

# Soft Sensor Modeling of Product Concentration in Glutamate Fermentation using Gaussian Process Regression

<sup>1,2</sup>Rongjian Zheng and <sup>1</sup>Feng Pan

<sup>1</sup>Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, China

<sup>2</sup>Faculty of Automation, Huaiyin Institute of Technology, Huaian, China

## Article history

Received: 01-06-2016

Revised: 30-08-2016

Accepted: 17-09-2016

Corresponding Author:

Rongjian Zheng

Key Laboratory of Advanced

Process Control for Light

Industry (Ministry of

Education), Jiangnan

University, Wuxi, China

Email: zhengrjian@163.com

**Abstract:** The on-line control of glutamate fermentation process is difficult, owing to the typical uncertainties of biochemical process and the lack of suitable on-line sensors for primary process variables. A prediction model based on Gaussian Process Regression (GPR) is presented to predict glutamate concentration online. First, Partial Least Squares (PLS) is applied to extract the features of the input secondary variables to reduce the number of the variables dimension and eliminate the correlation, through variables selection to reduce model complexity and improve model tracking performance. Validation was carried out in a 5 L fermentation tank for 10 batches glutamate fermentation process. Simulation results show that the proposed model outperforms the PLS and Support Vector Machine (SVM) model and the Root Mean Square Error (RMSE) are 1.59, 7.98 and 1.95, respectively. It can provide effective operation guidance for control and optimization of the glutamate fermentation process.

**Keywords:** Glutamate Fermentation, Gaussian Process Regression, Soft Sensor, Partial Least Squares, Input Variable Extraction, Modeling

## Introduction

The history of the species *Corynebacterium* as amino acid producer started in the 1950s when Dr Kinoshita was the first to discover that *Corynebacterium glutamicum* is a superior amino acid producer (Hermann, 2003). From then on, glutamate is one of the most important amino acids yielded mainly by fermentation using *Corynebacterium glutamicum*, which occupies about 53% of the world's amino acids market and its fermentative production amount exceeds 2.2 million tons per year (Khan *et al.*, 2005; Xiao *et al.*, 2006). Glutamate is widely used in food, pharmaceutical and other industries, especially used as a flavor enhancer with huge market requirement in oriental countries (China, Japan, Korea, etc.) (Cao *et al.*, 2013; Hasegawa *et al.*, 2008).

An accurate model is essential for further control and optimization of fermentation process. Generally, three type models are used for depicting the feature of fermentation process and for the purpose of process control and optimization: The unstructured dynamic model, the black box model and the fuzzy logic inference model (Zhang *et al.*, 2005). As we all know, fermentation process is a strong nonlinear, time-variant and correlating process (Wang *et al.*, 2010; Shi and Pan,

2010). Therefore, it is necessary to grasp the status information of fermentation process timely. Unfortunately, some crucially primary variables, such as biomass, substrate or product concentration which are more complex and important, can not be measured on-line yet in practical fermentation process. This seriously influences the execution of optimization strategy (Gu and Pan, 2015).

Soft sensor appeared as a reliable and helpful method to perform on-line observation of difficult to measure the primary variables in the past decades (Acuña *et al.*, 2014). It generally makes use of available process measurement data or prior knowledge on process mechanism to build predictive model for estimating the primary variables that cannot be easily measured by hardware based sensor in a real-time fashion (Kaneko and Funatsu, 2014; Chen *et al.*, 2014; Shokri *et al.*, 2015; Liu *et al.*, 2015). Many soft sensor methods have been presented, including Partial Least Squares (PLS), Artificial Neural Networks (ANN), Support Vector Machine (SVM) and Gaussian Process Regression (GPR) (Baffi *et al.*, 1999; Wang *et al.*, 2014; Acuña *et al.*, 2014; Chen *et al.*, 2014).

