

## Research Article

# Using Bayesian Inference Framework towards Identifying Gas Species and Concentration from High Temperature Resistive Sensor Array Data

Yixin Liu,<sup>1</sup> Kai Zhou,<sup>2</sup> and Yu Lei<sup>3</sup>

<sup>1</sup>ABB Corporate Research, Windsor, CT 06095, USA

<sup>2</sup>Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269-3139, USA

<sup>3</sup>Department of Chemical and Biomolecular Engineering, University of Connecticut, Storrs, CT 06269-3222, USA

Correspondence should be addressed to Yixin Liu; [yixin.liu@us.abb.com](mailto:yixin.liu@us.abb.com)

Received 31 March 2015; Accepted 4 May 2015

Academic Editor: Banshi D. Gupta

Copyright © 2015 Yixin Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High temperature gas sensors have been highly demanded for combustion process optimization and toxic emissions control, which usually suffer from poor selectivity. In order to solve this selectivity issue and identify unknown reducing gas species (CO, CH<sub>4</sub>, and CH<sub>8</sub>) and concentrations, a high temperature resistive sensor array data set was built in this study based on 5 reported sensors. As each sensor showed specific responses towards different types of reducing gas with certain concentrations, based on which calibration curves were fitted, providing benchmark sensor array response database, then Bayesian inference framework was utilized to process the sensor array data and build a sample selection program to simultaneously identify gas species and concentration, by formulating proper likelihood between input measured sensor array response pattern of an unknown gas and each sampled sensor array response pattern in benchmark database. This algorithm shows good robustness which can accurately identify gas species and predict gas concentration with a small error of less than 10% based on limited amount of experiment data. These features indicate that Bayesian probabilistic approach is a simple and efficient way to process sensor array data, which can significantly reduce the required computational overhead and training data.

## 1. Introduction

With increasing concerns over human health and environmental issues, gas sensors have received remarkable attentions for the past several decades and played important roles in many applications, including medical diagnostics [1, 2], food and drink analysis [3, 4], combustion process control [5], environmental monitoring [6], and industrial safety and public security [7]. High temperature gas sensors, which usually work in combustion related harsh environments, have been widely employed for monitoring the emission composition and optimizing the combustion process with improved energy efficiency, reduced toxic emission, and low operating cost [8]. Among well-developed gas analyzing instruments and state-of-the-art gas sensing technologies, solid-state gas sensors have seen growing usages in high temperature applications due to their high thermal stability

while subjecting them to aggressive high temperature environments [9]. However, a majority of solid-state gas sensors usually suffer from poor selectivity, indicating that they are simultaneously sensitive to a wide range of reducing and oxidizing gases. Selectivity has always been one of the most challenging issues for metal oxide-based gas sensors and there actually have been continuous and considerable efforts on improving it. Although novel sensing material and device design can enhance the sensor selectivity towards one specific gas to some extent, it is almost impossible for a single sensor to eliminate all interference or identify different gas species. Therefore, a sensor array, consisting of a group of sensors, was invented three decades ago and has been proven to be a more promising approach for gas identification.

A sensor array data set, which is collected from a cluster of nonselective sensors, could provide specific and unique response patterns (chemical fingerprints) for different

individual chemical species or mixtures of species. With subsequent data analysis, the gas sensor array could be used to qualitatively identify gas species using pattern recognition approaches and quantitatively determine gas composition based on regression methods [10]. It is worth noting that there are diverse statistical tools that can be utilized to facilitate the abovementioned data analysis. Commonly used pattern recognition schemes are principal component analysis (PCA) and cluster analysis [11], artificial neural networks (ANN) [12], and specific algorithms based on fuzzy logic [13]. Among them, PCA is the most prevalent technique to analyze sensor array data. A large number of literatures on sensor arrays have reported their results based on PCA [14–17], which shows its potential in clustering applications, especially in cases where data can be represented as clusters. However, PCA is merely a clustering technique and it requires additional classification methods, such as distance measures, Bayesian methods, support vector classifier, and neural networks, to identify specific species [18]. While PCA can qualitatively identify different gas species with the aid of the above classifiers, PCA is not able to quantify the gas composition. As mentioned above, regression methods are commonly used as quantitative methods to determine the gas concentration. Therefore, to further achieve the goal of gas composition identification, regression methods oftentimes are incorporated into general PCA framework, leading to a series of closely related techniques, for example, principal component regression (PCR) and partial least squares (PLS) [19]. For many of the applications presented, the separation provided by these methods was sufficient. It is also worth mentioning that PCA, PCR, and PLS are all linear methods, which require extra care in interpreting results when applying to nonlinear sensor array response data [10]. In contrast, cluster analysis and artificial neural networks offer the potential to accurately model nonlinearity presented in sensor array data. However, such intelligent algorithms yield prohibitive computational cost since usually a large number of training samples and considerable amount of training time are required [10].

In many cases where experimental efforts and time are largely required, it is almost impossible to obtain a comprehensive sensor response profile in a wide detection range. In such scenario, there is only a limited amount of sensor data for limited analyte concentrations, and curve-fitting techniques are generally applied to build calibration curves for those sensors in order to predict an unknown concentration based on a sensor response. In this work, we utilize Bayesian inference framework combined with curve-fitting techniques to simultaneously identify gas species ( $\text{CO}$ ,  $\text{CH}_4$ , and  $\text{C}_3\text{H}_8$ ) and concentrations (30–150 ppm) with limited sensor sample data, which greatly reduced the cumbersome efforts in classification and data training. Bayesian inference essentially is a statistical method of inference in which Bayes' rule is adopted to update the probability estimate for a hypothesis as additional evidence is acquired. Because of its formulation robustness and implementation convenience, Bayesian inference has been applied into a variety of engineering practices, and the results clearly validated its feasibility and effectiveness [20–22]. Another pronounced advantage of Bayesian

inference lies in its tolerance of uncertainty, which is inevitable in real sensor measurement. In this study, the sensor array data set was extracted from our previous studies using five metal oxides-based resistive sensors. Detailed methodology was presented and case study was then demonstrated. It is clearly found that Bayesian probabilistic approach can be tailored for simultaneous identification of gas species and composition, which can significantly reduce the required computational overhead and the amount of training data.

