variables $Z = (z_1, z_2, \ldots, z_{M(M-1)/2})$ for each of the t observations in the training dataset. These new variables are evaluated using a regularity criterion on the validation data (observation $t + 1$ to N), and the least effective variables are screened out. The regularity criterion r_j, which was used by Ivakhnenko [11], is defined as the square root of the mean-squared error of predictions for the j_{th} regression polynomial:

$$r_j^2 = \frac{\sum_{i=t+1}^{N} \left(y^i - z_j^i\right)^2}{\sum_{i=t+1}^{N} (y^i)^2}, \quad j = 1, 2, \ldots, M(M-1)/2. \tag{3}$$

Step 3: Order the new variables according to ascending values of the regularity criterion r_j, and retain those best-fitted variables $z_1, z_2, \ldots, z_{M_1}$ (where M_1 is the total number of the retained columns) in the vector Z that satisfy $r_j \le R$ (R is a predefined threshold value which affects modeling accuracy and therefore needs careful choosing, 1‰ in our computations) to replace the original columns $x_1, x_2, \ldots, x_M$ of X. From these new predictors in Z, we will combine two of them at a time exactly as we did for the previous layer. That is, a new set of $M_1(M_1 - 1)/2$ regression polynomials for predicting y^i are computed. The best-fitted of the new estimates are selected, and new independent variables are generated to replace the old because they have better predictive capability.

Step 4: The smallest r_j in each generation is determined, which is used as the criterion of terminating the GMDH training process. If the smallest r_j starts to increase at a certain layer, which indicates that the model fits the data with desired accuracy or the introduction of new layer does not induce a significant increase in the GMDH model performance, then stop the training process. Otherwise, return to step 1. Fig. 3 illustrates the stopping criterion for the GMDH algorithm during training. The other stopping criterion is that the GMDH algorithm terminates when the predetermined number of iterative layers is reached, as a carefully maintained balance between model accuracy and its complexity is always desired by the model developer.

At the end of the GMDH algorithm, an estimate of y is obtained as a quadratic function of two variables, which are themselves quadratic of two more variables, and so on. In other words, if we are to make the necessary algebraic substitutions, we will arrive at a fairly high-order polynomial that has the form shown in Eq. (1).

2.2. Addition of nonlinear input functions

Although only quadratic terms are used in the original GMDH method at each layer of computation (Eq. (2)), the enhanced GMDH algorithm used in this work includes more nonlinear functions of the basic measurements, such as ratio, trigonometric, logic terms, and others. The complete set of 11 different terms used by GMDH in this work is shown below.

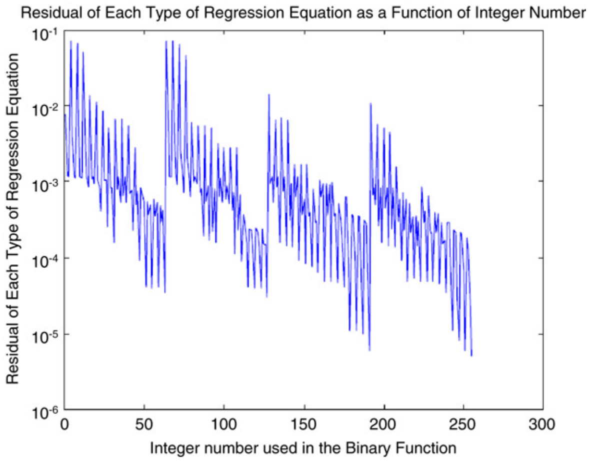

Fig. 4. Residuals of different combinations of input functions.

$$\left\{ \begin{array}{c} (1, x_1, x_2), (x_1^2, x_2^2), (x_1 x_2), \left(\frac{1}{x_1}, \frac{1}{x_2}\right), \left(\frac{1}{x_1^2}, \frac{1}{x_2^2}\right), \\ \left(\frac{1}{x_1 + x_2}\right), \left(\frac{1}{x_1 x_2}\right), \left(\frac{x_1}{x_2}, \frac{x_2}{x_1}\right), \\ \left(\frac{x_1}{x_1 + x_2}, \frac{x_2}{x_1 + x_2}\right), \left(\frac{x_1 + x_2}{x_1}, \frac{x_1 + x_2}{x_2}\right), \\ (\sin(x_1), \sin(x_2), \cos(x_1), \cos(x_2)) \end{array} \right\}. \tag{4}$$

