Korean J. Chem. Eng., **30**(3), 528-539 (2013)
DOI: 10.1007/s11814-012-0197-7

Adaptive neuro-fuzzy inference system based faulty sensor monitoring of indoor air quality in a subway station

Hongbin Liu*, Mingzhi Huang, Jeong Tai Kim***, and ChangKyoo Yoo*,†**

*Department of Environmental Science and Engineering, Center for Environmental Studies, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Korea
**College of Environmental Science and Engineering, South China University of Technology, Guangzhou 510640, P. R. China
***Department of Architectural Engineering, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Korea
(*Received 30 January 2012 • accepted 7 November 2012*)

Abstract−A new faulty sensor monitoring method based on an adaptive neuro-fuzzy inference system (ANFIS) is proposed to improve the monitoring performance of indoor air quality (IAQ) in subway stations. To enhance network performance, a data preprocessing step for detecting outliers and treating missing data is implemented before building the monitoring models. A squared prediction error (SPE) monitoring index based on the ANFIS prediction model is proposed to detect sensor faults, where the confidence limit for the SPE index is determined by using the kernel density estimation method. The proposed monitoring approach is applied to detect four typical kinds of sensor faults that may happen in the indoor space of a subway. The prediction results in the subway system indicate that the prediction accuracy of an ANFIS structure with 15 clusters is superior to that of an appropriate artificial neural network structure. Specifically, when detecting one kind of complete failure fault that happened within the normal range, the detection performance of ANFIS-based SPE outperforms that of a traditional principal component analysis method. The developed sensor monitoring technique could work well for other kinds of sensor faults resulting from a noxious underground environment.

Key words: Adaptive Neuro-fuzzy Inference System (ANFIS), Indoor Air Quality, Kernel Density Estimation, Sensor Fault Detection, Subway Systems

INTRODUCTION

Subways may be the most convenient form of commuter transportation, with more than eight million people choosing this mode of transportation daily in Seoul [1]. Given this volume of commuter traffic, the indoor air quality (IAQ) of subway stations is very important to public health. To effectively monitor IAQ measurements in subway stations, the Korean Ministry of Environment has established IAQ regulations for some hazardous pollutants, including carbon monoxide (CO) [1], particulate matter with diameters less than 10 (PM_{10}) and 2.5 μm ($PM_{2.5}$) [1,2], aromatic volatile organic compounds [3], airborne fungi [4] and carbonyl compounds [5]. Excessive exposure to these pollutants may cause serious consequences, such as respiratory symptoms and lung malfunction and even mortality [6]. Therefore, monitoring of IAQ in subway stations has become an important issue of public concern.

Sensors that monitor IAQ are important components in subway systems, but unfortunately, these sensors sometimes exhibit poor quality and low reliability because of the duration of usage and the hostile underground environment, such as low signal to noise ratio, unstable electricity supply, high speed trains, dusty grounds and so on. Therefore, sensors may experience bias, drifting, complete failure or precision degradation (Fig. 1) [7,8]. When inaccurate or even completely incorrect IAQ measurements are used for monitoring and controlling the IAQ and ventilation systems, faulty sensors can be quite detrimental to data-driven decision-making techniques. In a worst case scenario, due to sensor malfunction or communication problems, IAQ sensor data may not be available from the data collection system. These problems make it difficult to extract and inter-

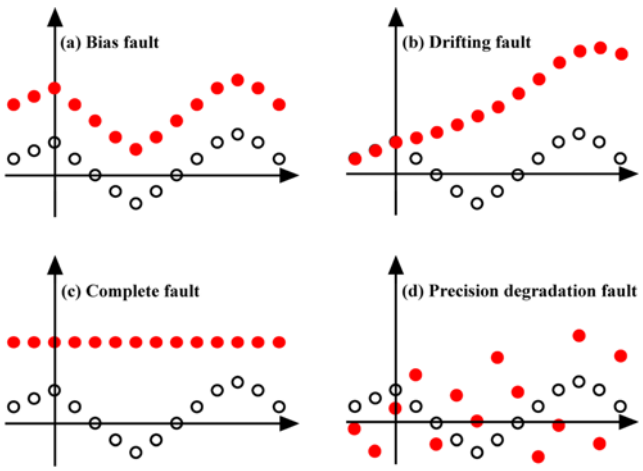

Fig. 1. Four types of sensor faults: (a) bias, (b) drifting, (c) complete failure and (d) precision degradation. The open circles represent normal sensor measurements and the filled dots represent faulty sensor measurements.

†To whom correspondence should be addressed.
E-mail: ckyoo@khu.ac.kr

pret information from process data. Consequently, monitoring or control based on incorrect IAQ measurements may become problematic or even dangerous. Therefore, the timely detection of these IAQ sensor faults plays a key role in the successful management and control of IAQ in subway systems. Recently, many methods have been developed to achieve early and accurate fault detection demand.

According to the classification of process fault detection and diagnosis by Venkatasubramanian et al. [9], data-driven process monitoring methods can be classified into statistical and neural networks-based methods. Multivariate statistical process monitoring techniques such as principal component analysis (PCA) and partial least squares (PLS) have been widely applied to the on-line process monitoring of air quality management systems [10-13]. However, when dealing with highly nonlinear and collinear data sets, the accuracy of these linear statistical models may decrease significantly [14,15]. To solve the problems posed by the nonlinear feature of process measurements, some nonlinear statistical process monitoring approaches, such as dynamic PCA (DPCA) [16,17], kernel PCA (KPCA) [18-20] and neural network-based PCA (NNPCA) [21-23], have been proposed. The key idea of KPCA is that the nonlinear data structure in the input space is more likely to be linear after high-dimensional nonlinear mapping. As a nonlinear extension of PCA, KPCA first maps the input space into a feature space, and then extracts the principal components in the feature space.

Conversely, neural networks-based methods are primarily used for fault diagnosis or for classification other than fault detection [9, 24]. However, neural network schemes still have several limitations resulting from possibly getting trapped in a local minimum and the choice of model architecture. Some recent studies have used an adaptive neuro-fuzzy inference system (ANFIS), which has proven to be an effective tool to approximate any nonlinear functions [25]. Combining the theories of fuzzy logic and neural networks can make effective use of the easy interpretability of fuzzy logic, as well as the superior learning ability and adaptive capability of neural networks. As a result, ANFIS has been successfully applied in many engineering fields [26-29]. Although there has been a great deal of monitoring research of IAQ, to our knowledge, there is no literature on the monitoring of IAQ in subway stations using ANFIS.

In this paper, an ANFIS-based monitoring method was developed to further improve sensor fault detection performance in subway stations. The remainder of this paper is organized as follows. In Section 2, we briefly present the computational formulas of ANFIS and PCA. Then, we introduce a monitoring index and a non-parametric method to determine the confidence limit. An integrated scheme showing the main steps to obtain better fault detection performance is highlighted at the end of this section. In Section 3, we evaluate the modeling and monitoring performance of ANFIS using a data set collected from a subway station in Seoul. The IAQ data pretreatment that mainly covers issues about dealing with outliers and missing data is also given. Finally, we present our conclusions.

METHODS

1. ANFIS Prediction Model

ANFIS is a multilayer feed-forward network that uses neural network learning algorithms and fuzzy reasoning to map inputs into an output. It is a fuzzy inference system (FIS) implemented in the framework of adaptive neural networks [30]. Fig. 2 shows the architecture of a typical ANFIS with two inputs, two rules and one output using the Takagi-Sugeno-Kang (TSK) model [31,32], where each input is assumed to have two membership functions (MFs).

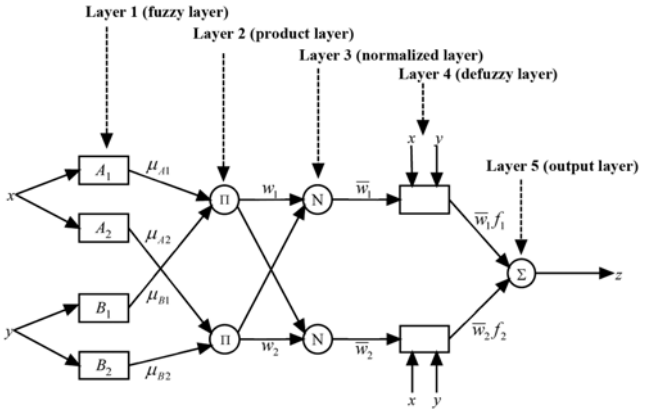

Fig. 2. ANFIS structure of a two-input TSK model with four rules.

