

# Soft Sensor Development for Multimode Processes Based on Semisupervised Gaussian Mixture Models <sup>\*</sup>

Weiming Shao, Zhihuan Song, Le Yao

*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: shaowm@zju.edu.cn; songzhihuan@zju.edu.cn; yaole\_frank@zju.edu.cn).*

---

**Abstract:** The Gaussian mixture models (GMM) is an effective tool for modeling processes with multiple operating modes that widely exist in industrial process systems. Traditional supervised version of GMM, namely the Gaussian mixture regression (GMR), for developing soft sensors merely relies on the labeled samples. However, labeled samples in the soft sensor application are usually very infrequent due to economical or technical limitations, which may lead the GMR to unreliable parameter estimation and finally poor performance for predicting the primary variable. To tackle this problem, a semisupervised GMM for regression purpose is proposed, where both labeled and unlabeled samples take effect, and the Gaussian parameters and regression coefficients are learned simultaneously based on the expectation-maximization algorithm. Two case studies are carried out using simulated dataset and real-life dataset collected from a primary reformer in an ammonia synthesis process, which demonstrates the effectiveness of the proposed method.

Keywords: Soft sensor, multimode process, semisupervised Gaussian mixture models, Gaussian mixture regression, expectation-maximization.

---

## 1. INTRODUCTION

In industrial processes, many quality-related variables called the primary variables are measured with large delay in the laboratory or with high price using analyzers. Soft sensors are popular alternatives of the lab analysis or analyser, since they are delay-free and very low-cost (Ge et al. (2017)). As process data that reflect the true process conditions become available, during past decades a variety of data-driven algorithms have been applied to soft sensor development, such as the principal component regression, partial least squares, artificial neural networks, support vector machines, and so forth. Extensive reviews of the development algorithms and application surveys of soft sensors in industrial processes can be found in (Kadlec et al. (2009); Kano and Ogawa (2009); Ge (2017)).

Processes with multiple operating modes widely exist in industrial process systems, which may result from multiple product grade demands, changes in feedstocks, load variations or seasonal operations (Souza and Araújo (2014)). Developing soft sensors for these processes needs to take into account the characteristics of the multiple operating modes that make the processes exhibit non-Gaussian behaviors. As a result, a single model may fail to perform satisfactorily and multiple models accounting for each mode are required. In addition, due to the measurement variations and transmission disturbances, industrial processes

are inherently random processes (Yuan et al. (2017)), thus probabilistic models are more proper to handle the uncertainties compared to their deterministic counterparts.

The Gaussian mixture models (GMM) is well known to be effective in modeling the multimode characteristics and accounting for the process uncertainties simultaneously, and has recently established itself as a popular tool for developing soft sensors for processes with multiple operating conditions. The GMM-based soft sensors can be generally categorized into two groups. In the first one, the secondary and primary variables are separately manipulated, and the unsupervised GMM is used for mode identification. Then localized models are trained for each mode using regression algorithms such as the kernel partial least squares (Yu (2012)) and Gaussian process regression (Grbić et al. (2013)). In the other group, the supervised GMM, namely the Gaussian mixture regression (GMR), is developed, which treats the secondary and primary variables together, and learns their joint probability distribution function (PDF) in each mode. The functional relationship for the soft sensor can be derived directly from the joint PDF. For instance, Yuan et al. (2014) trained soft sensors using the GMR for multimode/multiphase processes, which shows that the GMR outperforms the unsupervised GMM, because the GMR learns model parameters together instead of separately, and thus is not constrained by the number of samples in each mode. Zhu et al. (2017) developed a variational GMR to model non-Gaussian processes. Their studies indicate that the Bayesian treatment can not only

---

<sup>\*</sup> This work was supported by the National Key Research and Development Program of China (2017YFB0304203) and by the Natural Science Foundation of China (61703367).

automatically determine the number of Gaussians, but also improve the performance compared to the standard GMR.

Although the GMR outperforms the unsupervised GMM, it requires sufficient labeled samples (for which both the second and primary variables are known). However, in soft sensor applications, labelling samples could be expensive and infrequent due to certain economical or technical limitations. As a result, labeled samples are usually rare and the success of the GMR could not be guaranteed, because insufficient samples often lead to unreliable estimation of the PDF, especially when the dimensionality of process variables is high. On the contrary, there are large amounts of unlabeled samples (for which only the secondary variables are known), which also contain useful information yet have not been utilized in the GMR. Semisupervised learning exploiting both labeled and unlabeled samples, has been confirmed effective for soft sensor development in remedying the insufficiency of labeled samples (Yao and Ge (2018)). Moreover, the validity of semisupervised GMM for classification applications has also been demonstrated (Xing et al. (2013); Yan et al. (2017)). Unfortunately, to our best knowledge, no work on semisupervised GMM for regression task has been reported, in particular for soft sensor modeling.