## 2. Experimental

**2.1. Fabrication of Metal Oxides Nanofibers and Corresponding Sensors.** The sensor array consisted of 5 metal oxides-based resistive sensors, including  $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$  (n-type) NFs (Sensor I),  $\text{CeO}_2$  (p-type) NFs (Sensor II), p-LSMO/n- $\text{CeO}_2$  NFs composite (LSMO wt% = 20%) ( $\text{L}_{20}\text{C}_{80}$ ) (Sensor III), NiO (n-type) NFs (Sensor IV), and Ce-Ni-O NFs (Ce:Ni atom ratio = 1:1) (Sensor V). LSMO NFs,  $\text{CeO}_2$  NFs, NiO NFs, and Ce-Ni-O NFs were fabricated by electrospinning followed by a subsequent calcination process. The detailed preparation procedure and characterization were reported in previous studies [23–26]. LSMO/ $\text{CeO}_2$  nanofibers composite was prepared by physical mixing and sonication [25]. An individual metal oxide-based sensor was fabricated on  $\text{Al}_2\text{O}_3$  ceramic screws ( $4\text{--}40 \times 1/2''$ ), as reported elsewhere [27]. Briefly, the ceramic screw substrate was tightly tied by two Pt wires on two close threads, serving as two electrodes. Certain amount ( $\sim 2 \mu\text{g}$ ) of sensing metal oxide nanofibers after sonication in ethanol was casted onto the substrate and bridged the two Pt electrodes to complete the sensor fabrication.

**2.2. Sensor Measurements.** As reported previously, the sensing performance of each as-fabricated sensor at high temperature of  $800^\circ\text{C}$  was evaluated by measuring the resistance/conductance change of the sensor upon exposure to different concentrations of reducing gas ( $\text{CO}$ ,  $\text{CH}_4$ , and  $\text{C}_3\text{H}_8$ ) in a dynamic gas flow system. For illustration purpose, Figure 1(a) shows the proposed sensor array configuration based on five individual sensors and the test system using a multiple channel electrochemical workstation. At  $800^\circ\text{C}$ , reducing gas could react with  $\text{O}_2$ ; therefore, high purity nitrogen was used as the carrying gas of various reducing gases instead of air, and 1%  $\text{O}_2$  (in  $\text{N}_2$ ) was used as the sensor-recovering gas. All sensors were subject to a gas flow with a constant flow rate of 1.5 L/min, which were regulated by a computer-controlled gas mixing system (S-4000, Environics Inc., USA). The current output of each individual sensor at a fixed 1 V DC bias was continuously measured and the electric resistance of the sensor was calculated by applying Ohm's Law ( $R = V/I$ ). In a typical reducing gas sensing experiment, 100 ppm  $\text{C}_3\text{H}_8$ , for example, the sensor placed in furnace at  $800^\circ\text{C}$  was first exposed to  $\text{C}_3\text{H}_8/\text{N}_2$  mixture for 5 min (after baseline was stabilized), followed by 1%  $\text{O}_2$  for 10 min to recover the sensor, and then the "exposure/recovery" cycle was repeated.