An optimal model structure selection algorithm is utilized in this work to test different combinations of a given set of input functions shown in Eq. (4) in a systematic fashion. For each possible combination of the basic functions, the GMDH algorithm estimates the value of the dependent variable. An overall residual value between the test data points and the predicted values of the best regression equation found by GMDH is computed. This overall residual is then compared with residuals from other models with different combinations of input functional terms. Fig. 4 illustrates a sample plot of the residuals of each model as a function of binary numbers. The model that generates the lowest residual is selected for further use. A binary number generator is built into the GMDH computation to make this selection automatically. In the case of having defined a maximum of k candidate terms, this binary number goes from 1 to $2^k - 1$.

With the addition of more nonlinear input functions, a more complex polynomial can be written for two arbitrary input variables as

$$\begin{aligned} \hat{y}_i &= \hat{f}(x_{ip}, x_{iq}) = \hat{\beta}_0 + \sum_{j=1}^{g} \hat{\beta}_j \varphi_j(x_{ip}, x_{iq}) \\ &= \hat{\beta}_0 + (\hat{\beta}_1, \ldots, \hat{\beta}_g) \Phi_i(x_{ip}, x_{iq})^{\mathrm{T}} \end{aligned} \tag{5}$$

where the basis functions $\Phi_i(x) = (\varphi_1(x), \ldots, \varphi_g(x))^{\mathrm{T}}$ are nonlinear functions of the observations $\left\{(y_i, x_{ip}, x_{iq}), i = 1, 2, \ldots, N\right\}$ for different $p, q \in (1, 2, \ldots, M)$.

Using the above expression for each of N data triples $\left\{(y_i, x_{ip}, x_{iq}), i = 1, 2, \ldots, N\right\}$, the following matrix can be readily obtained as

$$\Phi \beta = Y \tag{6}$$

where $\beta = (\hat{\beta}_0, \ldots, \hat{\beta}_g)^{\mathrm{T}}$ is the vector of model coefficients and $Y = (y_1, \ldots, y_N)^{\mathrm{T}}$ is the vector of the observed output values. The Φ matrix has the form

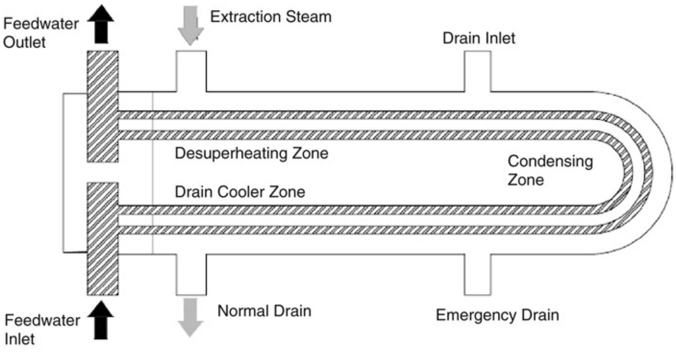

Fig. 5. Schematic of a horizontal closed feedwater heater.

$$\Phi = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}^2 & x_{1q}^2 & x_{1p}x_{1q} & \frac{1}{x_{1p}} & \frac{1}{x_{1q}} & \cdots \\ 1 & x_{2p} & x_{2q} & x_{2p}^2 & x_{2q}^2 & x_{2p}x_{2q} & \frac{1}{x_{2p}} & \frac{1}{x_{2q}} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{Np} & x_{Nq} & x_{Np}^2 & x_{Nq}^2 & x_{Np}x_{Nq} & \frac{1}{x_{Np}} & \frac{1}{x_{Nq}} & \cdots \end{bmatrix}. \quad (7)$$

The ordinary least squares estimator leads to the solution of the regression coefficients

$$\beta = (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}Y. \quad (8)$$

Rational functions are found to be especially useful in modeling dynamic processes [16]. However, the solution of Eq. (8) requires the inversion of a matrix whose dimensions are equal to the number of training samples. The problem may be ill-conditioned and thus leads to the unstable solutions. In this case, other regularization approaches such as ridge regression are recommended to obtain more stable estimators.