Therefore, in order to deal with the above analyzed deficiency of the GMR in the soft sensor application, this paper proposes a semisupervised GMM (S<sup>2</sup>GMM) for regression purpose, where both labeled and unlabeled samples take effect, and the PDF parameters and regression coefficients in each mode are learned simultaneously by the expectation-maximization (EM) algorithm.

## 2. SEMISUPERVISED GAUSSIAN MIXTURE MODELS

### 2.1 Formulation of the S<sup>2</sup>GMM

Let  $\mathbf{x} \in \mathbb{R}^d$  and  $y \in \mathbb{R}$  be the  $d$ -dimensional input variable and scalar output variable, respectively, and  $(\mathbf{X}_l, \mathbf{Y}_l) = (\mathbf{x}_i, y_i)_{i=1}^{n_l}$  and  $(\mathbf{X}_u) = (\mathbf{x}_j)_{j=1}^{n_u}$  denote the labeled and unlabeled dataset, respectively. Here  $n_l$  and  $n_u$  are the numbers of labeled and unlabeled samples, respectively. Assume there are a total of  $K$  Gaussian components, and for the  $k$ -th component, the PDF of  $\mathbf{x}$  and the functional dependence of  $y$  on  $\mathbf{x}$  are defined as:

$$P_k(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x) \quad (1)$$

$$y = \mathbf{w}_k^T \mathbf{x} + w_k^0 + \varepsilon_k$$

where  $P_k(\mathbf{x})$  means the PDF of  $\mathbf{x}$ ,  $\mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)$  stands for the Gaussian distribution over  $\mathbf{x}$  with mean vector  $\boldsymbol{\mu}_k^x$  and covariance matrix  $\boldsymbol{\Sigma}_k^x$ ,  $\mathbf{w}_k$  and  $w_k^0$  are the regression coefficients between  $\mathbf{x}$  and  $y$ ,  $\varepsilon_k \sim \mathcal{N}(0, \sigma_k^2)$  denotes the measurement noise of  $y$ , respectively, for the  $k$ -th component. By linear Gaussian operations, the joint PDF of  $\mathbf{x}$  and  $y$  for the  $k$ -th component can be obtained as

$$P_k(\mathbf{x}, y) = \mathcal{N}(\mathbf{x}, y | \boldsymbol{\mu}_k^{xy}, \boldsymbol{\Sigma}_k^{xy}) \quad (2)$$

where

$$\boldsymbol{\mu}_k^{xy} = \begin{bmatrix} \boldsymbol{\mu}_k^x \\ \mathbf{w}_k^T \boldsymbol{\mu}_k^x + w_k^0 \end{bmatrix}, \boldsymbol{\Sigma}_k^{xy} = \begin{bmatrix} \boldsymbol{\Sigma}_k^x & \boldsymbol{\Sigma}_k^x \mathbf{w}_k \\ \mathbf{w}_k^T \boldsymbol{\Sigma}_k^x & \mathbf{w}_k^T \boldsymbol{\Sigma}_k^x \mathbf{w}_k + \sigma_k^2 \end{bmatrix}.$$

Also, the conditional PDF of  $y$  given  $\mathbf{x}$  is calculated as

$$P_k(y | \mathbf{x}) = \mathcal{N}(y | \mathbf{w}_k^T \mathbf{x} + w_k^0, \sigma_k^2) \quad (3)$$

Therefore, in the S<sup>2</sup>GMM the PDF for labeled and unlabeled samples can be expressed as

$$P(\mathbf{x}_i, y_i) = \sum_{k=1}^K \alpha_k P_k(\mathbf{x}_i, y_i) \quad (4)$$

$$P(\mathbf{x}_j) = \sum_{k=1}^K \alpha_k P_k(\mathbf{x}_j) \quad (5)$$

where  $\alpha_k = P(\mathbf{z}_i = k) = P(\mathbf{z}_j = k)$  is the prior of the  $k$ -th component for  $i = 1, \dots, n_l$  and  $j = 1, \dots, n_u$ . Here  $\mathbf{z}_i$  and  $\mathbf{z}_j$  are the latent variables associated with the  $i$ -th labeled sample and  $j$ -th unlabeled sample, respectively. Again, using the linear Gaussian operation, the posterior distributions over the latent variables given the corresponding labeled and unlabeled sample are calculated as per