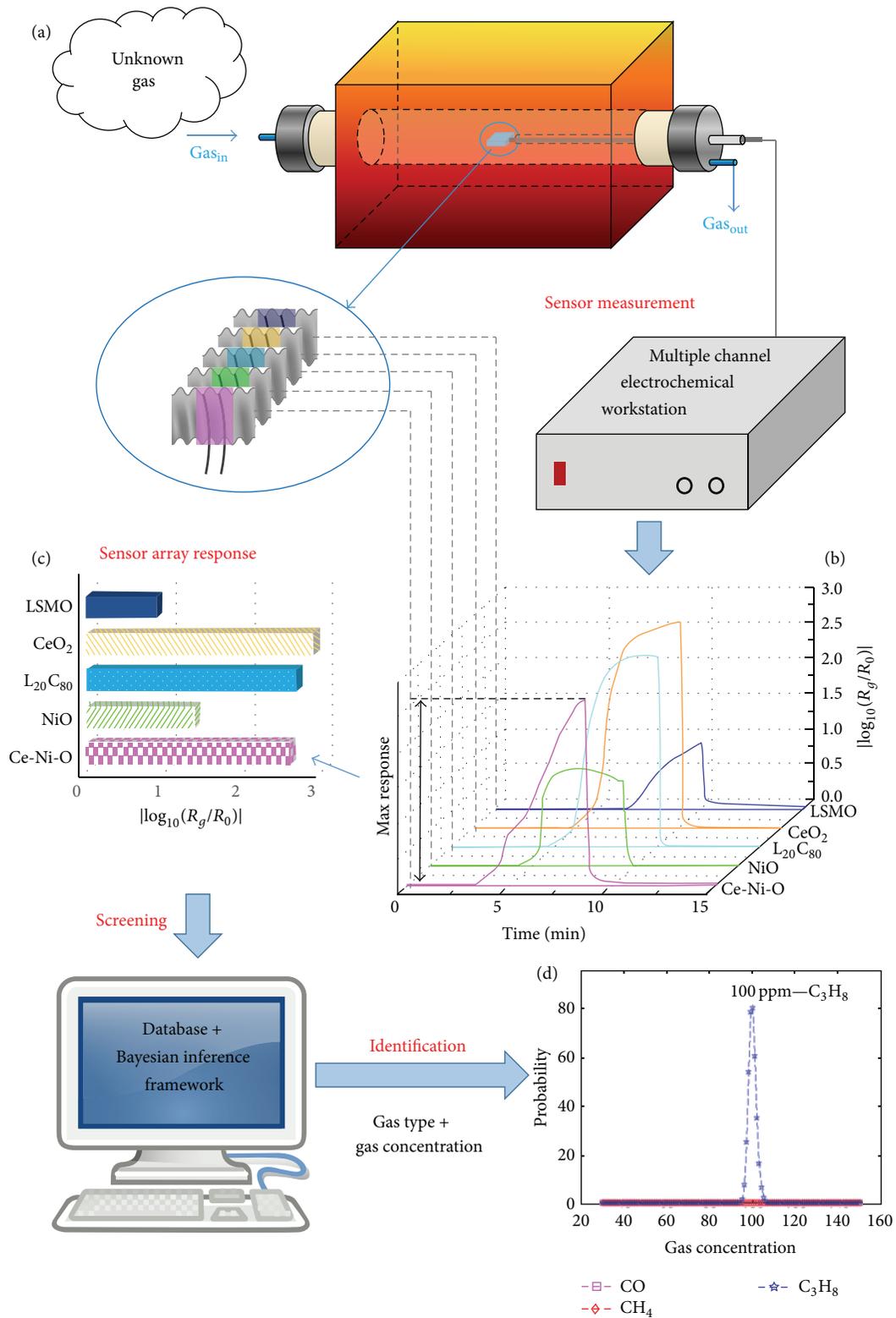


FIGURE 1: Illustration of gas identification process based on sensor array response data analysis: (a) the proposed configuration of sensor array and test system, (b) the combined real-time gas sensing profile of a sensor array based on five sensors' response measurement, (c) extracted sensor array response pattern (based on (b)), and (d) gas identification result using Bayesian inference framework.

### 3. Sensor Array Data Set

The sensor array data set used in this study was extracted and generated from the sensing performances of as-mentioned five metal oxides-based sensors in previous reports. As a demonstration, Figure 1(b) presents a typical combined real-time gas sensing profile of a sensor array based on five sensors' response measurements towards 100 ppm  $C_3H_8$  as an example. Upon the exposure to reducing gas, the resistance of the n-type sensors ( $CeO_2$  NFs,  $I_{20}C_{80}$  NFs composite, and Ce-Ni-O NFs based sensors) decreases, so the sensor response was defined as  $R_0/R_g$ , where  $R_0$  is the initial electrical resistance of the sensor in 1%  $O_2$  and  $R_g$  is the measured real-time resistance upon exposure to reducing gas/nitrogen mixture or 1%  $O_2$  recovering gas. In contrast, the resistance of the p-type sensors (LSMO NFs and NiO NFs based sensors) increases in the presence of reducing gas;  $R_g/R_0$  thus was used as sensor response to keep the number larger than 1. Due to the large variation of each sensor response towards different reducing gases, that is, ranging from less than 10 to several hundreds, the absolute value of  $\log_{10}(R_g/R_0)$  was used as an indicator when comparing the sensor array response patterns in different gas atmosphere. The maximum value of  $|\log_{10}(R_g/R_0)|$  of each individual sensor towards a certain concentration of gas was collected to build a sensor array response pattern at that specific gas concentration, for example, 100 ppm  $C_3H_8$ , as shown in Figure 1(c). The figures of real-time sensing response of each individual sensor can be found in previous reports.

Figure 2 summarizes the sensor array response patterns towards different reducing gases ( $CO$ ,  $CH_4$ , and  $C_3H_8$ ) with 5 selected concentrations. For each concentration of each reducing gas, the response of each sensor in a pattern was the calculated average of three parallel measurements. As one can notice in Figure 2, the sensor array exhibited good concentration dependent behavior towards each type of gas and showed overall highest response towards propane, followed by  $CO$  and  $CH_4$  successively. Meanwhile, for each concentration of gas, the sensor array showed a unique response pattern, which can be used as a specific ID in gas identification process.

### 4. Methodology

*4.1. Prestudy of Machine Learning Technique (Artificial Neural Network).* There are a number of well-developed machine learning techniques that have been employed to mimic the underlying relationship between the observed/training data input and output, that is, the experimental measurement. Among those techniques, nonparametric modeling methods, such as artificial neural network, Gaussian process, and Genetic algorithm, are widely applied into industries due to their inherent inference merits. Both neural network [28] and Gaussian processes [29] have been reported to establish gas recognition systems based on gas sensor arrays. While such methods show a good potential in parameter identification/prediction, their accuracy and efficiency highly depend on the amount of training data involved.

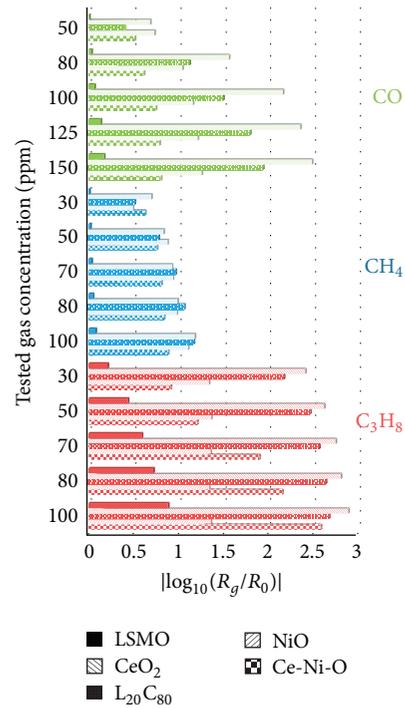


FIGURE 2: Summarized sensor array response patterns of reducing gas ( $CO$ ,  $CH_4$ , and  $C_3H_8$ ) with tested concentrations.

In the case that the sufficient amount of training data is used, high prediction fidelity can be ensured.

