

# On-line Calibration of Just in Time Learning and Gaussian Process Regression based Soft Sensor with Moving-Window Technology

Congli Mei<sup>a,\*</sup>, Xu Chen<sup>b</sup>, Yuhang Ding<sup>b</sup>, Yao Chen<sup>b</sup>, Jiangpin Cai<sup>a</sup>, Yunxia Luo<sup>a</sup>

<sup>a</sup>College of Electrical Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, P.R. China

<sup>b</sup>Department of Automation, Jiangsu University, Zhenjiang 212013, P.R. China  
 clmei@126.com

To handle time-varying, non-linear and multi-parameter characteristics of industrial processes, a new soft sensor modelling method by Gaussian process regression (GPR) with just in time learning (JITL) and moving window technology is proposed. Traditional soft sensors based on JITL only consider spatial characteristic of the query data point and select the best similar samples from a historical database for modelling, ignoring local temporal characteristics of industrial processes. That may result in some predictions relying too much on database. In the proposed soft sensor modelling method, firstly, JITL is used to build a GPR-based prediction model which gives output related to query data point. Then, a local temporal GPR-based model is built on the samples within the last given moving window. In the moving window, the prediction given by the JITL model is as the newest sample. Finally, the local GPR-based model is used to calculate output related to the query data point. This method takes into account not only spatial characteristic of a query data point but also local temporal characteristic of real-time process conditions. The proposed soft sensor is validated by an industrial Erythromycin fermentation process simulation. Results show that the proposed method has higher adaptability and predictive performance than traditional JITL based soft sensors.

## 1 Introduction

In fermentation processes, the main problem of quality control is caused by the lack of sufficient real-time feedback values of critical variables, due to lag time or some other technical or economic difficulties (Souza et al., 2016). To handle the problem, soft sensor technology was proposed to replace traditional instruments and has been widely used in process control (Mei et al., 2017a).

However, soft sensors usually have their internal drawbacks. The significant problem of existing soft sensors is that they will inevitably suffer the performance deterioration after serving for some service cycles due to the gradual model mismatch between soft sensors and the process (Kadlec et al., 2011). When deterioration starts, the sensor' performance becomes unreliable (Chen et al., 2015). Therefore, how to ensure long-term reliability is the key to successful application of soft sensors.

To handle the problem, just in time learning (JITL) was introduced to build soft sensors (Saprotor, 2014). JITL is inspired by the ideas from local modelling and database technology (Fujiwara et al., 2009). In the JITL model structure, a local model is built using the most relevant samples from historic data set around a query data point when the estimated value of the sample is required (Mei et al., 2017b). Different from traditional regression methods and recursive modelling methods, which can both be considered as global models, the JITL-based method exhibits a local model structure and it is built online with a lazy learning manner. Thus, the current state of the process can be tracked by the JITL model and then it can cope with the process nonlinearity directly. However, the performance of JIT-based soft sensors is affected by two factors which are similarity assessment and sampling scale (Liu et al., 2012). For simplicity, we take the process variables  $\mathbf{x}$  as input and the target variable  $y$  as output. The traditional way calculating similarity is to employ a distance- or angle-based similarity

index using the input information (Saporo, 2014). Mathematically two similar input values may correspond to two totally different output values due to nonlinearity and process dynamics. But the trade-off between the input's influence and output's influence on the final similarity remains undetermined. In fact, the JITL modelling method is a local modelling method, only depends on spatial similarity but ignores local temporal process characteristics. To enhance the reliability of JIT-based soft sensors, a new soft sensor modelling method is proposed, which combines JITL and moving window technology. The basic idea is that using a local temporal model on a moving window to calibrate predictions given by JIT-based soft sensor. To achieve the goal of calibration, the real-time prediction is used as the newest sample added to the sample window. In modelling, gaussian process regression (GPR) is used as the regression modelling tool for its superiority of describing complex processes with probability characteristics (Mei et al., 2016).

## 2 Proposed modelling strategies based on a moving window and JITL

### 2.1 Gaussian process regression based on a moving window(MWGPR)

In this section, only the parameter update algorithm of MWGPR is briefly introduction (Grbic et al., 2013). As the new complete input-output samples are acquired, the window slides along the data such that the oldest sample is discarded, and new ones are added to the window. In that way not only the new information is added to the model but also memory requirements are constrained by fixed window size.