$$P(\mathbf{z}_i = k | \mathbf{x}_i, y_i) = \frac{\alpha_k \mathcal{N}(\mathbf{x}_i, y_i | \boldsymbol{\mu}_k^{xy}, \boldsymbol{\Sigma}_k^{xy})}{\sum_{k=1}^K \alpha_k \mathcal{N}(\mathbf{x}_i, y_i | \boldsymbol{\mu}_k^{xy}, \boldsymbol{\Sigma}_k^{xy})} \equiv R_{ik}^l \quad (6)$$

$$P(\mathbf{z}_j = k | \mathbf{x}_j) = \frac{\alpha_k \mathcal{N}(\mathbf{x}_j | \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)}{\sum_{k=1}^K \alpha_k \mathcal{N}(\mathbf{x}_j | \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)} \equiv R_{jk}^u \quad (7)$$

where  $R_{ik}$  and  $R_{jk}$  represent the posterior responsibilities of the  $k$ -th component for generating the  $i$ -th labeled sample and  $j$ -th unlabeled sample, respectively.

### 2.2 Parameter learning for the S<sup>2</sup>GMM

For the S<sup>2</sup>GMM, the model parameters that need to be learned are denoted as  $\boldsymbol{\Theta} = (\alpha_k, \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x, \mathbf{w}_k, w_k^0, \sigma_k^2)_{k=1}^K$ . In this paper, we develop an efficient way of fulfilling such task based on the EM algorithm.

The complete log-likelihood function is formulated as

$$L(\boldsymbol{\Theta}) = \sum_{\mathbf{Z}} P(\mathbf{Z} | \mathbf{X}_l, \mathbf{Y}_l, \mathbf{X}_u) \ln P(\mathbf{X}_l, \mathbf{Y}_l, \mathbf{X}_u, \mathbf{Z})$$

$$= \sum_{i=1}^{n_l} \sum_{k=1}^K R_{ik}^l \ln P_k(y_i | \mathbf{x}_i) + \sum_{i=1}^{n_l} \sum_{k=1}^K R_{ik}^l \ln P_k(\mathbf{x}_i)$$

$$+ \sum_{j=1}^{n_u} \sum_{k=1}^K R_{jk}^u \ln P_k(\mathbf{x}_j) + \sum_{i=1}^{n_l} \sum_{k=1}^K R_{ik}^l \ln \alpha_k$$

$$+ \sum_{j=1}^{n_u} \sum_{k=1}^K R_{jk}^u \ln \alpha_k + \text{const} \quad (8)$$

where  $\mathbf{Z} = (\mathbf{Z}_l, \mathbf{Z}_u)$ , and  $\mathbf{Z}_l = (\mathbf{z}_i)_{i=1}^{n_l}$  and  $\mathbf{Z}_u = (\mathbf{z}_j)_{j=1}^{n_u}$  are the latent variable sets corresponding to  $(\mathbf{X}_l, \mathbf{Y}_l)$  and  $\mathbf{X}_u$ , respectively. Setting the derivatives of  $L(\boldsymbol{\Theta})$  with respect to those model parameters except  $(\alpha_k)_{k=1}^K$  to zeros can get their iteration equations. Specifically,

$$\frac{\partial L(\Theta)}{\partial \tilde{\mathbf{w}}_k} = \mathbf{0} \implies \tilde{\mathbf{w}}_k = \left( \tilde{\mathbf{X}}_l^T \Omega_k^l \tilde{\mathbf{X}}_l \right)^{-1} \tilde{\mathbf{X}}_l^T \Omega_k^l \mathbf{Y}_l \quad (9)$$

where  $\tilde{\mathbf{w}}_k = \begin{bmatrix} \mathbf{w}_k \\ w_k^0 \end{bmatrix}$ ,  $\Omega_k^l = \text{diag}(R_{1k}^l, \dots, R_{n_l k}^l)$ ,  $\tilde{\mathbf{X}}_l = (\mathbf{X}_l, \mathbf{1})$ , and  $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^{n_l}$ . Similarly, we can get

$$\sigma_k^2 = \sum_{i=1}^{n_l} R_{ik}^l (y_i - \mathbf{w}_k^T \mathbf{x}_i - w_k^0)^2 \Big/ \sum_{i=1}^{n_l} R_{ik}^l \quad (10)$$