In this study, due to practical limitation, we only have 15 sets of data available (3 types of reducing gases, 5 concentrations for each gas), which can be divided into two groups, that is, training and test data. The training data thus is in a very small amount and the prediction fidelity naturally will become a major concern. To verify this, we utilized a 3-layer BP neural network [30] for gas identification, in which 12 sets of data (3 types of reducing gases, 4 concentrations for each type of gas) were considered as training data and 1 set of data as test data for simple validation. It is worth nothing that the initial weights of different layers were randomly generated, which may affect the prediction result significantly. This issue will be compounded especially when only a small amount of training data is involved. In this context, we aim to evaluate its performance in terms of the prediction robustness by repeatedly running 1,000 times and acquiring the statistical distribution of prediction results. In this numerical analysis, the number of nodes in input layer was 5, equivalent to the number of sensor types. The number of nodes in hidden layer was 8, and the number of nodes in output layer was 2, respectively, indicating the gas type and concentration. The learning rate was selected as 0.1, and the iterations number of each run was set to 10,000 for convergence concern. As in this network model the output was set to being continuously varied values, the gas type identification result thus needed to be rounded (this network model can be modified under the same underlying framework). The nominal tested gas was  $CO$  (type 1 defined in this paper) with 125 ppm, which can be compared with result in supplementary. The mean of gas

type shown in Figure S.1 (in Supplementary Material available online at <http://dx.doi.org/10.1155/2015/351940>) is 1.21 that can be rounded to 1, indicating highest possibility of gas type 1, that is, CO. Similarly, the mean of concentration in Figure S.2 is 1.045 (equivalent 104.5 ppm according to the network model definition), yielding around 16.4% error. Although the result to some extent reveals the underlying properties of tested gas, the prediction error is still obvious.

On the other hand, the authors' previous work [23–26] has preliminarily investigated the gas sensing mechanism of each sensor individually. Except NiO based sensor towards  $C_3H_8$ , which showed no concentration dependence, all sensors exhibited increasing response upon exposure to increasing concentration of each gas, which will be discussed in Section 4.3. Therefore, the aforementioned machining learning techniques that are able to model the complicated relation between input and output may not be quite demanding. Alternatively, we can resort to general curve-fitting techniques to approximate such trend, from which a rich number of sensor response vectors can be parameterized. The sensor array database comprising those response vectors then will be established for Bayesian inference implementation.

**4.2. Bayesian Inference Framework.** Generally, Bayesian inference is one popular type of statistical inference method in which Bayes' rule is adopted to update the probability estimate for a hypothesis as additional evidence is acquired. Here a hypothesis is a prior distribution of concerned parameters depending on the prior knowledge and engineering judgment. Additional evidence indicates the actual output of the system that is characterized by concerned parameters to be identified. With additional evidence introduced, the prior distribution of concerned parameters can be updated, resulting in a so-called posterior distribution. According to the underlying features of posterior distribution, the candidate of parameters corresponding to the highest probability value can be considered as the best solution. Such framework allows one to statistically screen the candidates sampled from prior distribution. Specifically in this work, the gas species type and gas composition can be regarded as two concerned parameters. Meanwhile, the sensor array responses of objective gases are treated as the additional evidence. In conjunction with the sensor array database generated by surrogate sensor response model, the parameter updating process based on Bayesian inference then can be performed. In other words, Bayesian inference framework was employed to build a parameter (gas species and concentration) selection program, in which a measured sensor array response pattern towards an unknown concentration of an unknown gas can be considered as the input of this program. In combination with the benchmark sensor array response database, Bayesian inference-based program eventually yields the best solution of gas species and concentration, whose sensor array response pattern is in agreement with the measured one, thereby identifying the unknown gas, as illustrated in Figure 1(d). In this section, we will briefly outline the mathematical formulation of Bayesian inference framework as follows.

When applying Bayesian theorem for gas identification, the hypothesis  $\theta$  is interpreted as the vector of gas-related

parameters, that is, gas species and gas concentration in this case that need to be updated.  $E$  represents the measured sensor array response pattern/vector of an objective gas. All the notations are included in the following equation, shown as

$$p(\theta | E) = \frac{p(E | \theta) p(\theta)}{\int p(E | \theta) p(\theta) d\theta}. \quad (1)$$

The goal of Bayesian inference framework in this study is to determine the maximum probability of  $\theta$  (gas species and gas concentration) with input  $E$  (experimental measured sensor array response) based on the posterior distribution  $p(\theta | E)$ , which depends on the prior and likelihood distributions. The prior distribution  $p(\theta)$  describes the initial knowledge of concerned gas-related parameters. In general, this distribution is defined empirically as in many cases the knowledge is not sufficiently given. In this paper, we can simply specify this prior distribution as uniform distribution, in which each candidate  $\theta$  of sampled parameters from such distribution has equal probability to be the actual gas-related parameters. In other words, before introducing measured sensor array response  $E$ , the probabilities of sampled parameters are the same. This assumption agrees with the general understanding of an unknown system.  $p(E | \theta)$  is so called likelihood distribution that is used to evaluate the deviation between measured sensor array response pattern/vector  $E$  and a sampled sensor array response pattern/vector  $\tilde{E}(\theta)$  in benchmark sensor array database, which will be explained later. As a result, conditioned on the prior and likelihood distribution, the posterior distribution  $p(\theta | E)$  indicates the updated knowledge of  $\theta$ , which offers the profile to identify gas species and concentration.

In (1), the likelihood distribution remains unknown. In a broad sense, the likelihood distribution is a specified form to quantify the discrepancy between two different sets of data in a probabilistic manner. Here, we utilized the normal distribution for likelihood distribution that can be explained below. As illustrated by the profile of normal distribution, if arbitrary set of data  $\tilde{E}(\theta)$  in sensor array database approaches the measured sensor array response pattern/vector  $E$  of the objective gas, the gas parameters  $\theta$  corresponding to such sensor array data  $\tilde{E}(\theta)$  are much likely to be the actual parameters of the objective gas. In an extreme case, identical two data sets will yield identified gas parameters with absolutely maximum probability value. For certain vector of measured sensor response, we can have

$$p(E | \theta) = D e^{-(E - \tilde{E}(\theta))^T C^{-1} (E - \tilde{E}(\theta)) / 2}, \quad (2)$$

where  $\tilde{E}(\theta)$  is one sensor array response vector with parameters  $\theta$ , which is extracted from established benchmark sensor array response database.  $\tilde{E}(\cdot)$  is an operator to generate the benchmark sensor array response database that will be clarified in the next section.  $C$  is a covariance matrix of  $\theta$ , implying the visualization degree of sensor response deviation. For example, the smaller the values of  $C$  matrix are, the more pronounced the deviation will be presented.  $D$  is a constant related to determinant of the covariance matrix  $C$ .