Although the conventional PLS based soft sensors are efficient in handling high-dimensional process variables

collinearity, they are essentially linear models and thus may not interpret significant process nonlinearity. Though ANN is a powerful tool to characterize nonlinear relationships in process, it is liable to converging towards local minima and cannot undertake global optima while accessing nonlinear multi-peak functions. Additionally, it often suffers from the over-fitting question so that the model generalization capability can be very poor. Meanwhile, different machine learning techniques such as Support Vector Machine (SVM) has gained some success in soft sensor modeling and quality estimation. Compared to the traditional ANN approach, SVM method has excellent performance in handling small experimental data with strong forecast capability. Moreover, it can acquire the global optimal solution and efficiently avoid the issue of local minima for nonlinear processes (Yu, 2012). More recently, Gaussian Process Regression (GPR) appeared as a probabilistic tool for nonlinear regression in the late 90 s (Rasmussen and Williams, 2006; Grbić *et al.*, 2013). Non-parametric probabilistic models such as a Bayesian framework for Gaussian Process Regression (GPR) have received remarkable attention in the field of machine learning. In comparison to ANN, Gaussian process is easier to understand and fulfill in practice (Pal and Deswal, 2010). GPR models are closely related to SVM as a result of the use of kernel functions. Compared with other kernel-based regression models, GPR not only provides the estimated value of a variable of interest, but also the variance of computation which can be interpreted as a level of confidence of the model. Though it has some attractive properties beyond ANN and SVM, only recently have a limited number of soft sensor applications of GPR been reported (Chen *et al.*, 2014; Liu *et al.*, 2015; Yu, 2012; Pal and Deswal, 2010; Grbić *et al.*, 2013; Ge *et al.*, 2011). Due to its virtue for nonlinear system modeling, in this study, we are trying for the first time to use GPR to construct a soft sensor modeling of product concentration in glutamate fermentation process.

However, it is well known that a soft-sensor's satisfactory estimation performance is likely to be achieved if only those secondary variables that are most sensitive to the primary variables are used. In fact, the inappropriate choice of estimator inputs may lead to numerical problems, such as singularity and over-parameterization (Zamprognia *et al.*, 2005), or the computational load increases remarkably. Therefore, suitable input variables have to be selected. Although the inputs should be selected using process knowledge, it is difficult to see the effect of each input on outputs when the number of variables becomes large. Further more, in fermentation process, its physical phenomena has not been clarified yet, to select inputs of a soft-sensor based on physical knowledge is especially hard (Fujiwara *et al.*, 2012).

This study aims to build GPR model based soft sensors for prediction production concentration in glutamate fermentation. For dimension compression of process data and also to deal with correlations between different input secondary variables, conventional PLS method can be performed, PLS is used to extract the characteristic components of the input secondary variables to reduce the number of the variables dimension and eliminate the redundant information, which means the GPR model will be constructed upon score variables of the PLS model. Results of predictions are compared with experiment data in 5 L fermenter for glutamate fermentation process as well as SVM to validate effectiveness of the proposed soft sensor method.

## Materials and Methods

### *Microorganism and Fermentation Conditions*

*Corynebacterium glutamicum* S9114, kept by the key laboratory of industrial biotechnology, Jiangnan University, was used throughout this study (Zhang *et al.*, 2005). The seed microorganism was grown in a shaker at 32°C and 200 rpm for 8-10 h in liquid medium containing (in g/L): Glucose 25, K<sub>2</sub>HPO<sub>4</sub> 1.5, MgSO<sub>4</sub> 0.6, MnSO<sub>4</sub> 0.005, FeSO<sub>4</sub> 0.005, corn slurry 25 and urea 2.5 (separated sterilization). Initial pH was adjusted to 7.0-7.2.

*C. glutamicum* S9114 was cultivated for glutamate production at about 32°C in a 5L fermenter (BIOTECH-5BG, Baoxing Co., China) containing about 3.4 L medium. Fermentation medium contained (g/L): Glucose 140, K<sub>2</sub>HPO<sub>4</sub> 1.0, MgSO<sub>4</sub> 0.6, MnSO<sub>4</sub> 0.002, FeSO<sub>4</sub> 0.002, thiamine 0.00005, corn slurry 15 and urea 3.0 (separated sterilization). Concentrated glucose was added to fermenter based on requirement to ensure the substrate concentration above a suitable level (15 g L<sup>-1</sup>) throughout the fermentation period. pH was controlled at 7.1±0.1 by automatic addition of 25% (w/w) ammonia water which also supplied the nitrogen source required for glutamate synthesis. Dissolved Oxygen (DO) was controlled at various levels by automatically or manually controlling the agitation speed according to particular necessities. The CO<sub>2</sub> and O<sub>2</sub> concentrations (partial pressure) in the inlet and exhaust gas were measured online by a gas analyzer (LKM2000A, Lokas Co. Ltd., Korea). Glutamate fermentation apparatus is shown in Fig. 1 (Dong, 2008).