$$\boldsymbol{\mu}_k^x = \frac{\sum_{i=1}^{n_l} R_{ik}^l \mathbf{x}_i + \sum_{j=1}^{n_u} R_{jk}^u \mathbf{x}_j}{\sum_{i=1}^{n_l} R_{ik}^l + \sum_{j=1}^{n_u} R_{jk}^u} \quad (11)$$

$$\boldsymbol{\Sigma}_k^x = \frac{\sum_{i=1}^{n_l} R_{ik}^l \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T + \sum_{j=1}^{n_u} R_{jk}^u \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j^T}{\sum_{i=1}^{n_l} R_{ik}^l + \sum_{j=1}^{n_u} R_{jk}^u} \quad (12)$$

where  $\bar{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}_k^x$ ,  $\bar{\mathbf{x}}_j = \mathbf{x}_j - \boldsymbol{\mu}_k^x$ .

Using the Lagrange multiplier that combines  $L(\Theta)$  and the constraint  $\sum_{k=1}^K \alpha_k = 1$  leads to

$$\alpha_k = \frac{\sum_{i=1}^{n_l} R_{ik}^l + \sum_{j=1}^{n_u} R_{jk}^u}{n_l + n_u} \quad (13)$$

Detailed derivations for Eqs.(8)~(13) are omitted here, and one can refer to Bishop (2006) for fundamentals. The procedure for parameter learning for the S<sup>2</sup>GMM based on the EM algorithm are summarized as *Algorithm 1*.

*Algorithm 1* Pseudocode for parameter learning for the S<sup>2</sup>GMM

---

Step 1): Initialize  $\Theta = (\alpha_k, \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x, \mathbf{w}_k, w_k^0, \sigma_k^2)_{k=1}^K$ .  
 Step 2): Repeat 3)~4) until the convergence criterion is satisfied.  
 Step 3): *E*-step  
 for  $k = 1, \dots, K; i = 1, \dots, n_l; j = 1, \dots, n_u$   
     Calculate  $R_{ik}^l$  with Eq.(6).  
     Calculate  $R_{jk}^u$  with Eq.(7).  
 end for  
 Step 4): *M*-step  
 for  $k = 1, \dots, K$   
     Update  $\tilde{\mathbf{w}}_k$  with Eq.(9).  
     Update  $\sigma_k^2$  with Eq.(10).  
     Update  $\boldsymbol{\mu}_k^x$  with Eq.(11).  
     Update  $\boldsymbol{\Sigma}_k^x$  with Eq.(12).  
     Update  $\alpha_k$  with Eq.(13).  
 end for

---

In *Algorithm 1*, the convergence criterion can be either the increment of the log-likelihood function, or the maximum iteration number. Also, from *Algorithm 1* it can be seen that in the S<sup>2</sup>GMM, learning  $(\alpha_k, \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)_{k=1}^K$  involves explicitly both labeled and unlabeled samples, while unlabeled samples take effect implicitly in the learning of  $(\tilde{\mathbf{w}}_k, \sigma_k^2)_{k=1}^K$  through  $R_{ik}^l$  ( $k = 1, \dots, K; i = 1, \dots, n_l$ ).

### 3. SOFT SENSOR DEVELOPMENT BASED ON S<sup>2</sup>GMM

Based on the S<sup>2</sup>GMM, a soft sensor model can be developed for estimating the true value ( $y_q$ ) of the primary variable when a sample ( $\mathbf{x}_q$ ) of the secondary variable is available according to the following process.

The posterior distribution over the latent variable ( $\mathbf{z}_q$ ) given  $\mathbf{x}_q$  is calculated according to

$$\begin{aligned} P(\mathbf{z}_q = k | \mathbf{x}_q) &= \frac{P(\mathbf{z}_q = k) P(\mathbf{x}_q | \mathbf{z}_q = k)}{\sum_{k=1}^K P(\mathbf{z}_q = k) P(\mathbf{x}_q | \mathbf{z}_q = k)} \\ &= \frac{\alpha_k \mathcal{N}(\mathbf{x}_q | \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)}{\sum_{k=1}^K \alpha_k \mathcal{N}(\mathbf{x}_q | \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x)} \end{aligned} \quad (14)$$

For simplicity,  $P(\mathbf{z}_q = k | \mathbf{x}_q)$  is denoted as  $R_{qk}^u$ . Then, the conditional PDF of  $y_q$  given  $\mathbf{x}_q$  is computed as

$$\begin{aligned} P(y_q | \mathbf{x}_q) &= \sum_{k=1}^K P(\mathbf{z}_q = k | \mathbf{x}_q) P(y_q | \mathbf{x}_q, \mathbf{z}_q = k) \\ &= \sum_{k=1}^K R_{qk}^u \mathcal{N}(y_q | \mathbf{w}_k^T \mathbf{x}_q + w_k^0, \sigma_k^2) \end{aligned} \quad (15)$$