The denominator of (1) is an integral that can be accurately approximated by using the large number law. However, it is worth mentioning that this integral is a normalization constant that merely depends on the measurement. Therefore, (1) can be further simplified, in which the posterior distribution is expressed to be proportional to the numerator

$$p(\boldsymbol{\theta} | \mathbf{E}) \propto p(\mathbf{E} | \boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (3)$$

At this point, we can simply use the numerator of Bayes rule to acquire the posterior probability of all potential gas-related parameters  $\boldsymbol{\theta}$ . The distribution of numerator indeed reflects the identification result and its confidence level.

**4.3. Benchmark Sensor Array Response Database.** As mentioned in prestudy, 15 sets of sensor response data were extracted from experiment, as shown in Figure 2. Among those data sets, 12 sets of data (3 types of gases, 4 concentrations for each gas) are used as training data to build the benchmark sensor array database, based on which the Bayesian inference framework was implemented, and the remaining 3 sets of data (one concentration for each gas: 125 ppm CO, 70 ppm CH<sub>4</sub>, and 70 ppm C<sub>3</sub>H<sub>8</sub>) are treated as test data to validate the algorithm. This data classification strictly satisfies the data independence discipline.

As mentioned previously, we determine to employ general curve-fitting techniques for constructing sensor array database. Specifically, here we adopt least square fitting method [31]. The objective is to minimize the sum of the squares of errors between observed data and correspondence at specified curve function, which is expressed as

$$\varepsilon = \sum_{r=1}^n (f(x_r) - y_r)^2, \quad (4)$$

where  $y_r$  is the  $r$ th measured sensor response towards one gas type with certain concentration.  $f(\cdot)$  is a specified curve function, either polynomial or exponential function in this study.  $n$  is the number of measured sensor responses. The objective gives a measure of how well  $f(x_r)$  fits the data  $y_r$ .

The underlying idea of least square fitting method necessitates optimizing the key parameters of curve function by minimizing the above objective value. In general, this optimization problem can be easily achieved by using standard optimization toolbox in MATLAB, for example, “fminsearch” and “fmincon.” The well-calibration curves are given in Figure 3, which will be discussed in Section 5. As mentioned earlier,  $\tilde{\mathbf{E}}(\cdot)$ , which is the operator to generate the sensor array response database, consists of 15 calibration equations which can produce the sensor array response pattern/vector for each gas-related parameter (gas species and concentration). In this study, for each type of gas, the sensor array response in database was discretized from the calibration curves with every 1 ppm gas concentration change over entire detection range, providing sufficient resolution for subsequent gas identification.

## 5. Results and Discussion

**5.1. Calibration Curves Interpretation.** As mentioned previously, due to the simple tendency of the sensor response data, general curve fitting was used to build the comprehensive profile between sensor response and gas concentration. As shown in Figure 3, for 5 sensors and 3 types of gases, 15 calibration curves were generated based on experimental sensing response towards 4 concentrations of each gas, using curve-fitting technique discussed in Section 4.3. The concentration range of each gas was described by minimum and maximum concentration tested, that is, 50–150 ppm for CO and 30–100 ppm for CH<sub>4</sub> and C<sub>3</sub>H<sub>8</sub>.

Generally, p-LSMO NFs based sensor had the smallest sensitivity with a slightly exponential response increase with increasing gas concentration. p-NiO NFs based sensor showed a nearly liner relation between the sensor responses and the concentrations of CO/CH<sub>4</sub>, while it exhibited a concentration independent behavior with increasing concentration of C<sub>3</sub>H<sub>8</sub> with an already saturated response at low concentration (30 ppm). n-CeO<sub>2</sub> NFs based sensor showed a small response to CH<sub>4</sub>, a relatively large response towards CO with an exponential increase, and the highest sensitivity towards C<sub>3</sub>H<sub>8</sub> with a gradual saturation pattern. Therefore, it is difficult to differentiate the response from high concentration of CO and that from low concentration of C<sub>3</sub>H<sub>8</sub> based on single CeO<sub>2</sub> NFs based sensor. By adding p-LSMO NFs into n-CeO<sub>2</sub> NFs to prepare a L<sub>20</sub>C<sub>80</sub> NFs composite, the sensor showed a very similar response pattern with n-CeO<sub>2</sub> NFs based sensor with a reduced sensitivity and a slightly enhanced selectivity towards C<sub>3</sub>H<sub>8</sub> over CO. In contrast, by incorporating p-NiO and n-CeO<sub>2</sub> into coelectrospun Ce-Ni-O NFs, the sensor exhibited a totally different response pattern compared to n-CeO<sub>2</sub> NFs based sensor. Ce-Ni-O NFs greatly reduced the sensitivity towards CO and showed a response saturation pattern towards both CO and CH<sub>4</sub>, while it showed a large exponential increase with increasing concentration of C<sub>3</sub>H<sub>8</sub>. With such unique combination of 15 calibration curves based on 5 different sensors, the sensor array could provide a more comprehensive profile of the sensor array response towards each type of gas within the corresponding detection range. These 15 calibration curves were discretized to generate the benchmark sensor array response database for the subsequent Bayesian inference-based gas identification program.