Among the state variables in glutamate fermentation process, only physical variable and chemical variable, such as fermentation time (t), temperature (T), pH, fermentation tank pressure (P), fermentation volume (V), agitation rate (AG), ammonia water consumption rate (AR), CO<sub>2</sub> Evolution Rate (CER) and O<sub>2</sub> Uptake Rate (OUR), Dissolved Oxygen Concentration (DO), etc., can be measured on-line (Ding *et al.*, 2012).

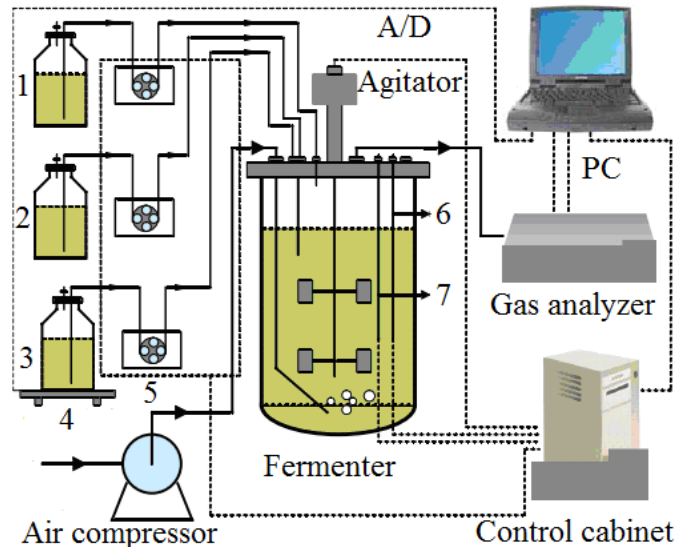


Fig. 1. Glutamate fermentation apparatus (1-Defoamer; 2-Glucose; 3-Ammonia water; 4-Electronic balance; 5-Peristaltic pump; 6-pH electrode; 7-DO electrode)

However the biomass concentration, substrate concentration and glutamate concentration cannot be measured on-line by hardware sensors. In general, the three variables can merely be assay off-line every two hours.

### Gaussian Process Regression

Suppose that we are given a training set  $D$  of  $n$  observations,  $D = \{(x_i, y_i) \mid i = 1, \dots, n\}$ , where  $x$  denotes an input vector of dimension  $D$  and  $y$  denotes a scalar output or target (dependent variable); the column vector inputs for design matrix all  $n$  cases are aggregated in the  $D \times n$  design matrix  $X$  and the targets are collected in the vector  $y$ , so we can write  $D = (X, y)$ . Consider the regression model with Gaussian noise:

$$y = f(x) + \varepsilon \quad (1)$$

where, white Gaussian noise  $\varepsilon \sim N(0, \sigma_n^2)$  with variance  $\sigma_n^2$  and the vector of regress  $x_i$  from operating space  $R^D$ . A Gaussian process is fully described by its mean function and covariance function. We define mean function  $m(x)$  and the covariance function  $k(x, x')$  of a real process  $f(x)$  as:

$$\begin{aligned} m(x) &= E[f(x)], \\ k(x, x') &= E[(f(x) - m(x))(f(x') - m(x'))], \end{aligned} \quad (2)$$