Therefore, the estimation of  $y_q$  is determined as

$$\hat{y}_q = \mathbb{E}[y_q | \mathbf{x}_q] = \sum_{k=1}^K R_{qk}^u (\mathbf{w}_k^T \mathbf{x}_q + w_k^0) \quad (16)$$

### 4. CASE STUDIES

In this section, the S<sup>2</sup>GMM is first investigated using a numerical example and then applied to developing soft sensor for a real-life industrial primary reformer in an ammonia synthesis process. For comparison purpose, the performance of the GMR (Yuan et al. (2014)) and the popular partial least squares (PLS) (Lindgren et al. (1993)) are also presented. Parameter initialization in the learning procedure for the GMR and S<sup>2</sup>GMM are aided by the  $k$ -means clustering method following the suggestion of Bishop (Bishop (2006)). In addition, certain proportion of training samples are chosen as labeled samples, and the rest of training samples are treated as unlabeled. To deal with the randomness in the initialization procedure for the GMR and S<sup>2</sup>GMM, 100 independent simulations are run for each labeling rate. The prediction accuracy is measured using the averaged root mean square error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n_t} (y_t - \hat{y}_t)^2 / n_t} \quad (17)$$

where  $y_t$  and  $\hat{y}_t$  represent the true and predicted labels of the  $t$ -th testing sample, respectively, and  $n_t$  stands for the number of testing samples.

#### 4.1 Numerical Example

Assume a 2-dimensional input vector  $\mathbf{x} = (x_1, x_2)^T$  and a scalar output  $y$  follow the relationship described in Eq.(1)

with 3 Gaussian components, where the configurations of each Gaussian component are listed in Table 1, and the data distributions are visualized in Fig. 1 in the input space, which presents clear multimode characteristics. In the simulation, a total of 200 samples are generated for model parameter learning and 1000 samples are generated for performance evaluation.

Table 1. Configurations of the three Gaussian components

	$k = 1$	$k = 1$	$k = 3$
$\alpha_k$	20%	30%	50%
$\mu_k^x$	$(0,2)^T$	$(4,6)^T$	$(4,0)^T$
$\Sigma_k^x$	$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$	$\begin{bmatrix} 3 & -1 \\ -1 & 1.5 \end{bmatrix}$
$w_k$	$(1,1)^T$	$(1,-1)^T$	$(-1,1)^T$
$w_k^0$	0	0	0
$\sigma_k^2$	0.5	0.5	0.5

Scatter plot comparisons among the three investigated soft sensing methods are illustrated in Fig. 2, where the labeling rate is 10%. It is found that in the prediction results obtained by the PLS, significant bias occurs in all the three modes, implying that the PLS doesn't model any mode well. The reason is that the PLS is a linear algorithm and can only deal with Gaussian or approximated Gaussian distributions, but this multimode example is strongly nonlinear and non-Gaussian. In contrast, the GMR and S<sup>2</sup>GMM show more powerful abilities in capturing the multimode characteristics. However, the GMR doesn't perform well in the area where the output is greater than 0, because this area corresponds to the first component where the data distribution is longer and narrower, and is easily disturbed by the other two components, which can be inferred from Fig. 1.

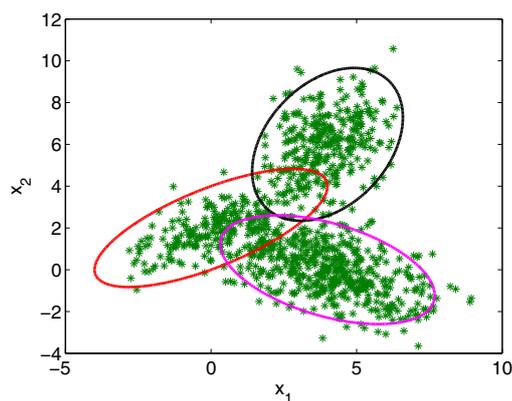


Fig. 1. Visualization of the numerical example in the input space using the testing dataset.