### 5.2. Prediction of Unknown Gas Species and Concentration.

For an unknown gas species with certain concentration, we can measure the response pattern/vector by the sensor array and then substitute it into Bayesian inference-based program for gas identification. As we discussed in the previous section, the benchmark database is fully represented by 15 calibration curves which is fitted based on the aforementioned 12 sets of training data. Three sets of test data for 125 ppm CO, 70 ppm CH<sub>4</sub>, and 70 ppm C<sub>3</sub>H<sub>8</sub> were introduced into the identification program as test data. The sensor array response patterns of those three sets of test data are compared in Figure 4(a). As one can see in Figures 4(b), 4(c), and 4(d), the output of the program essentially is the updated probability distribution

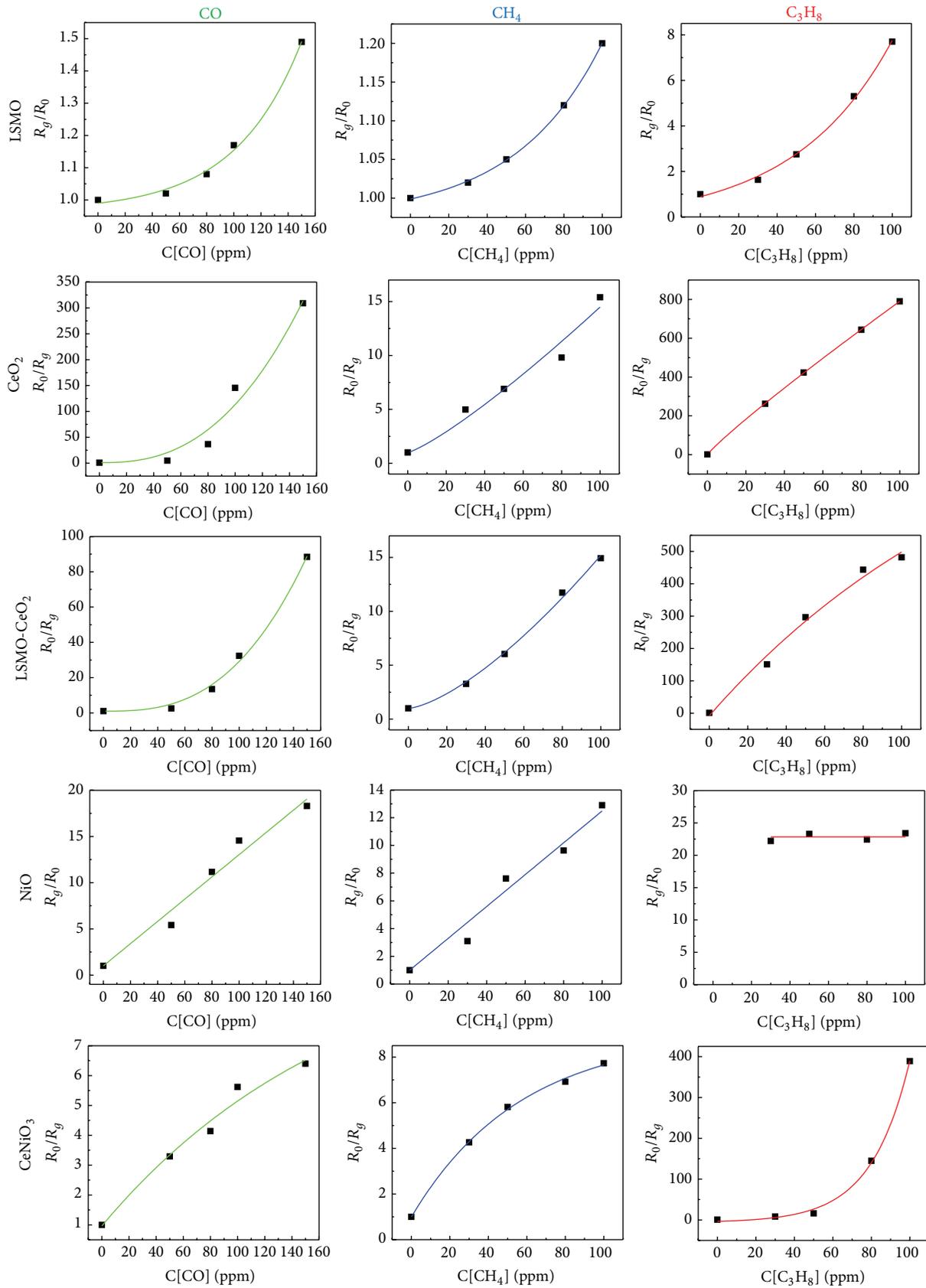


FIGURE 3: Calibration curves of sensor response with respect to the gas species and concentration.