2.3. Ridge regression

To deal with ill-conditioned problems, Tikhonov regularization [17] can be used to obtain a regularized solution to this problem. In this method, the minimization problem can be represented by the following functional:

$$\min \left\{ \|\Phi\beta - Y\|_2^2 + \lambda \, \|L\beta\|_2^2 \right\} \quad (9)$$

where $\lambda > 0$ is a regularization parameter that controls the trade-off between the smoothness of the solution and the fitness to the data. L is a well-conditioned matrix; for example, a discrete approximation of the derivative operator. The main assumption behind Tikhonov regularization is that the solution should be smooth or non-oscillating. In the case of $L = I$, where I is identity matrix, the Tikhonov's function (Eq. (9)) is said to be in standard form and is known in statistical literature as ridge regression. In this case, we can write the regularized solution as

$$\beta = (\Phi^{\mathrm{T}}\Phi + \lambda I)^{-1}\Phi^{\mathrm{T}}Y. \quad (10)$$

Ridge regression was proposed for the first time by Hoerl and Kennard [18]. This approach modifies the OLS method by influencing directly on the estimation of regression coefficients. In fact, ridge regression coefficients are calculated by introducing regularization parameter λ into the OLS equations. Thus, the calculated coefficients are biased but are more stable than those resulting from multiple regressions. In this work, ridge regression is integrated into the GMDH algorithm to estimate the best model structure and identify regression coefficients at each layer of the GMDH network. Potential ill-conditioned problems associated with collinearity can be avoided. More details concerning ridge regression and model coefficient calculation may be found in references [19–21].

3. Application to feedwater heater monitoring

3.1. System description and problem statement

Feedwater heaters are important process units in large-scale thermal power plants and serve two major functions. First, regenerative principles are applied to improve the efficiency of the cycle by extracting superheated steam at different stages of the turbine to the associated feedwater heaters to pre-heat the water before it feeds into the boiler. Thus, more electrical power output can be generated for the same heat power input. The second function of a feedwater heater system is to protect boiler internals from thermal stresses by raising feedwater temperature to a value close to the component temperatures [22].

Most feedwater heaters used in fossil power plants are of the shell and U-tube type, horizontal, and with three zone configuration. A typical configuration is given in Fig. 5. Feedwater entering the heater first passes through the drain-cooling region where single-phase convection is the leading heat transfer mechanism from the drains on the shell side to the feedwater on the tube side. The purpose of this section is to cool the drains to a temperature that is close to the shell side saturation temperature of the next heater where it will be mixed with the extraction steam. The feedwater is then passed to the condensing region where the majority of the heat transfer takes place. The feedwater temperature can be brought up within 5 °C of shell side saturation temperature in this section. Finally, a de-superheating region is used to raise the feedwater temperature even above saturation and cool the steam down to saturation.

A common and costly failure of feedwater heaters is that water or wet steam is inducted into the turbine from the extraction lines of the turbine as a result of rupture or extensive leakage from the heater interior tubing into the steam space of the feedwater heater. This is often caused by aging and thermal stress of the heater tubing. When the tubing starts rupturing, the rapid increase in steam pressure on the shell side will cause a reverse flow of the wet steam from the heater toward the turbine through the extraction line. As a result, wet steam may be allowed to enter the turbine and cause severe damage to the turbine blades. Thus, in cases of feedwater heater tube leakage, it is important to detect the problems as early as possible so that repair and shutdown, if necessary, can be scheduled well in advance.