At some point in online operation, scaled data contained in the window of size  $N$  can be written as

$$D_t = \{[\mathbf{x}_{t+1-N}, \dots, \mathbf{x}_t], [y_{t+1-N}, \dots, y_t]\} \quad (1)$$

With appropriate regularized kernel matrix  $\tilde{\mathbf{K}}_t = (\mathbf{K}_t + \sigma_n^2 \mathbf{1})$ . When new complete input-output sample is acquired  $\{\mathbf{x}_{t+1}, y_{t+1}\}$ , GPR model can be simply updated by removing first column and first row of covariance matrix due to marginalization property of GP:

$$\tilde{\mathbf{K}}_{temp} = [\tilde{\mathbf{K}}_t]_{-1,-1} \quad (2)$$

This new observed is added to the model by adding appropriate row and column in kernel matrix:

$$\tilde{\mathbf{K}}_{t+1} = \begin{bmatrix} \tilde{\mathbf{K}}_{temp} & \mathbf{k}_{t+1} \\ \mathbf{k}_{t+1}^T & k_{t+1} + \sigma_n^2 \end{bmatrix} \quad (3)$$

Where  $[\mathbf{k}_{t+1}]_i = k(\mathbf{x}_i, \mathbf{x}_{t+1})$ ,  $k_{t+1} = k(\mathbf{x}_{t+1}, \mathbf{x}_{t+1})$ , and  $\mathbf{x}_i$  are samples contained in the window. Values of output variable  $y$  contained in the new window should be rescaled (to satisfy zero mean process assumption) since the mean value of output variable  $y$  will change as the oldest sample is discarded and the newest is inserted. New mean value can be efficiently calculated by:

$$\mathbf{b}_{t+1} = \mathbf{b}_t - \frac{1}{N} (y_{t+1-N}^0 - y_{t+1}^0) \quad (4)$$

Where the superscript 0 denotes original values. Rescaled values of output variables in new window can be obtained by adding correction to each old value in a window:

$$\Delta \mathbf{b} = \frac{1}{N} (y_{t+1}^0 - y_{t+1}^0) \quad (5)$$

And the newest sample is simply centered:

$$y_{t+1} = y_{t+1}^0 - \mathbf{b}_{t+1} \quad (6)$$

Each model update requires inverse  $N \times N$  regularized kernel matrix  $\tilde{\mathbf{K}}_{t+1}$  which is computationally and memory demanding. This inverse can be calculated by known previous inverse  $\tilde{\mathbf{K}}_t$ . Generally, if  $i$ -th sample should be removed from the window and new sample is added, inverse of new regularized kernel matrix can be efficiently calculated by following expressions:

$$\tilde{\mathbf{K}}_{temp}^{-1} = \left[ \tilde{\mathbf{K}}_t^{-1} \right]_{-i,-i} - \left( \left[ \tilde{\mathbf{K}}_t^{-1} \right]_{-i,j} \left[ \tilde{\mathbf{K}}_t^{-1} \right]_{-i,i}^T \right) / \left( \left[ \tilde{\mathbf{K}}_t^{-1} \right]_{i,i} \right)$$

$$\tilde{\mathbf{K}}_{t+1}^{-1} = \begin{bmatrix} \tilde{\mathbf{K}}_{temp}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\sigma_{t+1}^2} \begin{bmatrix} \alpha \alpha^T & -\alpha \\ -\alpha^T & 1 \end{bmatrix}$$
(7)

Where  $\sigma_{t+1}^2 = k_{t+1} + \sigma_n^2 - \mathbf{k}_{t+1}^T \tilde{\mathbf{K}}_{temp}^{-1} \mathbf{k}_{t+1}$ ,  $\alpha = \tilde{\mathbf{K}}_{temp}^{-1} \mathbf{k}_{t+1}$ .

## 2.2 Just in time learning

The key procedure in JITL modelling is the selection of relevant samples for modelling. The selected sample set from the database is neighbouring data around the query data. The neighbourhood is defined as any data having similarity with the query data. To evaluate this similarity, distance-based measure and some variants are commonly used due to their simplicity. Assume that database consisting  $N$  process data  $(y_i, \mathbf{x}_i)_{i=1 \dots N}$ ,  $y_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathbb{R}^n$  collected. Given a specific query data  $\mathbf{x}_q \in \mathbb{R}^n$ , JITL model is used to predict the model output  $\hat{y}_q = f(\mathbf{x}_q)$  according to the known database  $(y_i, \mathbf{x}_i)_{i=1 \dots N}$ .

To achieve improvements in the selection of relevant data, many research works have been done (Saporo, 2014). Generally, similarity factors used for JITL modelling can be categorized into three groups, distance-based similarity, distance and angle-based similarity and correlation-based similarity. In this work, we used Gaussian function to describe the similarity of different data for avoiding setting sample scale.