Then the Gaussian process can be written as:

$$y \sim GP(m(x), k(x, x') + \sigma_n^2 \delta_{ij}) \quad (3)$$

where,  $\delta_{ij}$  is a Kronecker delta which is one if  $i = j$  and zero otherwise. Assuming additive independent

identically distributed Gaussian noise  $\varepsilon$  with variance  $\sigma_n^2$ , the prior on the noisy observations becomes:

$$\text{cov}(y) = K(X, X) + \sigma_n^2 I \quad (4)$$

We can write the joint distribution of the observed target values and the function values at the test locations under the prior as:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N \left( 0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \quad (5)$$

We arrive at the key predictive equations for Gaussian process regression:

$$f_* | X, y, X_* \sim N(\bar{f}_*, \text{cov}(f_*)), \quad (6)$$

Where:

$$\bar{f}_* = E[f_* | X, y, X_*] = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y, \quad (7)$$

$$\begin{aligned} \text{cov}(f_*) &= K(X_*, X_*) \\ &- K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \end{aligned} \quad (8)$$

This equation can also be regarded as a linear combination of  $n$  kernel functions, each one centered on a training point, by writing:

$$\bar{f}_*(x_*) = \sum_{i=1}^n \alpha_i k(x_i, x_*) \quad (9)$$

Where:

$$\alpha = (K + \sigma_n^2 I)^{-1} y$$

The Laplace approximation to the marginal likelihood can be obtained in the same way as for the binary case, yielding:

$$\log(p(y|X, \theta)) = -\frac{1}{2} y^T K_y^{-1} y - \frac{1}{2} \lg |K_y| - \frac{n}{2} \lg 2\pi \quad (10)$$

where,  $K_y = K_f + \sigma_n^2 I$  is the covariance matrix for the noisy targets  $y$  (and  $K_f$  is the covariance matrix for the noise-free latent  $f$ ).

A common choice for the covariance function is the radial basis kernel:

$$C(x_i, x_j) = Cov[f(x_i), f(x_j)] \\ = \nu \exp \left[ -\frac{1}{2} \sum_{d=1}^D \omega_d (x_i^d - x_j^d)^2 \right] + \delta_{ij} \nu_0 \quad (11)$$

where,  $D$  is the dimension of the input space of vector  $x$  and  $\Theta = [\omega_1; \dots; \omega_D; \nu; \nu_0]^T$  is a vector of parameters called hyper-parameters (Ažman and Kocijan, 2007).

To set the hyper-parameters by maximizing the marginal likelihood, we seek gradient the partial derivatives of the marginal likelihood w.r.t. the hyper-parameters, we then obtain:

$$\frac{\partial}{\partial \theta_j} = \frac{1}{2} y^T K^{-1} \frac{\partial K}{\partial \theta_j} K^{-1} y - \frac{1}{2} tr \left( K^{-1} \frac{\partial K}{\partial \theta_j} \right) \quad (12)$$

### Soft Sensor Modeling Based on GPR

The prediction model of soft sensor based on GPR for glutamate concentration is shown in Fig. 2.

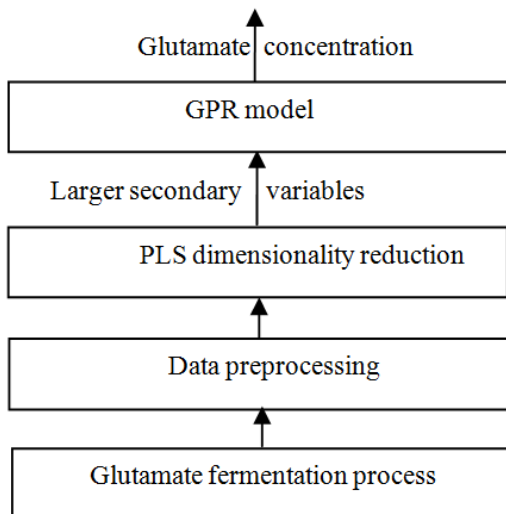


Fig. 2. Prediction model for glutamate concentration

Before soft sensing modeling, the secondary variables that are related with the primary variable are determined based on PLS and technologic analysis of fermentation process. In soft sensors, the secondary variables are used to act as the inputs of the soft sensing model and the primary variables such as glutamate concentration is employed to act as the output of soft sensor model.