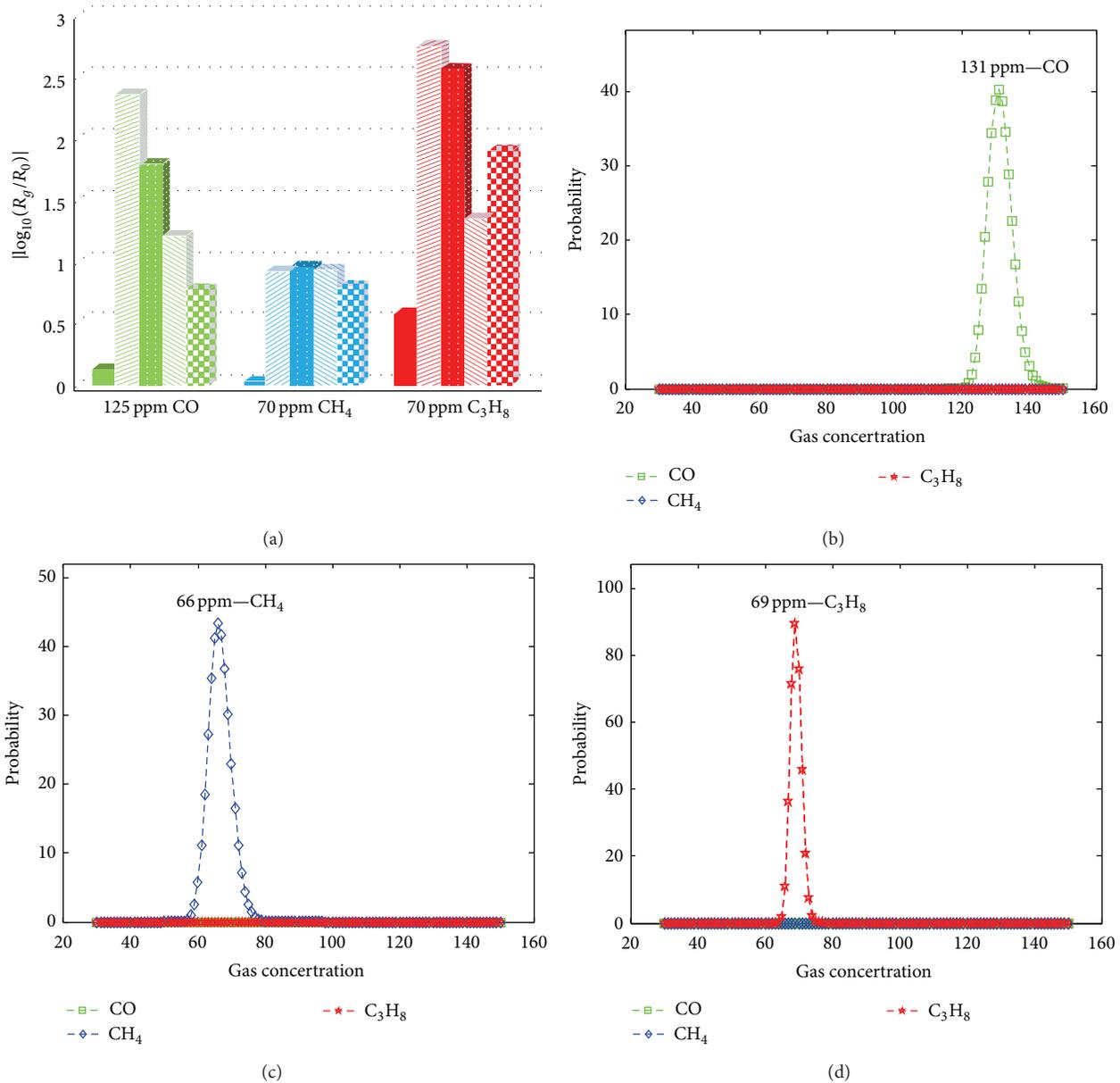


FIGURE 4: (a) The sensor array response input of 125 ppm CO, 70 ppm CH<sub>4</sub>, and 70 ppm C<sub>3</sub>H<sub>8</sub> and the corresponding resulting output of (b) 125 ppm CO, (c) 70 ppm CH<sub>4</sub>, and (d) 70 ppm C<sub>3</sub>H<sub>8</sub>.

of gas species and concentration, that is,  $\theta$ . Here, the gas concentration is uniformly parameterized with a resolution of 1 ppm in detection range. The result indicates that the combination of gas species and concentration resulting in the maximum probability is considered as the best parameters (gas species and concentration) of unknown tested gas. As presented in Figure 4, this gas identification program successfully identifies those three gas species and predicts the gas concentrations with a relatively high accuracy (error within  $\pm 10\%$ ). The program predicted 131 ppm CO, 66 ppm CH<sub>4</sub>, and 69 ppm C<sub>3</sub>H<sub>8</sub>, respectively, under experimental sensor array response input of 125 ppm CO, 70 ppm CH<sub>4</sub>, and 70 ppm C<sub>3</sub>H<sub>8</sub>. The authors feel quite confident that the proposed algorithm will show good feasibility and robustness if more other test data sets are involved. Even in the case that the test sensor responses are similar to a number of responses

(more than 1) in sensor array database, the identification with high confidence level will be achieved through adjusting certain implementation parameters of Bayesian inference algorithm, such as the covariance of likelihood function.

As verified by intensive results, the proposed Bayesian inference framework enables simultaneous identification of gas species and concentration with only limited sample data. With sufficient experimental sensor array response patterns of binary or ternary gas mixture, this framework also can be extended to analyze multiple gas compositions, which will be a future research topic.

## 6. Conclusions

A high temperature resistive sensor array data set was extracted from five reported metal oxides-based sensors,

which was employed to build the database of response patterns towards reducing gas ( $\text{CO}$ ,  $\text{CH}_4$ , and  $\text{C}_3\text{H}_8$ ) at  $800^\circ\text{C}$  by curve-fitting techniques. Bayesian inference framework was utilized to process the sensor array data and identify the gas species and concentration at the same time with limited sensor array data, which greatly reduced the cumbersome efforts in classification and data training. This algorithm accurately identified gas species and predicted gas concentration with good robustness and high accuracy, that is, small error less than 10%. Under this framework, the identification of gas mixture composition can be potentially accomplished as long as sufficient experimental sensor array response patterns of binary or ternary gas mixture are provided.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

The authors greatly appreciate the partial funding support from the Department of Energy (DOE).