3.2. Feedwater heater GMDH modeling for leak detection

In this study, a high-pressure feedwater heater (3A) at an operating coal-fired power plant is selected for performance monitoring using the proposed method. Analysis is carried out on a total of 9-month feedwater heater database acquired from the heater train A of the plant. Data are acquired at 1-minute

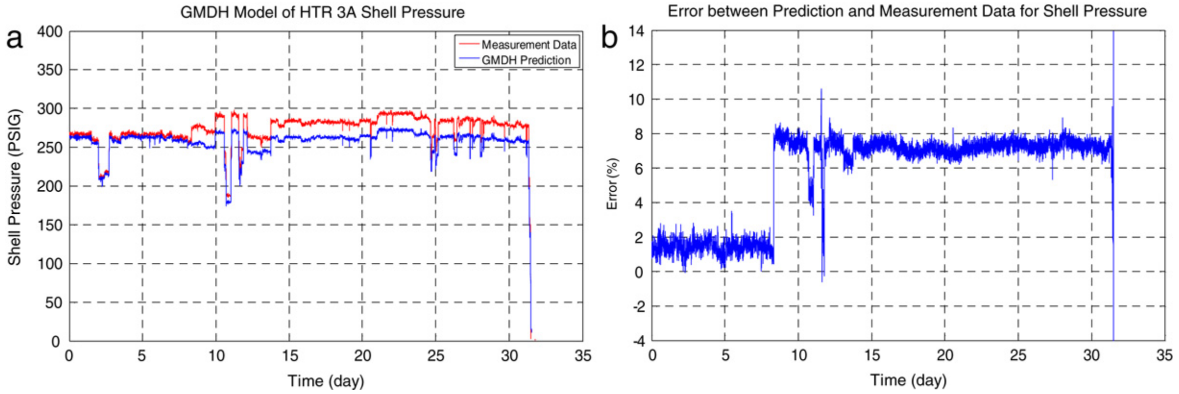

Fig. 6. GMDH modeling of heater shell pressure: (a) Comparison of measured and GMDH-estimated values. (b) Residuals between measured and GMDH-estimated values.

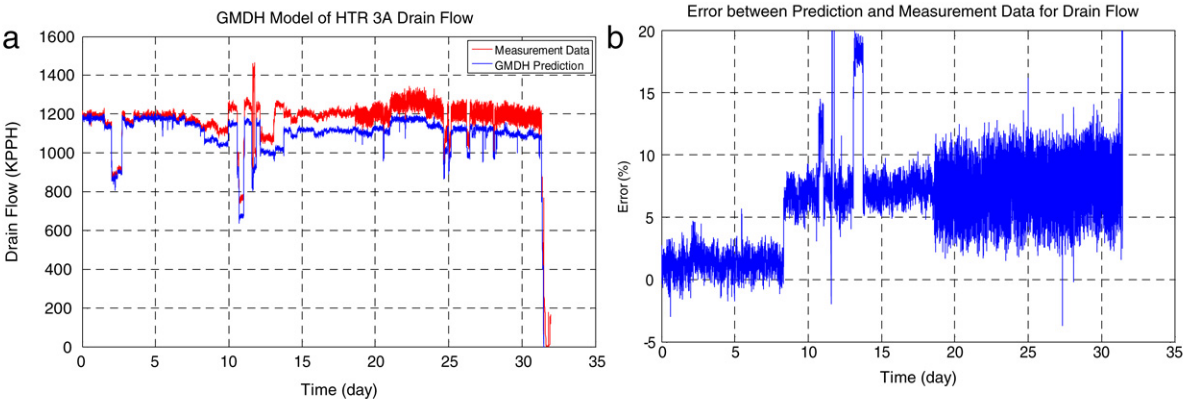

Fig. 7. GMDH modeling of heater drain flow: (a) Comparison of measured and GMDH-estimated values. (b) Residuals between measured and GMDH-estimated values.