The function of each ANFIS layer in Fig. 2 is summarized and explained as follows. For layer 1, all nodes are adaptive nodes that can generate membership values for inputs. The outputs of this layer are given by

$$Q^1_{Ai}=\mu_{Ai}(x),\ i=1,2$$
$$Q^1_{Bj}=\mu_{Bj}(y),\ j=1,2 \tag{1}$$

where x and y are crisp inputs, and A_i and B_j are fuzzy sets characterized by MFs with low, medium, and high values. The widely used MFs are triangular, trapezoidal, or Gaussian functions. The current study utilized Gaussian type MFs:

$$\mu_{Ai}(x)=e^{-\frac{1}{2}\left(\frac{x-a_i}{b_i}\right)^2},\ i=1,2$$
$$\mu_{Bj}(y)=e^{-\frac{1}{2}\left(\frac{y-a_j}{b_j}\right)^2},\ j=1,2 \tag{2}$$

For layer 2, the nodes are fixed, which are used as a simple multiplier. The outputs of this layer are represented by

$$O^2_{ij}=w_{ij}=\mu_{Ai}(x)\mu_{Bj}(y),\ i,j=1,2 \tag{3}$$

which represent the firing strength of each rule. The firing strength means the degree to which the antecedent part of the rule is satisfied.

For layer 3, the nodes are also fixed, indicating that they play a normalization role in the network. The outputs of this layer, which are called normalized firing strengths, can be represented as follows:

$$O^3_{ij}=\overline{w}_{ij}=\frac{w_{ij}}{w_{11}+w_{12}+w_{21}+w_{22}},\ i,j=1,2 \tag{4}$$

For layer 4, the parameters in this layer are referred to as consequent parameters. Each node is an adaptive node, and its output is simply the product of the normalized firing strength and a first-order polynomial. The outputs of this layer are given by

$$O^4_{ij}=\overline{w}_{ij}f_{ij}=\overline{w}_{ij}(p_{ij}x+q_{ij}y+r_{ij}),\ i,j=1,2 \tag{5}$$

where p_{ij}, q_{ij} and r_{ij} are consequent parameters of the first-order poly-

nomial.

For layer 5, the single node in this layer is a fixed node labeled **Σ**, which computes the overall output as the summation of all incoming signals

$$Z = O_j^5 = \sum_i \overline{w}_i f_i, \quad i = 1, 2 \tag{6}$$

By combining the gradient descent optimization method and the least squares method, the hybrid learning algorithm could effectively improve the prediction performance of ANFIS. Thus, this algorithm was used to tune the adjustable parameters in this study.

To assess the prediction capability of different ANFIS models, several performance indices including mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and squared correlation coefficient (R^2) are defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(a_i - p_i)^2 \tag{7}$$

$$RMSE = \sqrt{MSE} \tag{8}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|100 \times \frac{a_i - p_i}{a_i}\right| \tag{9}$$

$$R^2 = \frac{cov(a,p) \times cov(a,p)}{cov(a,a) \times cov(p,p)} \tag{10}$$

where a_i is the experimental value, p_i is the predicted value, N is the number of data and is the covariance between a and p sets.

2. Principal Component Analysis (PCA)

PCA models can take into account the correlation between the process variables in a sensor validation context. PCA decomposes the normalized matrix $\mathbf{X}$ into a score matrix $\hat{\mathbf{T}} \in R^{n \times d}$ and a loading matrix $\hat{\mathbf{P}} \in R^{m \times d}$ by eigenvalue decomposition or by singular value decomposition (SVD) as follows [15]:

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} \tag{11}$$

$$\hat{\mathbf{X}} = \hat{\mathbf{T}}\hat{\mathbf{P}}^T \tag{12}$$

$$\mathbf{E} = \tilde{\mathbf{T}}\tilde{\mathbf{P}}^T \tag{13}$$

$$\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{P}} \tag{14}$$

where $\hat{\mathbf{X}}$ is the principal component subspace (PCS) that represents the correct direction of the measured vectors and $\mathbf{E}$ is the residual subspace (RS) that is the direction of the faulty measurements. Value d is the number of principal components (PCs).

The columns of $\hat{\mathbf{P}}$ can be obtained by calculating the eigenvectors of the covariance matrix Σ, and $\hat{\mathbf{P}}$ contains d largest eigenvalues when using the eigenvalue decomposition. The columns of $\tilde{\mathbf{P}}$ are obtained by using the remaining m−d eigenvectors. Then the projection matrices $\mathbf{C}$ and $\tilde{\mathbf{C}}$ are calculated from Eqs. (15) and (16), respectively:

$$\mathbf{C} = \hat{\mathbf{P}}\hat{\mathbf{P}}^T \tag{15}$$

$$\tilde{\mathbf{C}} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \mathbf{I} - \mathbf{C} \tag{16}$$

After obtaining the PCA model, a new sample vector $\mathbf{x} \in R^m$ (this new sample vector needs to be normalized using the mean and variance of the training data set before further calculation) can be decomposed into two parts:

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e} \tag{17}$$

where

$$\hat{\mathbf{x}} = \mathbf{C}\mathbf{x} \tag{18}$$

is the projection of the sample vector in the principal component subspace, and

$$\mathbf{e} = \tilde{\mathbf{C}}\mathbf{x} = (\mathbf{I} - \mathbf{C})\mathbf{x} \tag{19}$$

is the projection of the sample vector in the residual subspace.

To detect sensor faults, three sets of statistics, including a squared prediction error (SPE) statistic and Hotelling's T^2 statistic, are normally used for fault detection. The SPE index, which is a measure of the faulty variations outside the PCA model space, can be calculated as follows:

$$SPE(\mathbf{x}) = \|\mathbf{e}\|^2 = \|\tilde{\mathbf{C}}\mathbf{x}\|^2 = \mathbf{x}^T(\mathbf{I} - \mathbf{C})\mathbf{x} \tag{20}$$

The upper confidence limit for the SPE can be calculated from an approximate distribution as follows [33]:

$$SPE_{lim} = \theta_1\left[\frac{c_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2}\right]^{1/h_0} \tag{21}$$

$$h_0 = 1 + \frac{2\theta_1\theta_3}{3\theta_2^2}, \quad \theta_1 = \sum_{i=d+1}^{m}\lambda_i, \quad \theta_2 = \sum_{i=d+1}^{m}\lambda_i^2, \quad \theta_3 = \sum_{i=d+1}^{m}\lambda_i^3 \tag{22}$$

where c_α is the standard normal deviate corresponding to the upper $(1-\alpha)$ percentile [34] and λ_j is the eigenvalue associated with the j^{th} loading vector of the covariance matrix Σ.

Another detection index commonly used is Hotelling's T^2 statistic, which is a measure of the variations in the PCS, and is defined as:

$$T^2 = \mathbf{x}^T\hat{\mathbf{P}}\Lambda^{-1}\hat{\mathbf{P}}^T\mathbf{x} = \mathbf{x}^T\mathbf{D}\mathbf{x} \tag{23}$$

where Λ=diag{λ_1, λ_2, ⋯, λ_m} is the diagonal matrix of the eigenvalues associated with the retained principal components. The upper confidence limit for T^2 can be computed from a χ^2 distribution with d degrees of freedom:

$$T_{lim}^2 = \chi_\alpha^2(d) \tag{24}$$

Because the T^2 statistic is a measure of the deviation in the residual space, it can be used to identify when the current process state deviates from the expected normal space. On the other hand, an SPE statistic will be more sensitive to the regular fluctuations that move the process away from a normal state. In practice, a fault is detected if either a T^2 or SPE statistic triggers an abnormal alarm. Note that the number of PCs could significantly affect the accuracy of sensor fault identification and reconstruction. The reconstruction-based unreconstructed variance (URV) method [35] was used in order to determine the optimal number of PCs for the best reconstruction.

3. A Monitoring Index Based on the ANFIS Model

The ANFIS modeling method cannot directly provide a monitoring index to detect sensor faults. Therefore, a new nonlinear monitoring index needs to be developed. In this study, a SPE type index based on the ANFIS prediction model was proposed. This monitoring index is calculated by using the following equation:

$$SPE_{ANFIS} = \|\mathbf{y}_{measured} - \mathbf{y}_{predicted}\|_2^2 \tag{25}$$

where SPE_{ANFIS} is a SPE value of the ANFIS prediction model, $\mathbf{y}_{measured}$ is a measured value of IAQ and $\mathbf{y}_{predicted}$ is a predicted value of the ANFIS model.