The average predictive RMSE comparisons are presented in Table 2, which indicates that the addition of labeled samples doesn't help the PLS too much, due to its limitation in dealing with the non-Gaussianity. Compared with the PLS, both the GMR and S<sup>2</sup>GMM demonstrate evident advantages, and when the labeling rates are 40% and 50%,

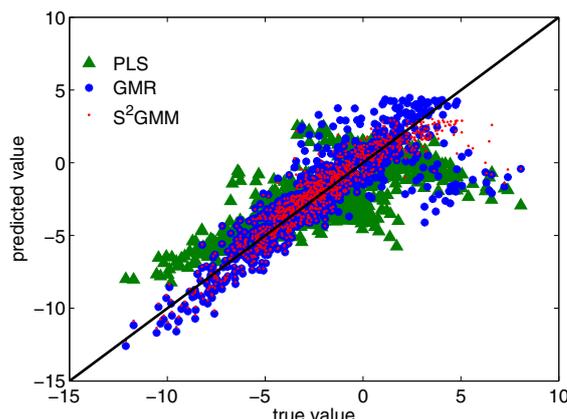


Fig. 2. Scatter plot comparisons among PLS, GMR and S<sup>2</sup>GMM.

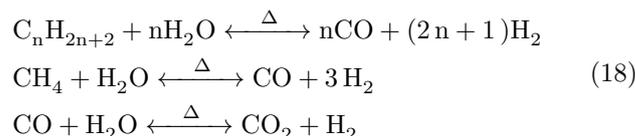
the predictive accuracies of the GMR and S<sup>2</sup>GMM are almost identical. However, when the labeling rate is reduced to be below 40%, the performance of the GMR starts to deteriorate, especially when the labeling rate decreases to 10%. By contrast, as the labeling rate decreases, the performance of the S<sup>2</sup>GMM deteriorates relatively slowly and the deterioration is much smaller. For examples, when the labeling rate decreases from 50% to 10%, the deteriorations for the GMR and S<sup>2</sup>GMM are 55.7% and 10.6%, respectively. Thus, in this numerical example it can be concluded that the S<sup>2</sup>GMM outperforms the GMR with small amount of labeled samples, which verifies the benefit of incorporating those unlabelled samples.

Table 2. Average predictive RMSE for the numerical example by PLS, GMR and S<sup>2</sup>GMM

labeling rate	PLS	GMR	S <sup>2</sup> GMM
10%	2.1943	1.6776	1.1956
20%	2.0632	1.3890	1.1839
30%	2.1567	1.3184	1.0876
40%	2.0939	1.0101	1.0372
50%	2.0617	1.0776	1.0815

#### 4.2 Primary Reformer

The primary reformer shown in Fig. 3 comes from the hydrogen manufacturing units in the ammonia synthesis process. It transforms the desulphurized natural gas into crude synthesis gas for ammonia production in a follow-up unit through the following reaction with nickel catalyst:



Reaction temperature is a pivotal factor to keep the chemical reaction described in Eq. (18) stable. In an ammonia production plant from the *China National Offshore Oil Company*, the reaction temperature is controlled at a certain level by manipulating the burning condition in the furnace, which is realized by monitoring the concentration of the oxygen at the top of the primary reformer

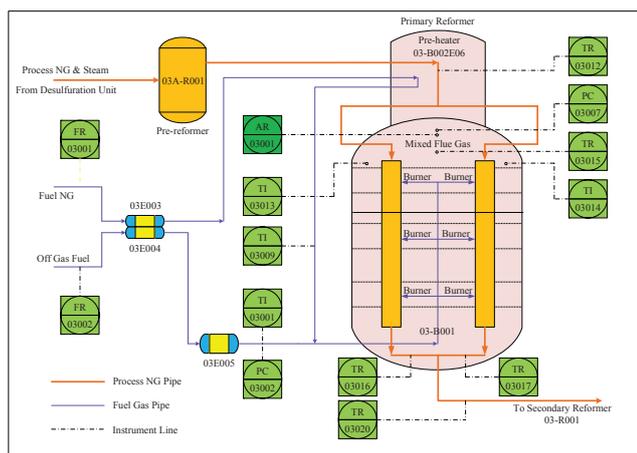


Fig. 3. Flow chart of the primary reformer.

measured by an expensive mass spectrometer (AR03001). Thus, a soft sensor is desired for online estimating the oxygen concentration at the top of the primary reformer. Secondary variables of the soft sensor are selected using expert knowledge from the field engineers (Yao and Ge (2017)), which are presented in Table 3.