sampling interval. Ten important variables are grouped together for Heater 3A monitoring, as shown in Table 1. The selection of these variables from the entire feedwater heater database is based on the understanding of the system characteristics and the expertise from plant personnel. In addition, correlation coefficients are also used as measures of selection. The response variables of the GMDH models and the associated fault signatures, that are expected residual deviation patterns, are shown in Table 1 as well. If the actual value of a signal is greater than (less than) the model prediction, then the residual value is positive with "↑" sign (or negative with "↓" sign). Heater shell side steam pressure and heater drain flow serve as the two most important indicators of a rupture in the interior tubing of the feedwater heater. When a rupture of the tubing occurs, the first thing that happens is that the steam pressure increases almost instantly because the feedwater, which flows through the tubing of the heater, will be at a pressure several times higher than that of the steam space on the shell side. When the feedwater enters the steam space, as a result of rupture or leakage, the difference in pressures causes the feedwater to flash into wet steam and causes a rapid pressure build up on the shell side of the heater. Drain flow, in the mean time, increases due to additional water entering the shell side.

The first step in the analysis is to collect training data from the system for normal operating conditions. The training data we use in this study cover a time-span of 30 days right after a major overhaul of the plant; and it is assumed that the system is under normal operating conditions. Table 2 shows the response variables and the corresponding predictors selected for GMDH model development. For each chosen set of input functional terms given in Eq. (4), the GMDH algorithm is processed to find the model that best maps the input/output relationships. The functional form of input terms used in each model, as well as the number of layers used to build the model, are also shown in Table 2.

Table 1
Fault signature chart for feedwater heaters.

Description	Units	Tube leak
Gross Generation	MW	
HTR 3A Drain Temp	DEGF	
HTR 3A Shell Pressure	PSIG	↑
HTR 3A Drain Flow	KPPH	↑
HTR 3A Extraction Pressure	PSIG	
Extraction to HTR 1A Temp	DEGF	
Extraction to HTR 2A Temp	DEGF	
HTR 3A Inlet Water Temp	DEGF	
HTR 3A Outlet Water Temp	DEGF	
HTR 3A DCA	DEGF	

After an optimal model is obtained, we can apply the model to the data that are not used for training and compute the residual values between the estimated values and the actual measurements. Fig. 6 and 7 show the GMDH estimates and the actual measurements as well as the residuals between them for heater shell side steam pressure and heater drain flow, respectively.

It is observed that significant residuals start registering on shell-side pressure on the 8th day of monitored operation period. The residuals between the measured shell pressure of HP heater 3A and the model estimates increase immediately by 10%, which is about 25 pounds per square inch. At the same time, there are high residuals on drain flow. These abnormal conditions persisted for about three weeks before the unit was eventually shut down for maintenance. It was later revealed that the high-pressure feedwater heater 3A had experienced ruptures in the interior tubing. Therefore, an extensive leak from the tubes had occurred, causing a process upset.

It is interesting to note that in Fig. 7(b) there are larger fluctuations in the residuals of drain flow starting around the 18th

Table 2
Process measurements used in GMDH models for feedwater heaters.

Modeled variable	Input measurements	Input functional terms	Layer number
Shell-side steam pressure	Gross generation, extraction steam pressure, and extraction steam temperature	$\left\{ \begin{array}{c} (1, x_1, x_2), (x_1^2, x_2^2), \left(\frac{1}{x_1}, \frac{1}{x_2}\right), \\ \left(\frac{x_1}{x_2}, \frac{x_2}{x_1}\right), \left(\frac{x_1}{x_1+x_2}, \frac{x_2}{x_1+x_2}\right), (x_1x_2) \end{array} \right\}$	3
Heater drain flow	Gross generation, drain temperature, extraction steam pressure, and feedwater inlet temperature	$\left\{(1, x_1, x_2), (x_1^2, x_2^2), \left(\frac{1}{x_1}, \frac{1}{x_2}\right), \left(\frac{1}{x_1+x_2}\right)\right\}$	7