Because SPE_{ANFIS} values do not usually follow Gaussian distribution, their confidence limit (CL), if calculated from the assumption of Gaussian distribution, may result in imprecise monitoring performance. The kernel density estimation (KDE) method, which can tackle a non-Gaussian distribution problem, is used to determine the confidence limit of SPE_{ANFIS} [36,37]. The idea behind KDE is that the point covering a 99% range of density function will be read off and can be used as a CL. In this study, a univariate kernel density estimator was used to calculate the CL. This estimator is defined as follows:

$$\hat{f}(x,h)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right) \quad (26)$$

where x is the value used to calculate KDE, h is the bandwidth matrix (also called the smoothing parameter) and K is a kernel function (Gaussian kernel function, which is often used in practice, was hence applied in this work).

4. Integrated Scheme for Monitoring IAQ Sensors Using ANFIS

The framework of the proposed nonlinear faulty sensor monitoring of the ANFIS model in a subway station is shown in Fig. 3. The implementation of the proposed method was divided into two main parts: build an ANFIS model and monitor IAQ using this model to detect the sensor faults. Within the ANFIS modeling part, the measured data were first collected from a subway station. Then, the collected original data were pre-processed by checking the outliers and missing data. The checked data needed to be normalized and labeled as training data. This normalized data were used to develop an ANFIS model as a training dataset, where the residual signals are generated by comparing the predicted values with the measured data. Within the sensor monitoring part, the squared prediction error (SPE) was used to detect sensor faults by comparing its current value with the confidence limit determined by kernel density estimation. Note that this study mainly concentrated on the sensor fault detection step. After sensor faults are detected by our proposed method, a PCA or PLS model can be used for fault identification and reconstruction steps.

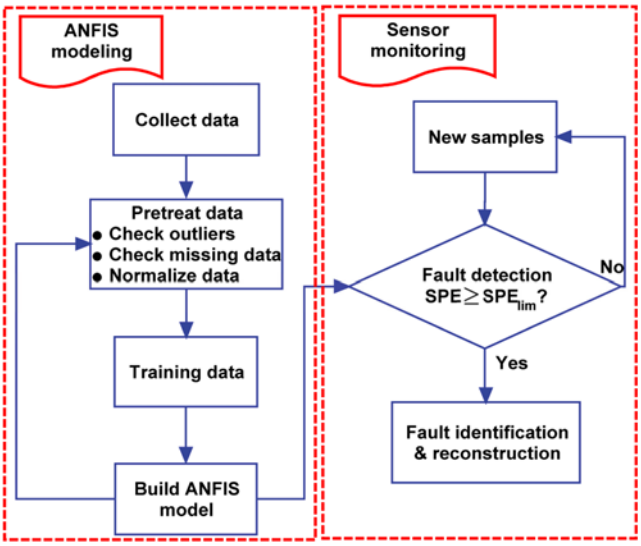

Fig. 3. Steps used in ANFIS monitoring to enhance the monitoring performance of faulty sensors.

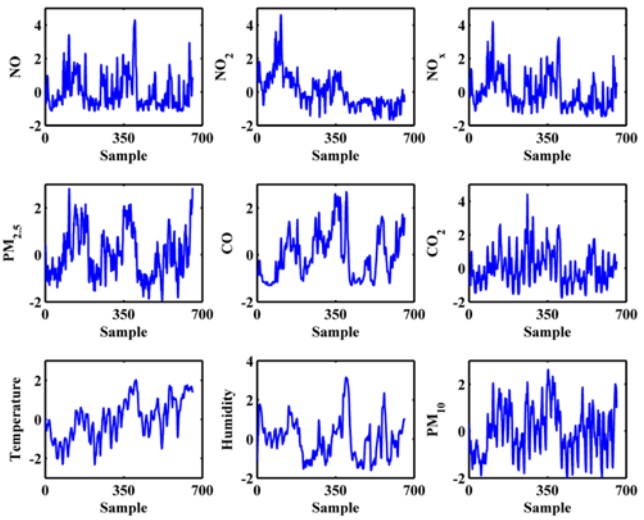

Fig. 4. Variations of IAQ data from a Seoul subway station.

RESULTS AND DISCUSSION

1. Pre-processing of IAQ Data

This subsection gives a short description of the IAQ data collected by a tele-monitoring system (TMS). Four weeks of air pollutant data were collected in real-time from a TMS in a Seoul subway station. The measured variables were the concentrations of nitrogen monoxide (NO), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), PM_{10}, $PM_{2.5}$, CO, carbon dioxide (CO_2), temperature and humidity on the platform (Fig. 4). All of the variables were scaled to zero mean and unit variance to improve network performance.

Outliers and missing data could have an uncertain effect on model accuracy, especially on data-driven models such as an ANFIS model. Although the PCA method can be used directly to extract outliers by observing Hotelling's T^2 and SPE plots [14], a more advanced detection method, namely, Jolliffe's three-parameter method [38], was effectively used for detecting outliers. Dealing with missing data is another important step before developing the ANFIS model to detect sensor faults. There are many methods that can be used to treat missing data [39-41]. Although a listwise deletion method is the simplest approach, it can delete some important information. On the other hand, a substitution method that replaces missing data with reasonable approximations could be regarded as a more suitable method. Therefore, the substitution method was adopted to deal with possible missing data in this study.

The hourly mean values for each variable from January 4 to 31, 2010, with a total number of 650 observations, were evaluated in this study. Of the total data, 70% were used as training data, the next 15% of total data were used as validation data and the remaining data were used as test data to verify the proposed sensor fault monitoring method. Jolliffe's three-parameter method was used to find the potential outliers. Fig. 5 clearly shows that sample 162 was detected as one outlier because all of the Jolliffe parameters were extremely high compared to their threshold values. After substituting

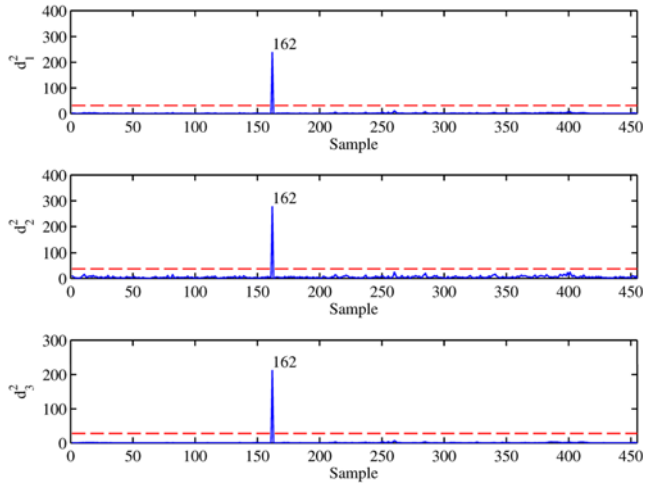

Fig. 5. Outlier detection by Jolliffe's three-parameter method.

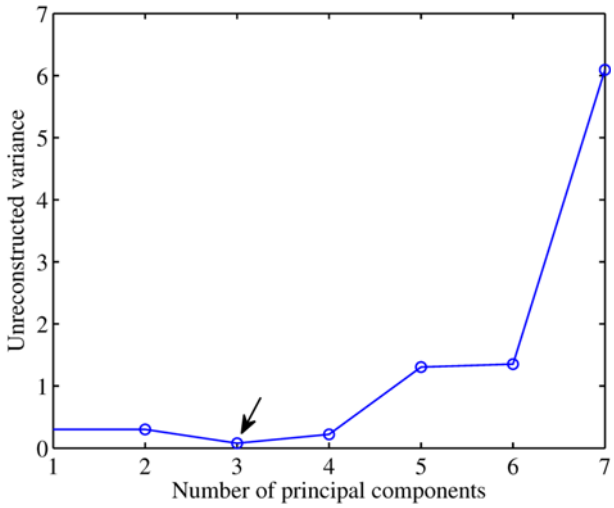

Fig. 6. Determination of the optimal number of principal components based on the unreconstructed variance.

this abnormal point with a reasonable value (for example, a mean value), new training data was available for building the ANFIS model.