## 2.3 Process of the proposed soft sensor modeling

As above mentioned, the JITL-based soft sensor cannot guarantee the accuracy of predictions. It is natural to think of using a model to calibrate predictions. In this study, the basic idea is using a local temporal model on a local temporal window to calibrate predictions of JITL model. The local temporal model is built by the MWGPR algorithm. Detail modelling procedures of the proposed soft sensor are depicted as follows:

Step 1, Collect and standardize training data and test data.

Step 2, Search the relevant samples to construct a similar set from historical database with Gaussian function-based similarity criterion. Set value of the Gaussian function-based similarity criterion, which is within 0 and 1. In this work, the value is set to 0.9.

Step 3, Build a JITL model using the relevant samples.

Step 4, Predict the required output online for the query data and then discard the JITL model.

Step 5, Add the prediction of the last JITL model to the window and discard the oldest sample in the window.

Step 6, Update MWGPR on the new data in the window according to (2) - (7)

Step 7. Calculate the output responding to query data using the updated MWGPR model

## 3 Case study

Erythromycin is a macrolide antibiotic. In terms of structure, this macrocyclic compound contains a 14-membered lactone ring with ten asymmetric centres and two sugars (L-cladinose and D-desosamine), making it a compound very difficult to produce via synthetic methods. Erythromycin is produced from a strain of the actinomycete *Saccharopolyspora erythraea*. Erythromycin fermentation process is with high oxygen consumption. The simplified structure and parameters of reactor are shown in Figure 1.

For an Erythromycin fermentation process, biomass concentration plays a decisive role in the final product (Erythromycin) concentration. The curve of microbial growth is nonlinear (Figure 2). In autecological studies, microbial growth in batch culture can be modelled with four different phases: lag phase (A), log phase or exponential phase (B), stationary phase (C), and death phase (D). This basic batch culture growth model draws out and emphasizes aspects of microbial growth which may differ from the growth of any other creatures. In reality, even in batch culture, the four phases are not well defined. The cells do not reproduce in synchrony and their growth rate during exponential phase is often not a constant, instead of a slowly decaying rate. Therefore, it is difficult to accurately predict the growth of microorganisms, especially during the exponential phase.

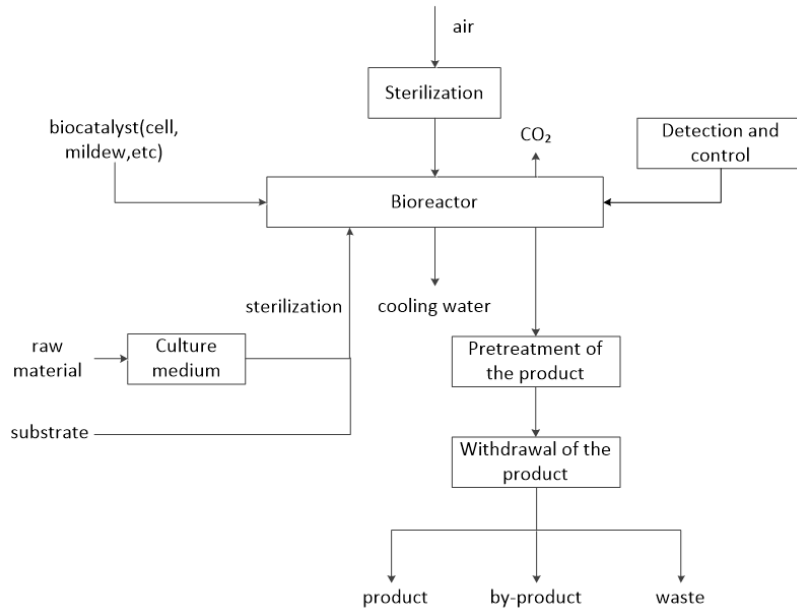


Figure 1: The flow chart of fermentation reaction

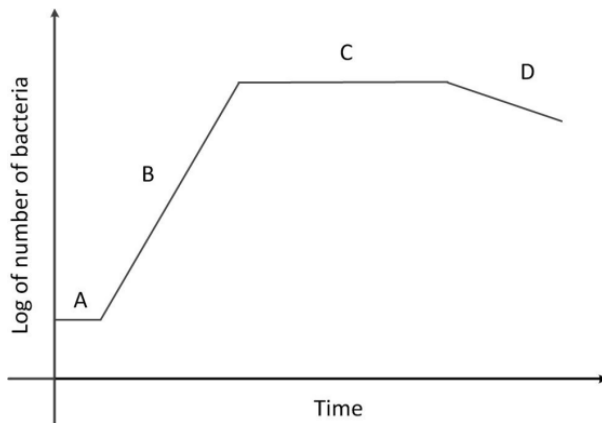


Figure 2: Microbial growth curve

Table 1: The meaning of each label

No.	Variables
1	Time
2	Dissolved oxygen tension
3	pH value
4	dextrin flow
5	soybean oil flow,
6	isopropanol flow
7	water flow
8	volume of dextrin
9	volume of soybean
10	volume of isopropanol
11	volume of water
12	temperature
13	air pressure
14	stirring speed
15	air flow