Table 3. Descriptions of selected secondary variables for soft sensing  $O_2$  concentration

Tags	Descriptions
FR03001.PV	Flow of fuel natural gas
FR03002.PV	Flow of fuel off gas
PC03002.PV	Pressure of fuel off gas
PC03007.PV	Pressure of furnace flue gas
TI03001.PV	Temperature of fuel off gas
TI03009.PV	Temperature of fuel natural gas
TR03012.PV	Temperature of process gas
TI03013.PV	Temperature of furnace flue gas
TI03014.PV	Temperature of furnace flue gas
TR03015.PV	Temperature of mixed furnace flue gas
TR03016.PV	Temperature of transformed gas
TR03017.PV	Temperature of transformed gas
TR03020.PV	Temperature of transformed gas

Around 1500 samples for soft sensor development have been selected from the database of the distributed control systems in a real-life ammonia production plant. Those samples are evenly partitioned as training ones and testing ones. The primary reformer is a multimode process due to the operating conditions, which are shown in Fig. 4 using the testing samples.

The number of Gaussian components for both the GMR and  $S^2$ GMM are roughly pre-defined as 3 according to the operating conditions. Note that some criteria, such as the Akaike information criterion (Yan et al. (2017)) and the absolute increment of the log-likelihood (Yuan et al. (2014)), can also be used for automatically determining the number of Gaussian components. Predictions achieved by the PLS, GMR and  $S^2$ GMM-based soft sensors for the oxygen concentration at their average performance

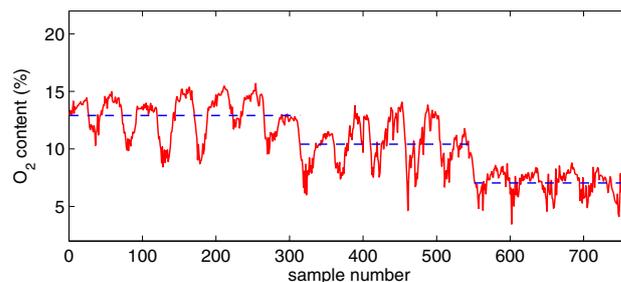


Fig. 4. Illustration of the multimode characteristics of primary reformer in terms of  $O_2$  content.

levels with labeling rate = 20% are presented as Fig. 5(a), Fig. 5(b) and Fig. 5(c), respectively.

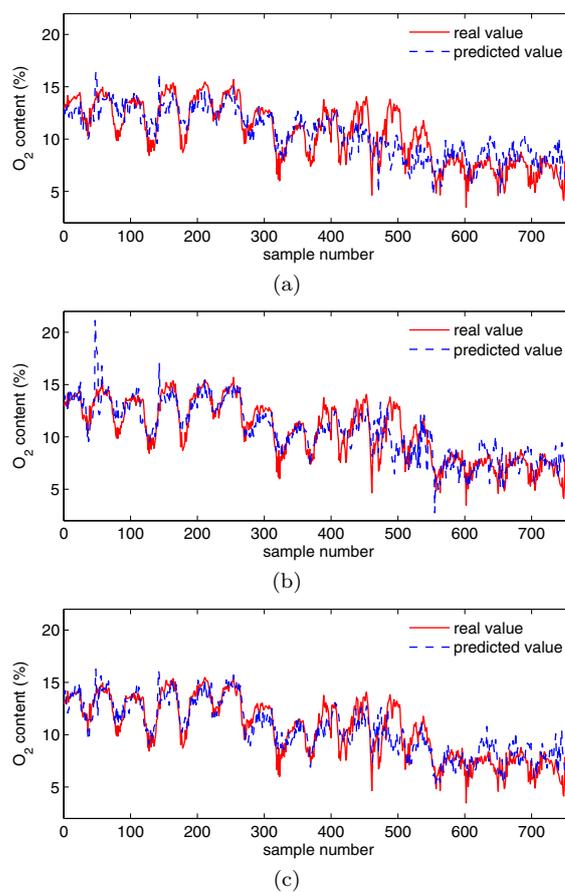


Fig. 5. Predictions of the oxygen concentration by: (a) PLS; (b) GRM; (c)  $S^2$ GMM.

It can be seen from Fig. 5(a) that like in the previous numerical example, the PLS could not deal with any mode pertinently, but tends to find a balance among the three modes, which is because of its failure in dealing with the multimode characteristics. In contrast, both the GMR and  $S^2$ GMM perform better than the PLS in capturing multimode characteristics. However, in some areas such as those around 50-th and 550-th sample, the  $S^2$ GMM shows apparent predictive advantages over the GMR.

For further investigations, quantitative prediction accuracies achieved by the three soft sensors for the oxygen content are tabulated in Table 4. Once more, there seems no substantial performance improvement in the PLS through

increasing the labeling rate. But the performance of both the GMR and S<sup>2</sup>GMM get improved as the labeling rate is increased from 20% to 50%, and the S<sup>2</sup>GMM accomplishes higher predictive accuracy than the GMR. These results confirm the effectiveness of the semisupervised learning strategy for the GMM.