### Preparation of Training Dataset

There are 10 batches experimental data of glutamate fermentation in 5L fermenter are used, fermentation period of 34 h, each batch includes 18 sample points, the batches data are divided randomly into two subsets, 9 batch used for training and the other one batch for testing. All the input variables and output variables are rescaled to the interval  $[-1, 1]$  by using the following equations:

$$x^* = \frac{2 \cdot (x - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (13)$$

$$y^* = \frac{2 \cdot (y - y_{\min})}{y_{\max} - y_{\min}} - 1 \quad (14)$$

where,  $(x, y)$  are the original values of input variables and output variables,  $(x^*, y^*)$  are normalized values of inputs and outputs.

### Select of Input Variables Based on PLS

During the process of building a production concentration model for glutamate fermentation, the secondary variables are chose as input variables of GPR by using PLS. PLS provides a bilinear decomposition of the secondary variables  $X$  and primary variable  $Y$  matrices into a number of rank-one matrices in a similar mode to that of Principal Component Analysis (PCA) for a single data matrix. The kernel of the PLS procedure is the Non-linear Iterative Partial Least-Squares (NIPALS) arithmetic (Baffi *et al.*, 1999; Park *et al.*, 2010). This arithmetic unceasingly extracts each pair of corresponding latent variables as a linear combination of the secondary and primary variables (Equation 15 and 16 respectively, prior to matching the inside linear regression (Equation 17) and then assessing the linear prediction of the output scores (Equation 18):

$$t_j = \sum_{m=1}^M x_m \omega_{mj} = X \omega_j \quad (15)$$

$$u_j = \sum_{k=1}^K y_k c_{kj} = Y c_j \quad (16)$$

$$b_j \frac{u_j^T t_j}{u_j^T u_j}, j = 1, 2, \dots, a \quad (17)$$

$$\hat{u}_j = t_j b_j, j = 1, 2, \dots, a \quad (18)$$

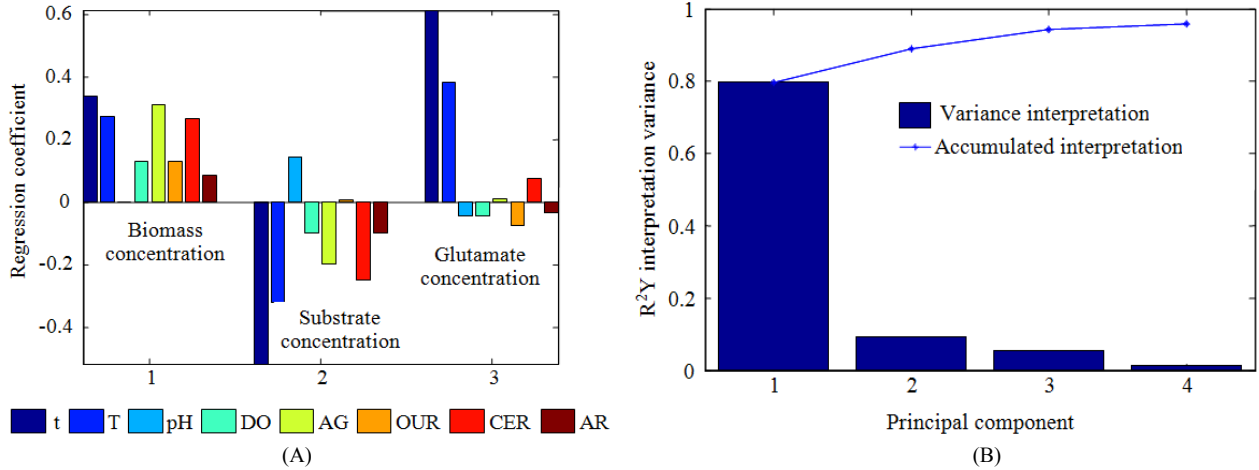


Fig. 3. The selection result of input variables by Partial Least Squares (PLS). (A) Regression coefficient chart by using Partial Least Squares (PLS), (B) Variance interpretation of principal component