An important step in the reconstruction-based sensor validation approach is the determination of the number of principal components (PCs) for the best reconstruction. Fig. 6 shows the variation in the URV versus the number of PCs. The URV was minimal when three PCs were chosen. This indicates that the optimal number of PCs would be three. More specifically, the first three PCs corresponding to the three largest eigenvalues could explain 85.68% of the total variance in the system. Therefore, three PCs were used to develop the PCA-based detection and reconstruction model for faulty sensors.

2. Development of the ANFIS Prediction Model

The subtractive clustering algorithm [30] of Matlab's fuzzy logic toolbox was used to determine the initial parameters for membership functions and fuzzy if-then rules. Compared to the grid partition algorithm, the subtractive clustering algorithm is more suitable for modeling complex systems with a large number of input variables. By setting the range of the cluster center, several cluster centers were obtained. The cluster centers represent the initial parameters of input MFs. Using this subtractive clustering method, the total number of fuzzy rules only depends on the number of clusters.

Particulate matter (PM) from subway systems has been proven to be relatively toxic and their presence may have potential health effects on commuters [42]. Therefore, particulate matter should be monitored by using real-time IAQ sensors in subway stations. In this study, PM_{10} was used as a dependent variable (output variable), and the other eight variables were chosen as the independent variables (input variables). The structure of ANFIS was identified to represent the process model for detecting sensor faults. The structure parameters of the ANFIS model are given in Table 1. This model has five layers with eight nodes in the input layer and one node in the output layer. The second layer that contains 120 (8×15) nodes is used to calculate the MFs, the third layer with 15 nodes is the rule layer and the fourth layer with 15 nodes is the normalization calculating layer. In this study, the Gaussian type membership function, which is one of the most widely used fuzzy membership functions for modeling of high-dimensional systems, was utilized to train the network. After the ANFIS model was trained with the parameters listed in Table 1, the inference was performed according to 15 fuzzy linguistic rules for modeling PM_{10}. After determining the initial value of the premise parameter and the architecture of the predictive model, we trained the network by a hybrid algorithm. Then the premise and consequent parameters of the network were pruned. Membership functions of the variables were drawn after the premise parameter was obtained. Fig. 7(a) shows the shapes of eight input MFs before training, whereas Fig. 7(b) shows those shapes after training. Fig. 8 illustrates the three-dimensional graphic surfaces of the defuzzified results.

Because the number of clusters has a significant effect on the

Table 1. Basic structures of ANFIS and ANN

ANFIS		ANN	
Basic structure		Basic structure	
No. of total layers	5	No. of total layers	3
No. of layers without input and output layers	3	No. of hidden layers	1
No. of nodes in input layers	8	No. of neurons in input layers	8
No. of nodes in output layers	1	No. of neurons in output layers	1
No. of fuzzy rules	15	No. of neurons in hidden layers	10
Shape of MFs	Gaussian	Training algorithm	Trainlm
No. of training	1000	No. of training	2000

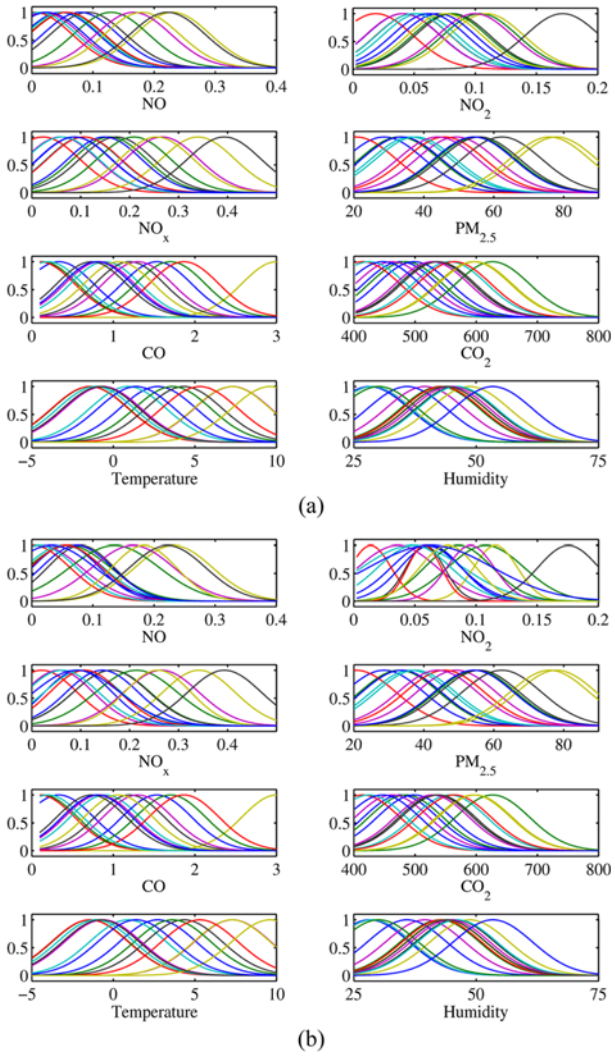

Fig. 7. The input membership functions of IAQ measurements: (a) before training and (b) after training.

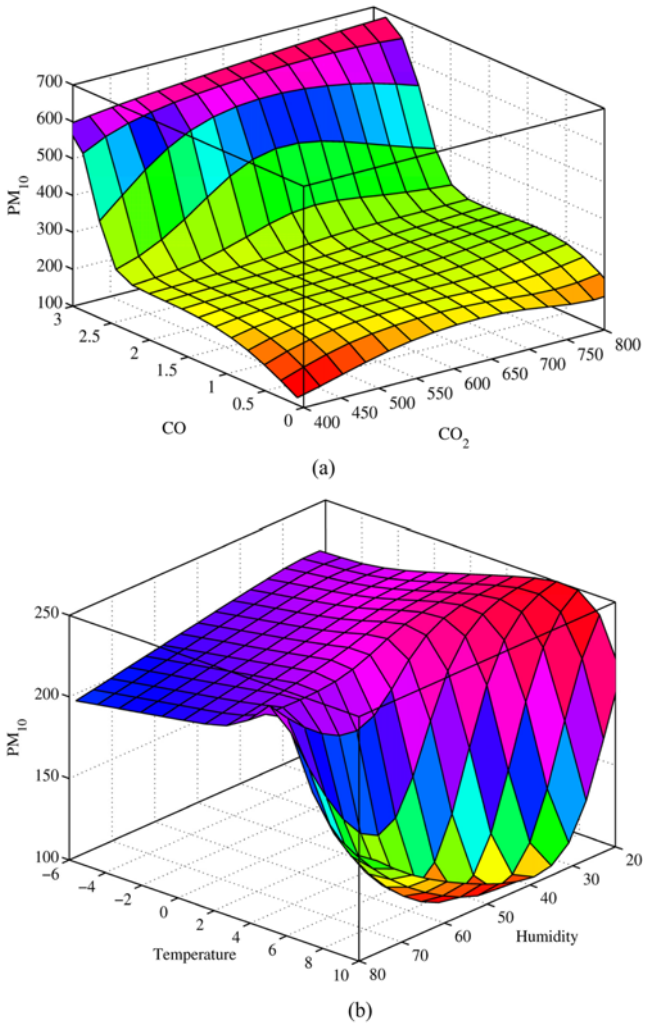

Fig. 8. Three-dimensional representation of the PM_{10} response surface graph in terms of: (a) CO_2 and CO and (b) humidity and temperature.