The main way of ensuring product quality of Erythromycin is to control biomass concentration which can be affected by many process factors. In this study, process data were collected from 10 independent fermentation processes. 182 samples were collected from each batch. Samples from 3 batches are selected as the query data, and remaining samples are used as the historical database. Every sample contains fifteen input variables and one output variable. After variable selection by a principal component analysis (PCA)-based method (Mei et al., 2016), five input variables, i.e. DO saturation, pH, Temperature, Agitator power, Aeration rate, are selected as secondary variables, and the output variable is biomass concentration.

Figure 3 gives predictions of the proposed soft sensor (JIT-MWGPR based soft sensor). For comparisons, the JIT-based model is also studied. From Figure 3, it can be observed that predictions by JIT-MWGPR models tracked more closely to actual data points than those of the JIT-GPR models. Note that actual curve is obtained by fitting collected samples for visual display.

The quantitative assessment of the JIT-MWGPR is also given by using the root-mean-square error (RMSE) criterion (see Table 2). From Table 2, for all three batches, it is obvious that JIT-MWGPR models have smaller RMSE values than JIT-GPR models. It means that JIT-MWGPR models have better predictive accuracy than traditional JIT-GPR models. The improvement is attributed to the calibration of local temporal models.

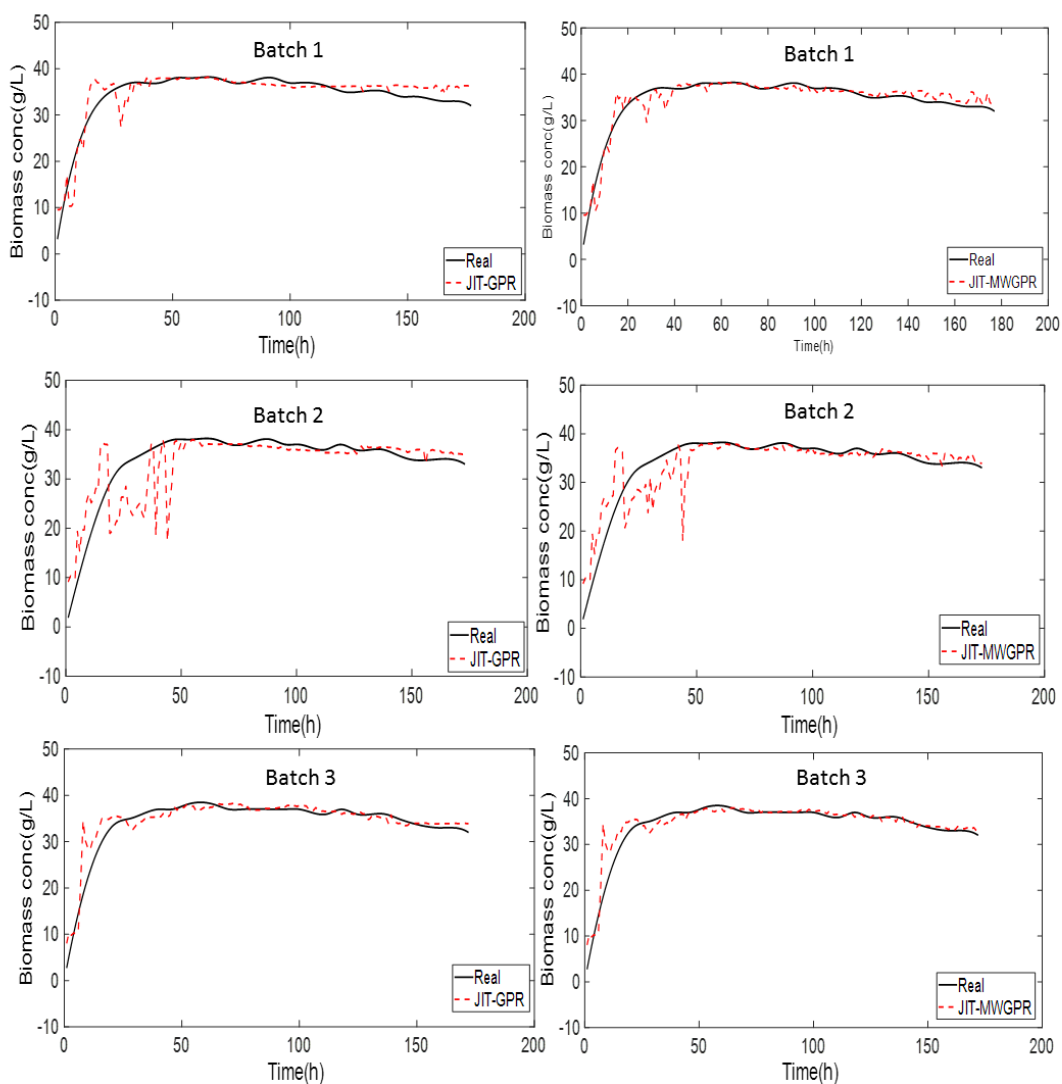


Figure 3: Comparisons of JIT-GPR and JIT-MWGPR based soft sensors in the Erythromycin fermentation process

Table 2: Comparisons of RMSE values of JIT and JIT-MWGPR-based soft sensors

Methods	Batch 1	Batch 2	Batch 3
JIT	2.1276	3.5296	1.9608
JIT-MWGPR	1.6540	3.1887	1.8058

#### 4. Conclusions

The growth rates of microorganisms in different phases of fermentation processes vary greatly, resulting in significant biomass concentration changes. Therefore, traditional soft sensors cannot be applied successfully in real world. It has been proved that JITL has ability of coping with those processes with strong nonlinearity. However, for a complex and dynamic process, similar inputs may result in total different outputs due to complex and nonlinear process dynamics. A new prediction-calibration strategy for soft sensor modeling is proposed in this work. In the new strategy, spatial characteristics of process data and temporal characteristics of the local fermentation process are both considered. To evaluate the proposed method, an industrial Erythromycin fermentation process was used. Results show that the proposed JIT-MW-based soft sensor performs better than traditional JIT-based soft sensor.