The input and output loadings vectors are calculated from the least-squares regression of the  $X$  matrix on  $t_j$  and the  $Y$  matrix on  $\hat{u}_j$ :

$$p_j^T = \frac{t_j^T X}{t_j^T t_j}, j = 1, 2, \dots, a \quad (19)$$

$$q_j^T = \frac{\hat{u}_j^T Y}{\hat{u}_j^T \hat{u}_j}, j = 1, 2, \dots, a \quad (20)$$

The final step in the iterative procedure is to deflate the  $X$  and  $Y$  matrices:

$$E = X - t_j p_j^T, j = 1, 2, \dots, a \quad (21)$$

$$F = Y - \hat{u}_j q_j^T, j = 1, 2, \dots, a \quad (22)$$

A detailed specification of the algorithm can be found in Geladi and Kowalski (1986) and Höskuldsson (1988).

The selection result of input variables using PLS is shown in Fig. 3, from Fig. 3A, we can see that fermentation time  $t$ , fermentation temperature  $T$ ,  $OUR$  and  $CER$  are important secondary variables to interpret glutamate concentration, these input variables have a large influence on output variable. As shown in Fig. 3B, Variance interpretation of the top-scored four principal component reaches upwards of 95%.

According to glutamate fermentation technique, fermenter pressure ( $P$ ) and fermentation volume ( $V$ ) changes are little during the entire fermentation period, besides, there is coupling between agitation rate ( $AG$ ) and dissolved oxygen concentration ( $DO$ ),

therefore, the fermenter pressure ( $P$ ), fermentation volume ( $V$ ) and agitation rate ( $AG$ ) are neglected as secondary variables in soft sensor model (Ding *et al.*, 2012). Mapping relationship of the secondary variables to the primary variable, i.e.,  $y = f(t, T, OUR, CER)$ , is implemented by GPR prediction model.

#### Performance Index

In this study, Root Mean Square Error (RMSE) and coefficient of determination ( $R^2$ ) are used as index of accuracy and the predictive ability of soft sensor model and were defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (23)$$

$$R^2 = \frac{\left( n \sum_{i=1}^n \hat{y}_i y_i - \sum_{i=1}^n \hat{y}_i \sum_{i=1}^n y_i \right)^2}{\left( n \sum_{i=1}^n \hat{y}_i^2 - \left( \sum_{i=1}^n \hat{y}_i \right)^2 \right) \left( n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right)} \quad (24)$$

where,  $y_i$  is the off-line measured value;  $\hat{y}_i$  is the prediction value and  $n$  is the number of samples.

## Results and Discussion

The simulations are implemented in MATLAB running on a laptop with 2.2GHz processor and 8GB RAM.

In order to evaluate the performance of the soft sensor modeling based on GPR, the conventional PLS and SVM models are equally constructed to predict production concentration with the same fermentation data. The input secondary variables number of PLS

model is selected as 8; the Gaussian kernel is used in SVM, the hyper-parameters of SVM are tuned by grid search,  $c = 45.255$ ,  $\gamma = 0.1768$ ,  $\varepsilon = 0.01$ , whereas hyper-parameters of GPR are  $\log\Theta = [2.5187, 6.2309, 4.7544, 5.9008, 3.4810, 1.0961]^T$ , both of which can get the best prediction performance.

Soft sensors based on PLS, PLS-SVM and PLS-GPR are constructed through training data. Prediction residuals and predicted values by PLS-GPR for glutamate concentration are shown in Fig. 4-5, it can be seen that most of the residuals range is between -2~2, the maximum residuals reaches 3.775 at 28 h, this is because it is production phase at the moment. Predicted values of the soft sensor based on PLS and PLS-SVM for glutamate concentration are shown in Fig. 6-7, respectively. Prediction error comparison of three different methods are shown in Figure 8, to make this clearer, the error bar for each real values are shown in Fig. 9. From Fig. 5-9, we can see that prediction by PLS-GPR exhibits better tracking