Table 2. Results of prediction performance using ANFIS and ANN models

		ANN	ANFIS with 6 clusters	ANFIS with 9 clusters	ANFIS with 15 clusters
MSE	Train	87.8715	79.6714	64.6345	45.1917
	Validation	103.4634	74.9187	78.7594	82.0601
	Test	90.6563	108.5579	127.6964	102.6187
	Overall	90.6421	83.3100	76.2719	59.4085
RMSE	Train	9.3740	8.9259	8.0396	6.7225
	Validation	10.1717	8.6556	8.8746	9.0587
	Test	9.5214	10.4191	11.3003	10.1301
	Overall	9.5206	9.1274	8.7334	7.7077
MAPE	Train	11.2831	10.6023	9.1254	7.7823
	Validation	12.1548	11.4657	10.1981	10.3895
	Test	11.4553	12.1069	12.2770	10.8822
	Overall	11.4405	10.9593	9.7623	8.6427
R^2	Train	0.8985	0.9073	0.9247	0.9473
	Validation	0.8740	0.9088	0.9048	0.9016
	Test	0.8989	0.8785	0.8629	0.8889
	Overall	0.8953	0.9031	0.9116	0.9312

prediction capability of ANFIS, three ANFIS models with different numbers of clusters were implemented. Furthermore, an artificial neural network (ANN) model with appropriate parameters (refer to Table 1) was also implemented to predict PM_{10} levels to compare prediction performance between ANFIS and ANN. The ANN had three independent layers, that is, an input, hidden and output layer. The hidden layer was comprised of 10 neurons. To achieve comparable prediction performance, the number of trainings in ANN was twice what it was in ANFIS. Table 2 shows the results of MSE, RMSE, MAPE and R^2 calculated by ANN and ANFIS. These prediction performance criteria have different emphases on modeling results. RMSE gives the dispersion of measurement data, MAPE

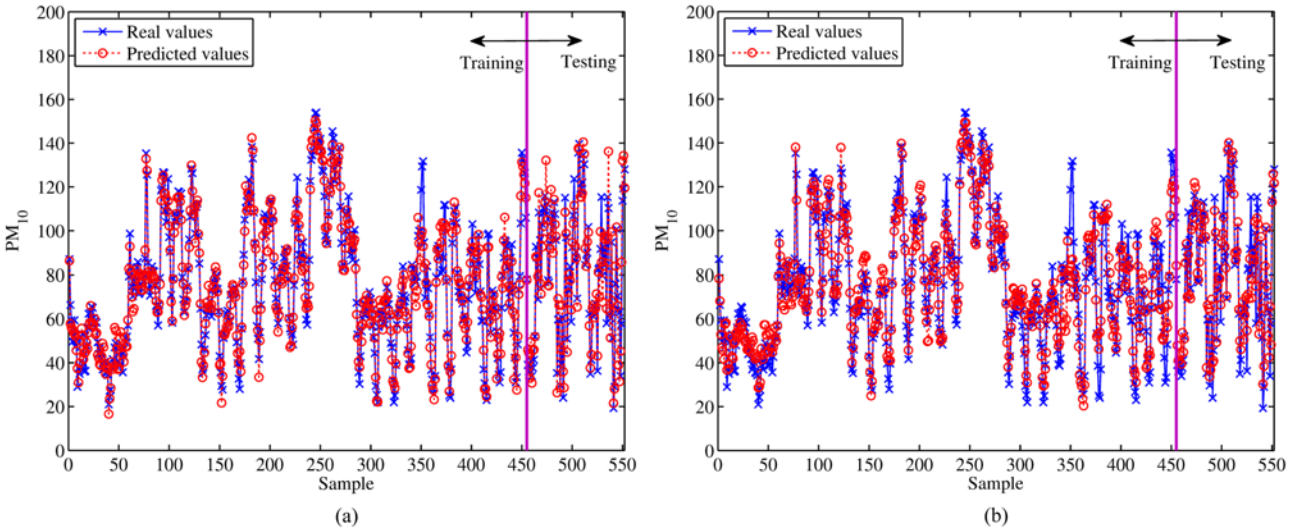

Fig. 9. Prediction results of PM_{10} using: (a) ANFIS with 15 clusters and (b) ANN.

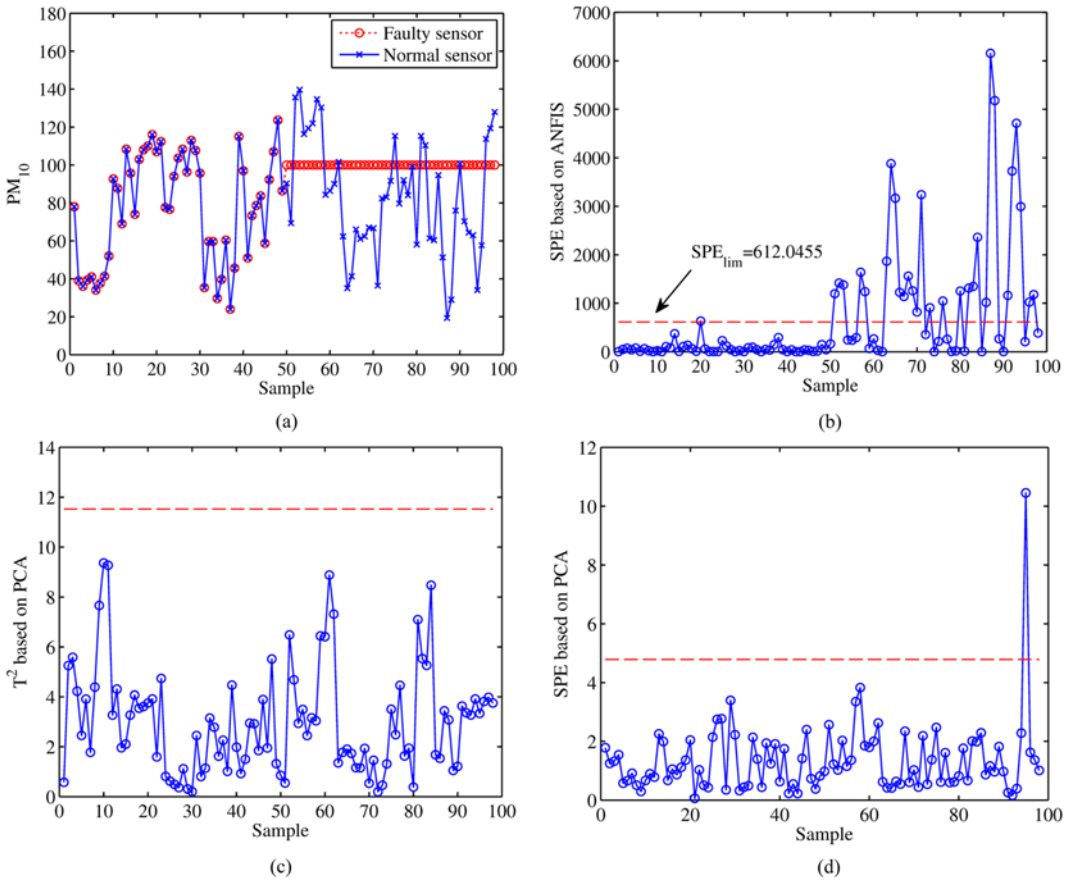

Fig. 10. Comparison of fault detection performances for (a) complete fault of the PM_{10} sensor appearing at sample 50, based on (b) ANFIS-based SPE, (c) PCA-based T^2 and (d) PCA-based SPE.

shows the mean ratio between the error and measurement values, and R^2 is an indicator of how well the variations performed in the predicted outputs. The R^2 values closer to the unit indicated a very good fitting performance. In Table 2, the terms "Train," "Validation," "Test" and "Overall" represent the use of the training, validation, testing and whole data sets to calculate the four performance criteria, respectively.

While training and validating, the MSE, RMSE and MAPE performance criteria values using ANFIS are lower than those using ANN. On the other hand, the R^2 values using ANFIS are higher than those using ANN. During testing, the modeling performance of ANN outperforms that of all three ANFIS structures. When considering the overall prediction performance, the MSE, RMSE and MAPE values using ANFIS are lower than those using ANN, while the R^2 values using ANFIS are higher than those using ANN. In view of the various models, the ANFIS with 15 clusters shows the best modeling performance among all three ANFIS models. Furthermore, the training and predicting results using ANFIS with 15 clusters and ANN are shown in Fig. 9(a) and Fig. 9(b), respectively. The performance of the ANFIS model with 15 clusters is better than that of ANN especially in the training part (Fig. 9). Therefore, the ANFIS model with 15 clusters was adopted to monitor four typical sensor faults in the next monitoring step.

3. Faulty Sensor Monitoring Using ANFIS

Sensor failures may decrease sensor reliability, which may result in an inaccurate or a dangerous control action on the performance of the monitored system. Therefore, a powerful sensor monitoring technique is an important issue.