#### Acknowledgement

The authors gratefully acknowledge the financial support provided by Natural Science Foundation of Jiangsu Province of China (grant no. BK 20160540) and Natural Science Research Foundation of Higher Education of Jiangsu Province (grant no.17KJB510008).

#### References

- Chen, K., Castillo, I., Chiang, L.H., Yu, J., 2015, Soft sensor model maintenance: a case study in industrial processes. *IFAC-PapersOnLine*, 48, 427-432.
- Fujiwara, K., Kano, M., Hasebe, S., Takinami, A., 2009, Soft - sensor development using correlation-based just-in-time modeling. *AIChE Journal*, 55, 1754 - 1765.
- Grbic, R., Sliskovic, D., Kadlec, P., 2013, Adaptive soft sensor for online prediction based on moving window Gaussian process regression. In: *International Conference on Machine Learning and Applications*, 428-433.
- Kadlec, P., Grbić, R., Gabrys, B., 2011, Review of adaptation mechanisms for data-driven soft sensors. *Computers & Chemical Engineering*, 35, 1-24.
- Liu, Y., Gao, Z., Li, P., Wang, H., 2012, Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. 51, 4313-4327.
- Mei, C., Chen, Y., Jiang, H., Ding, Y., Chen, X., Liu, G., 2017a, Just-in-time modeling with a combination of input and output similarity criteria for soft sensor modeling in fermentation processes. *Chemical engineering transaction*, 61, 1045-1050.
- Mei, C., Chen, Y., Zhang, H., Chen, X., Liu, G., 2017b, Development of a multi-model strategy based soft sensor using gaussian process regression and principal component analysis in fermentation processes. *Chemical engineering transaction*, 61, 385-390.
- Mei, C., Ming, Y., Shu, D., Hui, J., Liu, G., Liao, Z., 2016, Soft sensor based on Gaussian process regression and its application in erythromycin fermentation process. *Chemical Industry & Chemical Engineering Quarterly*, 22(2), 127-135.
- Saporo, A., 2014, State of the art in the development of adaptive soft sensors based on just-in-time models. *Procedia Chemistry*, 9, 226-234.
- Souza, F. A.A., Rui, A., Mendes, J., 2016, Review of soft sensor methods for regression applications. *Chemometrics & Intelligent Laboratory Systems*, 152, 69-79.