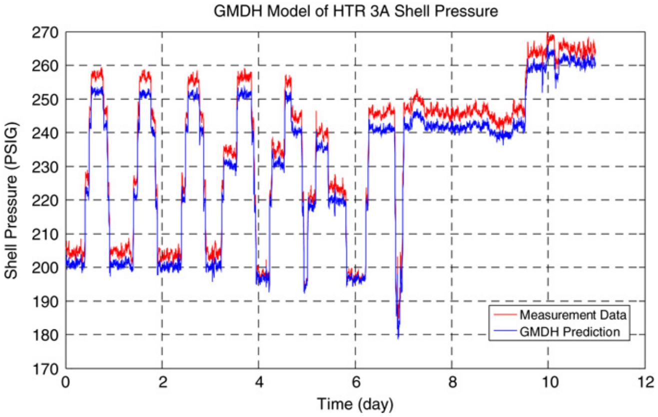

Fig. 8. GMDH estimates of shell-side pressure for post-maintenance data.

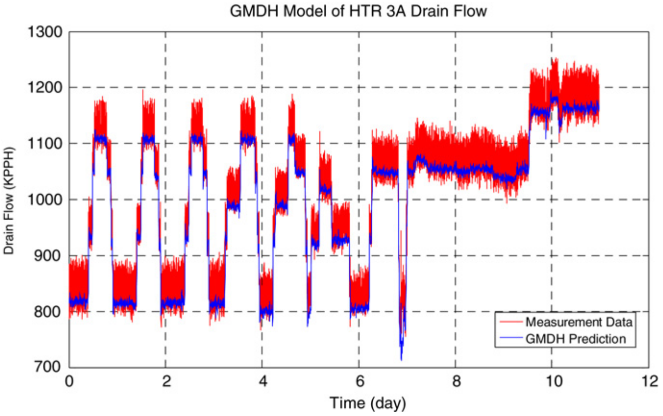

Fig. 9. GMDH estimates of drain flow for post-maintenance data.

day during the period. Since the analysis of the data following the maintenance shows that fluctuations still exist on the drain flow measurements, it is concluded that fluctuations are not caused by tube leakage.

Now that the results of using GMDH for heater tube leak detection have been presented, the GMDH models are applied to the data after this problem has been fixed. Fig. 8 and 9 shows that the estimation of the drain flow signal closely follows the measurement, even though there is a small offset, which is about 2% in residuals compared to the residual of less than 0.5% in initial model training results. In fact, there is no problem associated with the feedwater heater being monitored. Relatively high residuals result from a new operating strategy used by the plant, that is, the power output is adjusted as demand for electricity fluctuates throughout a day. This probably indicates a need for model retraining for each of the indicated variables.

Retraining for a new operating mode either expands the existing training dataset with additional data representing the new operating state, or completely replaces the existing training data with the newly acquired data if a permanent change in the process operation is achieved. For fossil power plant applications, an on-line monitoring system acquires operation data from the data historian at periodic intervals for each model and stores them in a designated location. If model retraining is needed, the on-line monitoring system combines the new data with previously acquired data, automatically runs each model with the new data included, and stores the run results. For the sake of computational efficiency, the polynomial structures (the chosen combination of input functional terms) of the previous GMDH models may be retained. Only the model coefficients need to be updated for the new process operating mode in order to take into account both the previous network knowledge and the current knowledge extracted from the new training data.

4. Concluding remarks

A model-based fault diagnosis method using an enhanced version of the GMDH approach has been developed for monitoring process units of a fossil power plant. The proposed algorithm performs a fault detection task by examining the residuals between the actual behavior of the process measurements and their estimates using a GMDH model of normal operation of the unit. The fault interpretation is conducted by incorporating fault signature analysis, which is based on patterns of residual deviation. Ridge regression is integrated into the conventional GMDH algorithm to estimate the best model structure. The improved GMDH can significantly avoid potential ill-conditioned problems associated with collinearity.

The performance of the proposed fault diagnostic system has been tested for the diagnostics of the feedwater heater units of a fossil power plant using real, operational data acquired by the power plant DCS system. It is shown that, using the proposed leak detection method and the feedwater heater GMDH models, water leakage into heater steam space can be easily detected, and an early warning can be provided to plant operators. The fault detection results of feedwater heater application demonstrate the capability of the proposed data-based technique for modeling long-duration process data, characterization of their relationships, and for continuous monitoring of process units.