Table 3. Summary of five fault scenarios and their fault detection rates of the PCA and ANFIS-based monitoring methods in a subway station

	Complete failure 1	Bias	Drifting	Complete failure 2	Precision degradation
Faulty sensor	PM_{10}	PM_{10}	Temperature	PM_{10}	NO_x
Fault size	100 μg/m^3	60 μg/m^3	0.5 °C/hour	180 μg/m^3	σ=0.2
Fault time	50	50	50	50	50
PCA (T^2)	0%	2%	31%	2%	0%
PCA (SPE)	2%	41%	69%	84%	27%
ANFIS (SPE)	**59%**	**100%**	**73%**	**100%**	**90%**

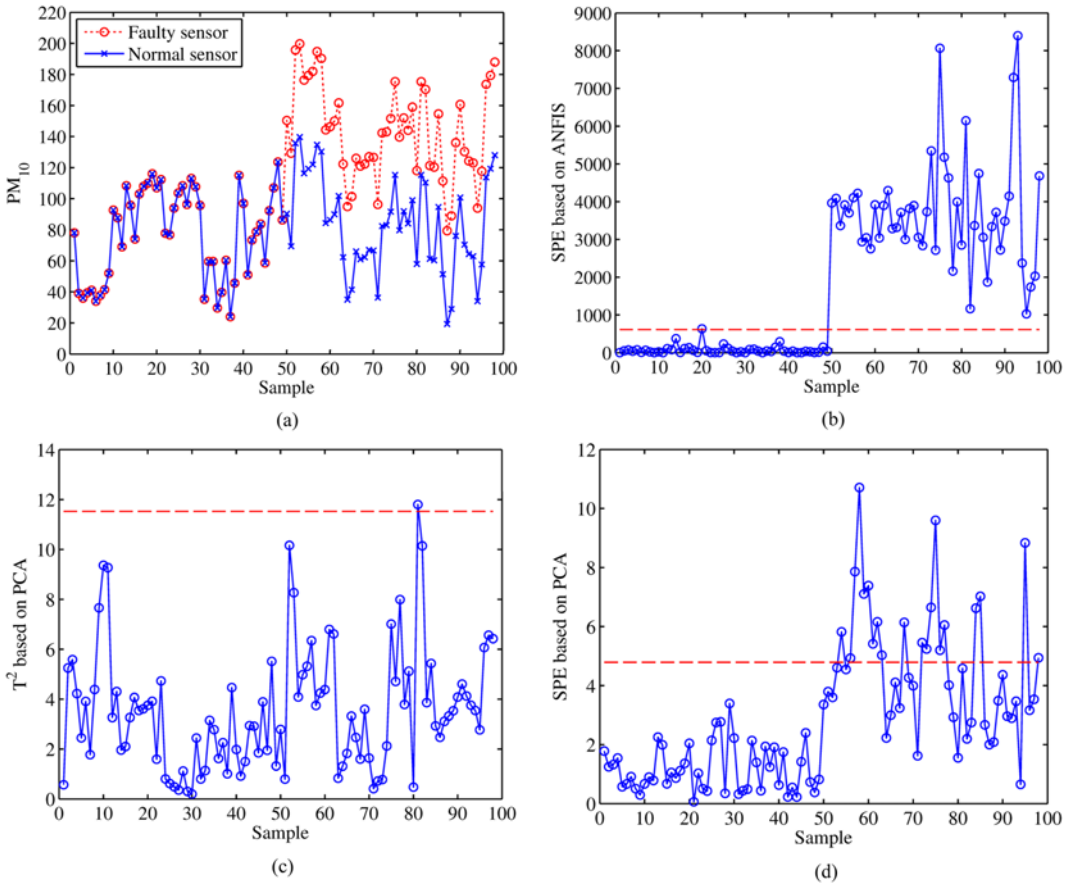

Fig. 11. Comparison of fault detection performances for (a) bias fault of the PM_{10} sensor appearing at sample 50, based on (b) ANFIS-based SPE, (c) PCA-based T^2 and (d) PCA-based SPE.

With respect to the ability of sensor fault detection, a complete failure fault of 100 μg/m^3 at sample 50 of the PM_{10} sensor was introduced to underline the superiority of ANFIS over PCA (Fig. 10(a)). The traditional PCA-based monitoring indices, such as T^2 (Fig. 10(c)) and SPE (Fig. 10(d)), fail to detect any such kind of sensor fault, whereas the ANFIS-based monitoring method could detect it. After the complete failure fault was introduced, the values of SPE in Fig. 10(b) begin to immediately exceed their CL calculated by using kernel density estimation and more than half of the points are successfully detected as sensor faults during the failure period. Because ANFIS can model the dynamics of IAQ data more precisely, it is more sensitive to this kind of sensor fault. Conversely, the PCA-based monitoring method cannot detect the abnormal measurements varying within the normal variation range. This kind of sensor fault has been considered undetectable by any statistical process monitoring technique (like PCA) because there are no changes in the mean, the variance or the higher moments [43].

To verify the superiority of an ANFIS-based sensor monitoring scheme over conventional PCA-based monitoring methods, four types of sensor faults consisting of bias, drifting, complete failure and precision degradation (Table 3) were tested using both ANFIS-based and PCA-based monitoring indices. A bias and complete failure were introduced to sample 50 of the PM_{10} sensor. The drifting and precision degradation IAQ sensor faults were introduced to sample 50 of the temperature sensor and sample 50 of the NO_x sensor, respectively. For the computation of fault detection rates, a 99% confidence limit was used in this study. The fault detection rates of the PCA and ANFIS-based monitoring methods are summarized in Table 3. As seen, the maximum fault detection rate value obtained for each of the IAQ sensor faults is highlighted in bold face. The ANFIS-based SPE monitoring index successfully recognizes all of the faults and outperforms the PCA-based monitoring indices (T^2 and SPE). In particular, for the faults of complete failure 1, bias and precision degradation, the detection rates of the ANFIS-based SPE monitoring method are significantly better than those of the PCA-based monitoring indices, which demonstrates that the ANFIS-based SPE chart is capable of detecting small events that are difficult to detect by linear statistical process monitoring methods like PCA.

To further illustrate the superiority of the ANFIS-based SPE monitoring method, the fault detection results of ANFIS-based SPE and PCA-based monitoring indices (T^2 and SPE) in bias, drifting, complete failure and precision degradation sensor faults are presented in Fig. 11, Fig. 12, Fig. 13 and Fig. 14, respectively. The 99% control limits are also shown in Fig. 11, Fig. 12, Fig. 13 and Fig. 14. For the bias fault, as shown in Fig. 11(a), a bias fault of 60 μg/m^3 was introduced to the PM_{10} sensor. The ANFIS-based SPE index can consistently detect all of the faults once they occurred after sample 50, as shown in Fig. 11(b), whereas the PCA-based T^2 index fails