## References

- [1] A. P. F. Turner and N. Magan, "Electronic noses and disease diagnostics," *Nature Reviews Microbiology*, vol. 2, no. 2, pp. 160–166, 2004.
- [2] E. R. Thaler and C. W. Hanson, "Medical applications of electronic nose technology," *Expert Review of Medical Devices*, vol. 2, no. 5, pp. 559–566, 2005.
- [3] M. Falasconi, I. Concina, V. Sberveglieri, G. Zambotti, and E. Gobbi, "Application of chemical gas sensors and data analysis to food analysis," *Sensor Letters*, vol. 12, no. 6-7, pp. 1119–1122, 2014.
- [4] E. Schaller, J. O. Bosset, and F. Escher, "Electronic noses' and their application to food," *LWT—Food Science and Technology*, vol. 31, no. 4, pp. 305–316, 1998.
- [5] N. Docquier and S. Candel, "Combustion control and sensors: a review," *Progress in Energy and Combustion Science*, vol. 28, no. 2, pp. 107–150, 2002.
- [6] D.-D. Lee, "Environmental gas sensors," *IEEE Sensors Journal*, vol. 1, no. 3, pp. 214–224, 2001.
- [7] O. A. Sadik, A. K. Wanekaya, and S. Andreescu, "Advances in analytical technologies for environmental protection and public safety," *Journal of Environmental Monitoring*, vol. 6, no. 6, pp. 513–522, 2004.
- [8] Y. Liu, J. Parisi, X. Sun, and Y. Lei, "Solid-state gas sensors for high temperature applications—a review," *Journal of Materials Chemistry A*, vol. 2, no. 26, pp. 9919–9943, 2014.
- [9] S. Akbar, P. Dutta, and C. Lee, "High-temperature ceramic gas sensors: a review," *International Journal of Applied Ceramic Technology*, vol. 3, no. 4, pp. 302–311, 2006.
- [10] P. C. Jurs, G. A. Bakken, and H. E. McClelland, "Computational methods for the analysis of chemical sensor array data from volatile analytes," *Chemical Reviews*, vol. 100, no. 7, pp. 2649–2678, 2000.
- [11] J. W. Gardner, "Detection of vapours and odours from a multisensor array using pattern recognition Part 1. Principal component and cluster analysis," *Sensors and Actuators B: Chemical*, vol. 4, no. 1-2, pp. 109–115, 1991.
- [12] L. Zhang, F. Tian, X. Peng et al., "Standardization of metal oxide sensor array using artificial neural networks through experimental design," *Sensors and Actuators B: Chemical*, vol. 177, pp. 947–955, 2013.
- [13] W. Maziarz, P. Potempa, A. Sutor, and T. Pisarkiewicz, "Dynamic response of a semiconductor gas sensor analysed with the help of fuzzy logic," *Thin Solid Films*, vol. 436, no. 1, pp. 127–131, 2003.
- [14] Y. Lu, C. Partridge, M. Meyyappan, and J. Li, "A carbon nanotube sensor array for sensitive gas discrimination using principal component analysis," *Journal of Electroanalytical Chemistry*, vol. 593, no. 1-2, pp. 105–110, 2006.
- [15] G. Peng, U. Tisch, O. Adams et al., "Diagnosing lung cancer in exhaled breath using gold nanoparticles," *Nature Nanotechnology*, vol. 4, no. 10, pp. 669–673, 2009.
- [16] J. Chen, K. Wang, and W. Zhou, "Vertically aligned ZnO nanorod arrays coated with  $\text{SnO}_2$ /noble metal nanoparticles for highly sensitive and selective gas detection," *IEEE Transactions on Nanotechnology*, vol. 10, no. 5, pp. 968–974, 2011.
- [17] S. Vallejos, P. Umek, T. Stoycheva et al., "Single-step deposition of au- and pt-nanoparticle-functionalized tungsten oxide nanoneedles synthesized via aerosol-assisted CVD, and used for fabrication of selective gas microsensor arrays," *Advanced Functional Materials*, vol. 23, no. 10, pp. 1313–1322, 2013.
- [18] A. A. Tomchenko, G. P. Harmer, B. T. Marquis, and J. W. Allen, "Semiconducting metal oxide sensor array for the selective detection of combustion gases," *Sensors and Actuators B: Chemical*, vol. 93, no. 1–3, pp. 126–134, 2003.
- [19] P. Ciosek and W. Wróblewski, "The analysis of sensor array data with various pattern recognition techniques," *Sensors and Actuators, B: Chemical*, vol. 114, no. 1, pp. 85–93, 2006.
- [20] K. Zhou, Q. Shuai, and J. Tang, "Adaptive damage detection using tunable piezoelectric admittance sensor and intelligent inference," in *Proceedings of the ASME Conference on Smart Materials, Adaptive Structures and Intelligent Systems (SMASIS '14)*, 2014.
- [21] K. Zhou, G. Liang, and J. Tang, "Efficient model updating using Bayesian probabilistic framework based on measured vibratory response," in *Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*, vol. 9063 of *Proceedings of SPIE*, The International Society for Optical Engineering, 2014.
- [22] J. L. Beck and L. S. Katafygiotis, "Updating models and their uncertainties. I. Bayesian statistical framework," *Journal of Engineering Mechanics*, vol. 124, no. 4, pp. 455–461, 1998.
- [23] Y. X. Liu, Y. Ding, H. Y. Gao et al., " $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$  nanofibers for in situ, real-time, and stable high temperature oxygen sensing," *RSC Advances*, vol. 2, no. 9, pp. 3872–3877, 2012.
- [24] Y. Liu, Y. Ding, L. Zhang, P. X. Gao, and Y. Lei, " $\text{CeO}_2$  nanofibers for in situ  $\text{O}_2$  and  $\text{CO}$  sensing in harsh environments," *RSC Advances*, vol. 2, no. 12, pp. 5193–5198, 2012.
- [25] Y. Liu, X. Sun, B. Li, and Y. Lei, "Tunable p-n transition behaviour of a p- $\text{La}_{0.67}\text{Sr}_{0.33}\text{MnO}_3$ /n- $\text{CeO}_2$  nanofibers heterojunction for the development of selective high temperature propane sensors," *Journal of Materials Chemistry A*, vol. 2, no. 30, pp. 11651–11659, 2014.
- [26] Y. Liu, X. Sun, Z. Zhou, and Y. Lei, "Electrospun Ce–Ni–O composite nanofibers for highly selective propane detection at high temperature based on its rapid reaction kinetics," *Journal of Materials Chemistry A*, vol. 2, no. 34, pp. 14038–14047, 2014.

- [27] Y. Liu and Y. Lei, "Pt-CeO<sub>2</sub> nanofibers based high-frequency impedancemetric gas sensor for selective CO and C<sub>3</sub>H<sub>8</sub> detection in high-temperature harsh environment," *Sensors and Actuators, B: Chemical*, vol. 188, pp. 1141–1147, 2013.
- [28] H.-K. Hong, H. W. Shin, H. S. Park et al., "Gas identification using micro gas sensor array and neural-network pattern recognition," *Sensors and Actuators B: Chemical*, vol. 33, no. 1–3, pp. 68–71, 1996.
- [29] A. Bermak and S. B. Belhouari, "Bayesian learning using Gaussian process for gas identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 3, pp. 787–792, 2006.
- [30] Q. Li, J.-Y. Yu, B.-C. Mu, and X.-D. Sun, "BP neural network prediction of the mechanical properties of porous NiTi shape memory alloy prepared by thermal explosion reaction," *Materials Science and Engineering A*, vol. 419, no. 1-2, pp. 214–217, 2006.
- [31] P. G. Guest, *Numerical Methods of Curve Fitting*, Cambridge University Press, Cambridge, UK, 2012.