However, when the amount of labeled samples are small to certain extent, for example the labeling rate is set as 10% in the primary reformer, both the GMR and S<sup>2</sup>GMM suffer from the drawback of the standard GMM model, i.e., the numerical issue caused by ill-conditioned covariance matrices. As a result, the inverse of the covariance matrices of the PDF for the Gaussian components become unavailable, which disables the use of the GMR and S<sup>2</sup>GMM as marked with the symbol ‘-’ in Table 4.

Table 4. Average predictive RMSE for the primary reformer by PLS, GMR and S<sup>2</sup>GMM

labeling rate	PLS	GMR	S <sup>2</sup> GMM
10%	1.7271	-	-
20%	1.6708	1.5064	1.4374
30%	1.6525	1.4341	1.4147
40%	1.6772	1.4271	1.4070
50%	1.6570	1.4011	1.3897

## 5. CONCLUSION

In this paper, we have proposed a semisupervised version of the GMM (S<sup>2</sup>GMM) for regression applications, which is able to mine the information contained in both labeled and unlabeled samples. The S<sup>2</sup>GMM has been applied to developing soft sensor for the multimode process, and two case studies including a numerical example and a real-life chemical process have confirmed the superiorities of the S<sup>2</sup>GMM over the popular PLS and the supervised GMM (GMR).

However, in the second case study we have encountered a drawback of the S<sup>2</sup>GMM and GMR with too small amount of labeled samples and high-dimensional process variables, namely the numerical issue. A potential way of solving or alleviating this problem is to form the variational Bayesian S<sup>2</sup>GMM or a simpler S<sup>2</sup>GMM with Bayesian regularization, which both treat the regression coefficients for each Gaussian component as random variables. This will be our future work.

## REFERENCES

Bishop C.M., *Pattern recognition and machine learning*. Springer, New York, 2006.

Ge Z., Song Z., Ding S.X. and Huang B., Data mining and analytics in the process industry: the role of machine learning. *IEEE Access*, 5(99):20590-20616, 2017.

Ge Z., Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 171:16-25, 2017.

Grić R., Slišković D. and Kadlec P., Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. *Computers and Chemical Engineering*, 58(22):84-97, 2013.

Kadlec B., Gabrys B. and Strandt S., Data-driven soft sensors in the process industry. *Computers and Chemical Engineering*, 33(4):795-814, 2009.

Kano M. and Ogawa M., The state of the art in chemical process control in Japan: good practice and questionnaire survey. *Journal of Process Control*, 20(9):969C982, 2010.

Lindgren R., Geladi P. and Wold S., The kernel algorithm for PLS. *Journal of Chemometrics*, 7(1):45-59, 1993.

Souza A.A.F. and Araújo R., Mixture of partial least squares experts and application in prediction settings with multiple operating modes. *Chemometrics and Intelligent Laboratory Systems*, 130(2):192-202, 2014.

Xing X., Yu Y., Jian H. and Du S., A multi-manifold semisupervised Gaussian mixture model for pattern classification. *Pattern Recognition Letters*, 34(16):2118-2125, 2013.

Yan H., Zhou J. and Pang C., Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories. *IEEE Transactions on Instrumentation and Measurement*, 64(4):723-733, 2017.

Yao L. and Ge Z., Moving window adaptive soft sensor for state shifting process based on weighted supervised latent factor analysis. *Control Engineering Practice*, 61:72-80, 2017.

Yao L. and Ge Z., Deep learning of semi-supervised process data with hierarchical extreme learning machine and soft sensor application. *IEEE Transactions on Industrial Electronics*, 65:1490-1498, 2018.

Yu J., Multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes. *Industrial and Engineering Chemistry Research*, 51(40):13227-13237, 2012.

Yuan X., Ge Z. and Song Z., Soft sensor model development in multiphase/multimode processes based on Gaussian mixture models. *Chemometrics and Intelligent Laboratory Systems*, 138(6):97-109, 2014.

Yuan X., Ge Z., Huang B. and Song Z., A probabilistic just-in-time learning framework for soft sensor development with missing data. *IEEE Transactions on Control Systems and Technology*, 25(3):1124-1132, 2017.

Zhu J., Ge Z. and Song Z., Variational Bayesian Gaussian mixture Regression for soft sensing key variables in non-gaussian industrial processes. *IEEE Transactions on Control Systems and Technology*, 25(3):1092-1099, 2017.