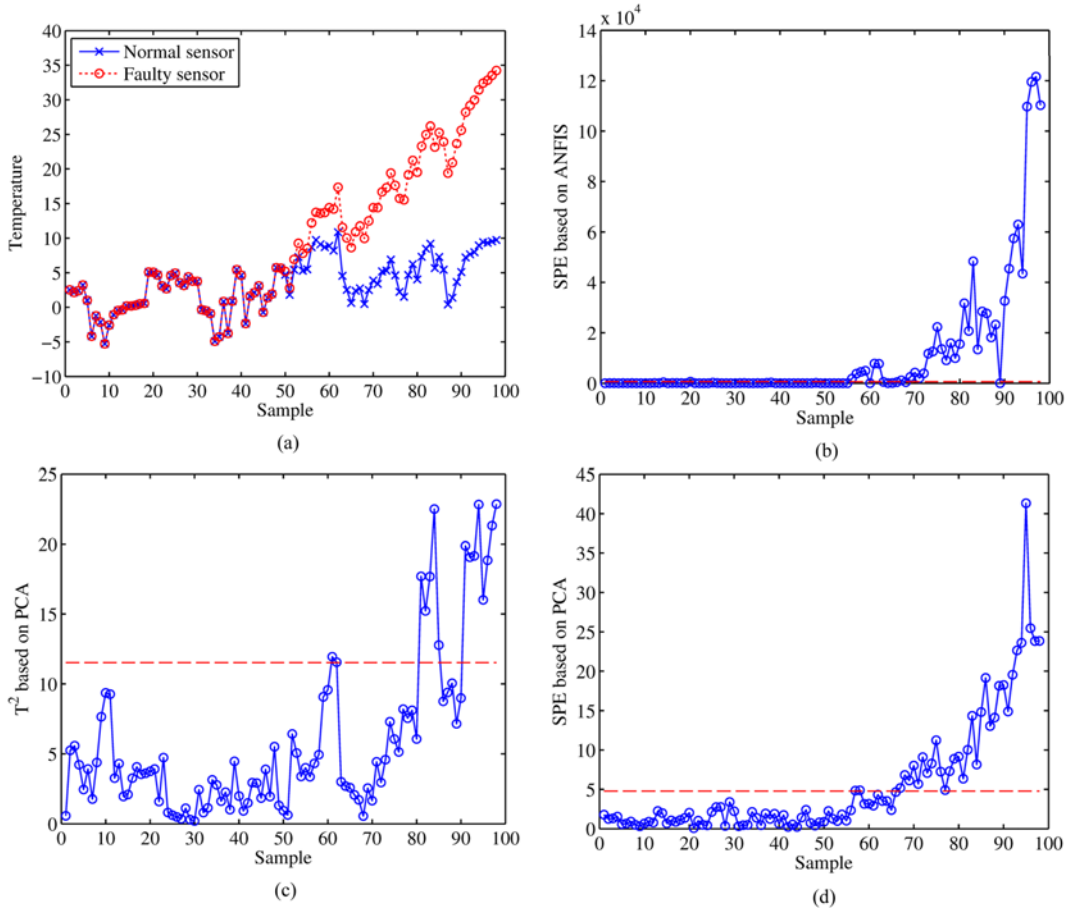

Fig. 12. Comparison of fault detection performances for (a) drifting fault of the temperature sensor appearing at sample 50, based on (b) ANFIS-based SPE, (c) PCA-based T^2 and (d) PCA-based SPE.

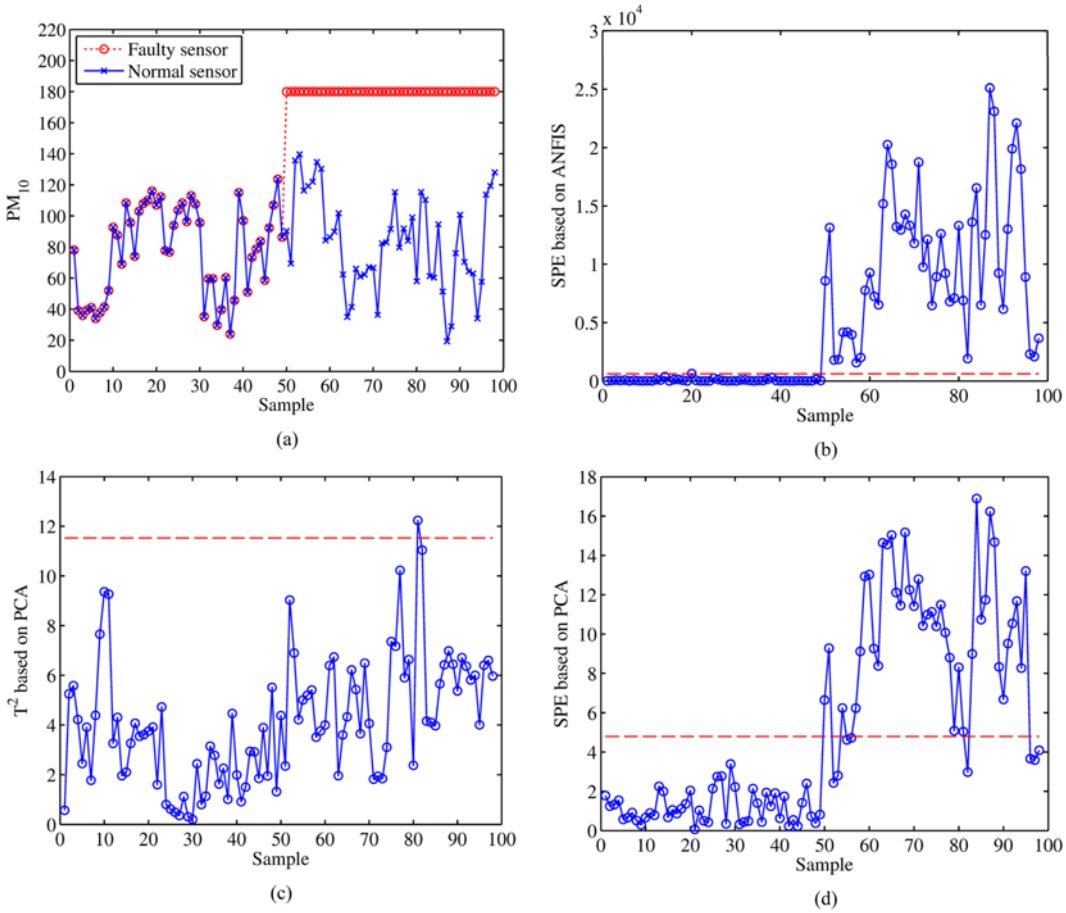

Fig. 13. Comparison of fault detection performances for (a) complete fault of the PM_{10} sensor appearing at sample 50, based on (b) ANFIS-based SPE, (c) PCA-based T^2 and (d) PCA-based SPE.

to detect these faults and the PCA-based SPE index only detects 41% of the faults (see Fig. 11(c), Fig. 11(d) and Table 3).

For the drifting fault, as shown in Fig. 12(a), a drifting fault with a magnitude of 0.5 °C/hour was introduced to the temperature sensor. The ANFIS-based SPE index effectively detected this kind of sensor fault (Fig. 12(b)). The changing tendency of SPE values is almost the same as that of drifting fault measurements. In contrast, the PCA-based T^2 index only detects the ending part of the fault (Fig. 12(c)). Although the PCA-based SPE index shows similar detection capabilities to the ANFIS-based SPE index, when comparing Fig. 12(b) and Fig. 12(d), it has a longer detection time for this sensor fault. For the complete failure fault, as shown in Fig. 13(a), an abrupt complete failure fault of 180 $\mu g/m^3$ was introduced to the PM_{10} sensor. This complete sensor failure fault is different from the previous complete failure fault in terms of different fault sizes. All the sensor faults are successfully detected during the failure period using the ANFIS-based SPE monitoring index as shown in Fig. 13(b). However, the PCA-based T^2 index again fails to detect these faults and the PCA-based SPE index detects 84% of the faults (see Fig. 13(c), Fig. 13(d) and Table 3). For the precision degradation fault as shown in Fig. 14(a), a precision degradation fault of σ=0.2 was added to the NO_x sensor. Being highly sensitive to the introduced external noise signals, the ANFIS-based SPE monitoring index (Fig. 14(b)) shows high fault detection capabilities compared to the PCA-based T^2 (Fig. 14(c)) and SPE indices (Fig. 14(d)). To sum up, the ANFIS-based SPE monitoring method could be successfully applied to detect all of the four kinds of sensor faults that commonly exist in IAQ monitoring systems.

Having high prediction accuracy is the main advantage of the ANFIS modeling method. However, when implementing this ANFIS model to monitor the variations of IAQ in a subway station, the problem of over-fitting, which is common in neural networks and ANFIS training, should be considered seriously during the course of the modeling phase. A possible solution to this problem is to periodically or adaptively maintain the ANFIS models.

CONCLUSIONS

Some sensor faults may have adverse effects on the maintenance, calibration and monitoring of IAQ measurements in subway systems. We have proposed an ANFIS-based faulty sensor monitoring method to enhance the monitoring capability of IAQ sensors. The prediction and monitoring performance of ANFIS was evaluated using the real-time IAQ data collected from a subway station. For the prediction accuracy performance, the ANFIS prediction model could predict PM_{10} variation better than the ANN model with similar structure, and the ANFIS with 15 clusters achieves the best modeling performance. The minimum performance criteria values for MSE,