Although the GMDH is a powerful modeling tool, it does have certain drawbacks in comparison with other soft computing techniques. The large computational overhead associated with enumerating all the possible nodes in each layer could hinder some real-world implementations. The use of a genetics-based GMDH network construction algorithm is currently being investigated. It is anticipated that the computational leverage of the genetic algorithm will provide a means to improve the computational efficiency associated with the GMDH procedure.

Acknowledgments

The research and development reported here was supported in part by Emerson Process Management through a contract with the University of Tennessee, Knoxville.

References

[1] Chow EY, Willsky AS. Analytical redundancy and the design of robust detection system. IEEE Trans Automat Control 1984;29:603–14.

[2] Chen J, Patton RJ. Robust model-based fault diagnosis for dynamic systems. Dordrecht: Kluwer Academic Publishers; 1999.

[3] Patton RJ, Chen J. Observer based fault detection and isolation: Robustness and application. Control Eng Pract 1997;5:671–82.

[4] Gertler JJ. Fault detection and diagnosis in engineering systems. New York: Marcel Dekker; 1998.

[5] Ljung L. System identification: Theory for the user. Englewood Cliffs, NJ: Prentice-Hall; 1987.

[6] Zhang X, Polycarpou MM, Parisini T. A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. IEEE Trans Autom Control 2002;47(4):576–93.

[7] Kaistha N, Upadhyaya BR. Incipient fault detection and isolation of field devices in nuclear power systems using principal component analysis. Nucl Technol 2001;136:221–30.

[8] Zhao K, Upadhyaya BR. Adaptive fuzzy inference causal graph approach to fault detection and isolation of field devices in nuclear power plants. Prog Nucl Energy 2005;46:226–40.

[9] Hild CR. Development of the group method of data handling with information-based model evaluation criteria: A new approach to statistical modeling. Ph.D. Dissertation. University of Tennessee, Knoxville, 1998.

[10] Hines JW, Seibert R. Technical review of on-line monitoring techniques for performance assessment. Report prepared for the Nuclear Regulatory Commission. NUREG/CR-6895. May 2008.

[11] Ivakhnenko AG. Group method of data handling — a rival of the method of stochastic approximation. Soviet Autom Control 1966;13:43–71.

[12] Farlow SJ. Self-organizing methods in modeling: GMDH-type algorithms. New York: Marcel-Dekker; 1984.

[13] Lu B, Upadhyaya BR. Monitoring and fault diagnosis of the steam generator system of a nuclear power plant using data-driven modeling and residual space analysis. Ann Nucl Energy 2005;32:897–912.

[14] Upadhyaya BR, Li F, Samardzija N, Kephart R, Coffey L. Development of data-driven modeling methods for monitoring coal pulverizer units in power plants. In: Proceedings of ISA POWID symposium. 2007.

[15] Abdel-Aal RE. Predictive modeling of mercury speciation in combustion flue gases using GMDH-based abductive networks. Fuel Processing Technol 2007; 88:483–91.

[16] Ferreira PB. Incipient Fault Detection and Isolation of Sensors and Field Devices. Ph.D. dissertation. Knoxville: University of Tennessee; 1999.

[17] Tikhonov AN, Arsenin VA. Solution of Ill-posed Problems. Washington: Winston & Sons; 1977.

[18] Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 1970;12:55–67.

[19] Draper NR, Van Nostrand RC. Ridge regression and James–Stein estimation: Review and comments. Technometrics 1979;21:451–66.

[20] Hoerl AE, Kennard RW, Baldwin KF. Ridge regression: Some simulations. Comm Statist Theory Methods 1975;4:105–23.

[21] Lawless JF, Wang P. A simulation study of ridge and other regression estimators. Comm Statist Theory Methods 1976;5:307–23.

[22] Kavaklioglu K. Optimal fuzzy control design using simulated annealing and application to feedwater heater control. Ph.D. dissertation. Knoxville: University of Tennessee; 1996.