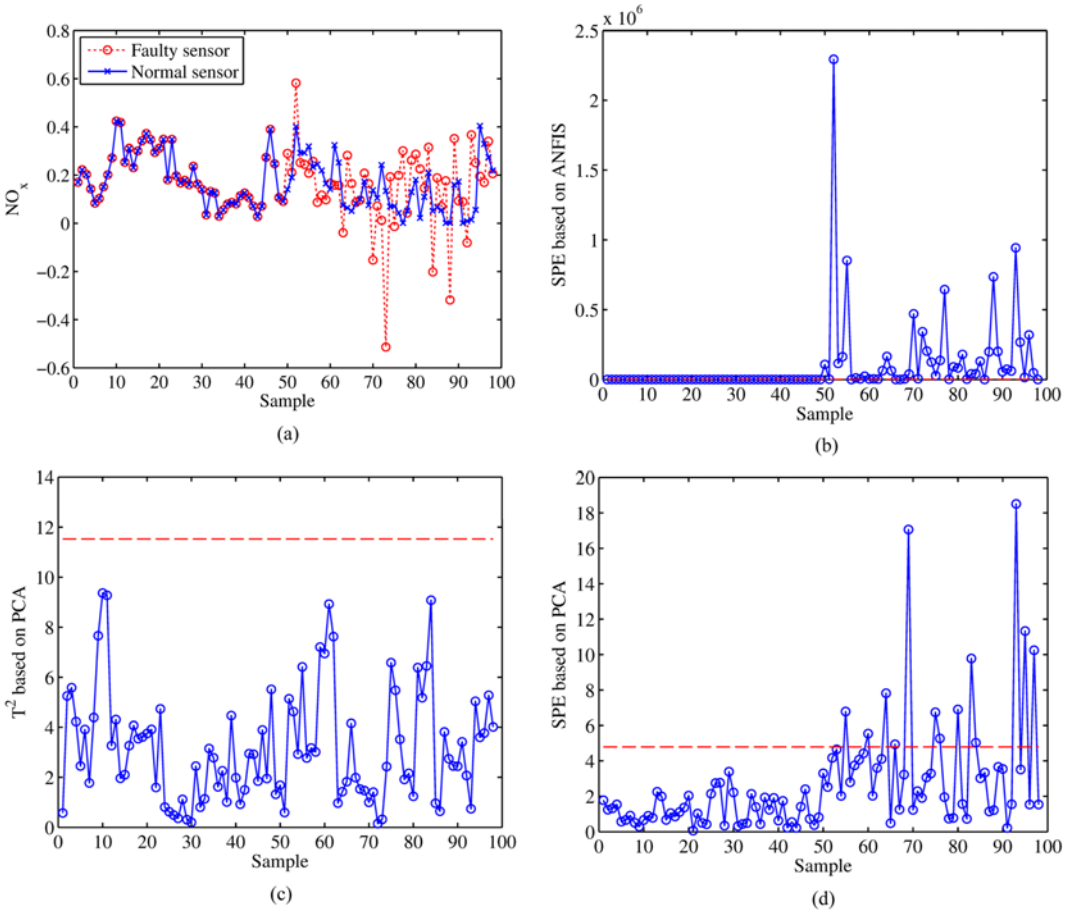

Fig. 14. Comparison of fault detection performances for (a) precision degradation fault of the NO_x sensor appearing at sample 50, based on (b) ANFIS-based SPE, (c) PCA-based T^2 and (d) PCA-based SPE.

RMSE and MAPE obtained are 45.20, 6.72 and 7.78, respectively. The maximum R^2 value is 0.95. For five IAQ sensor fault scenarios, the monitoring performance of the ANFIS with 15 clusters is better than that of the conventional PCA method because of the high precision of the ANFIS model.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2012-0000609).

REFERENCES

1. D.-U. Park and K.-C. Ha, *Environ. Int.*, **34**, 629 (2008).
2. X. Ye, Z. Lian, C. Jiang, Z. Zhou and H. Chen, *Environ. Monit. Assess*, **167**, 643 (2010).
3. L. Y. Chan, W. L. Lau, X. M. Wang and J. H. Tang, *Environ. Int.*, **29**, 429 (2003).
4. J. H. Cho, K. Hee Min and N. W. Paik, *Int. J. Hyg. Environ. Heal.*, **209**, 249 (2006).
5. Y. Feng, C. Mu, J. Zhai, J. Li and T. Zou, *J. Hazard. Mater.*, **183**, 574 (2010).
6. J. S. Pandey, R. Kumar and S. Devotta, *Atmos. Environ.*, **39**, 6868 (2005).
7. C. K. Yoo, K. Villez, S. W. Van Hulle and P. A. Vanrolleghem, *Environmetrics*, **19**, 602 (2008).
8. R. Dunia, S. J. Qin, T. F. Edgar and T. J. McAvoy, *AIChE J.*, **42**, 2797 (1996).
9. V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri and K. Yin, *Comput. Chem. Eng.*, **27**, 327 (2003).
10. J. C. M. Pires, S. I. V. Sousa, M. C. Pereira, M. C. M. Alvim-Ferraz and F. G. Martins, *Atmos. Environ.*, **42**, 1249 (2008).
11. J. Lau, W. T. Hung and C. S. Cheung, *Atmos. Environ.*, **43**, 769 (2009).
12. Y. Kim, J. T. Kim, I. Kim, J.-C. Kim and C. Yoo, *Environ. Eng. Sci.*, **27**, 721 (2010).
13. H. Liu, M. Kim, O. Kang, B. Sankararao, J. T. Kim, J.-C. Kim and C. K. Yoo, *Indoor Built Environ.*, **21**, 205 (2012).
14. L. H. Chiang, E. L. Russell and R. D. Braatz, *Fault detection and diagnosis in industrial systems*, Springer-Verlag (2001).
15. S. J. Qin, *J. Chemometrics*, **17**, 480 (2003).
16. W. Ku, R. H. Storer and C. Georgakis, *Chemometr. Intell. Lab*, **30**, 179 (1995).
17. J. Chen and K.-C. Liu, *Chem. Eng. Sci.*, **57**, 63 (2002).
18. S. Haykin, *Neural networks*, Prentice-Hall (1999).
19. J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem and I.-B. Lee, *Chem. Eng. Sci.*, **59**, 223 (2004).

20. Z. Ge, C. Yang and Z. Song, *Chem. Eng. Sci.*, **64**, 2245 (2009).
21. M. A. Kramer, *AIChE J.*, **37**, 233 (1991).
22. D. Dong and T. J. McAvoy, *Comput. Chem. Eng.*, **20**, 65 (1996).
23. J. Chen and C.-M. Liao, *J. Process. Contr.*, **12**, 277 (2002).
24. J. Zhang, *Comput. Chem. Eng.*, **30**, 558 (2006).
25. J. S. R. Jang, *IEEE Trans. Syst. Man Cybern.*, **23**, 665 (1993).
26. T. Y. Pai, T. J. Wan, S. T. Hsu, T. C. Chang, Y. P. Tsai, C. Y. Lin, H. C. Su and L. F. Yu, *Comput. Chem. Eng.*, **33**, 1272 (2009).
27. C. K. Lau, Y. S. Heng, M. A. Hussain and M. I. M. Nor, *ISA Trans.*, **49**, 559 (2010).
28. M. Huang, J. Wan, Y. Ma, H. Zhang, Y. Wang, C. Wei, H. Liu and C. Yoo, *Ind. Eng. Chem. Res.*, **50**, 13500 (2011).
29. M. Huang, J. Wan, Y. Wang, Y. Ma, H. Zhang, H. Liu, Z. Hu and C. Yoo, *Korean J. Chem. Eng.*, **29**(5), 636 (2012).
30. J. S. R. Jang, C. T. Sun and E. Mizutani, *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*, Prentice Hall (1997).
31. T. Takagi and M. Sugeno, *IEEE Trans. Syst. Man Cybern.*, **15**, 116 (1985).
32. M. Sugeno and G. T. Kang, *Fuzzy Set. Syst.*, **28**, 15 (1988).
33. J. Jackson, *A user's guide to principal components*, Wiley (1991).
34. B. M. Wise and N. B. Gallagher, *J. Process. Contr.*, **6**, 329 (1996).
35. S. J. Qin and R. Dunia, *J. Process. Contr.*, **10**, 245 (2000).
36. Q. Chen, R. J. Wynne, P. Goulding and D. Sandoz, *Control Eng. Pract.*, **8**, 531 (2000).
37. E. B. Martin and A. J. Morris, *J. Process. Contr.*, **6**, 349 (1996).
38. I. Jolliffe, *Principal component analysis*, Springer-Verlag (2002).
39. P. R. C. Nelson, P. A. Taylor and J. F. MacGregor, *Chemometr. Intell. Lab*, **35**, 45 (1996).
40. I. Stanimirova, M. Daszykowski and B. Walczak, *Talanta*, **72**, 172 (2007).
41. T. Chen, E. Martin and G. Montague, *Comput. Stat. Data. An.*, **53**, 3706 (2009).
42. M. J. Nieuwenhuijsen, J. E. Gómez-Perales and R. N. Colvile, *Atmos. Environ.*, **41**, 7995 (2007).
43. J. Yu, *J. Process. Contr.*, **22**, 1358 (2012).