performance than that by PLS and PLS-SVM, the testing samples by PLS-GPR have been well fitted, The prediction results of all soft sensors are shown in Table 1, as can be seen that GPR has got the greatest prediction performance, since it has the smallest Root Mean Square Error (RMSE) 1.32 among all soft sensors, the RMSE of PLS-GPR model increases to 1.59 compared to the GPR without inputs selection by PLS, however, the number of input secondary variables is down from eight to four by 50%,  $R^2$  is down from 0.997 to 0.996, it also has better tracking performance. Meanwhile the prediction performance of PLS-SVM model is clearly relatively superior to SVM model, such as RMSE is down from 3.65 to 1.95 by about 46.58% and  $R^2$  is up from 0.989 to 0.996. Table 1 also shows the same R squared values for PLS-SVM and PLS-GPR models, since PLS-GPR model has smaller Root Mean Square Error (RMSE) 1.59, but PLS-SVM model is 1.95, PLS-GPR is superior to PLS-SVM model.

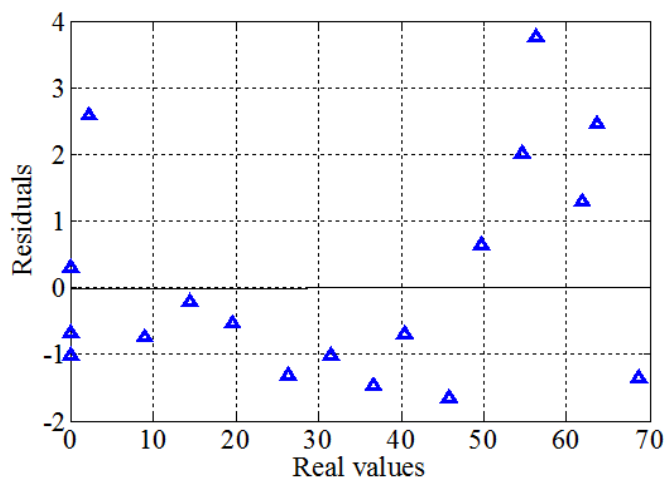


Fig. 4. Prediction residuals by PLS-GPR for glutamate concentration

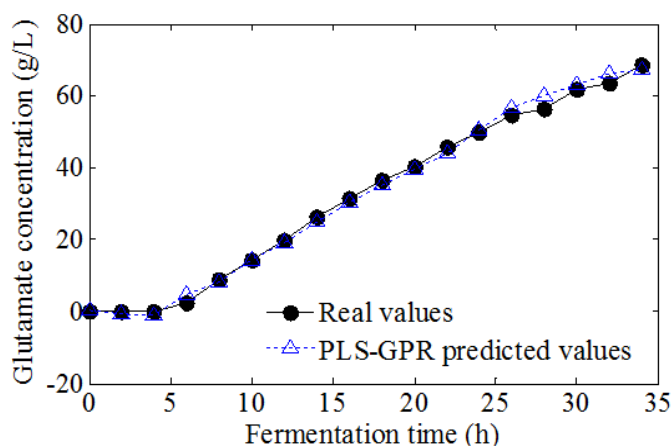


Fig. 5. Prediction results by PLS-GPR

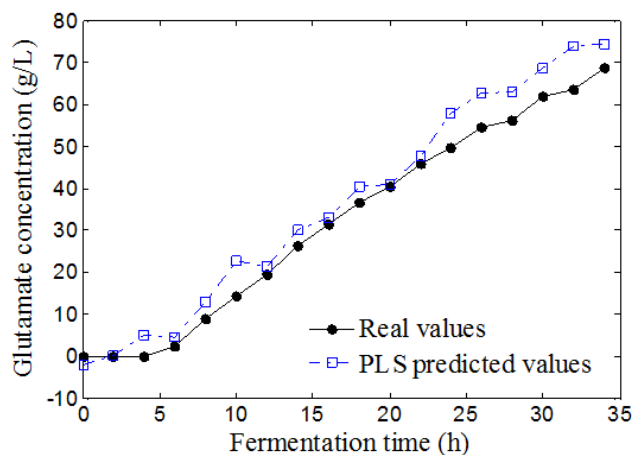


Fig. 6. Prediction results by Partial Least Squares (PLS)

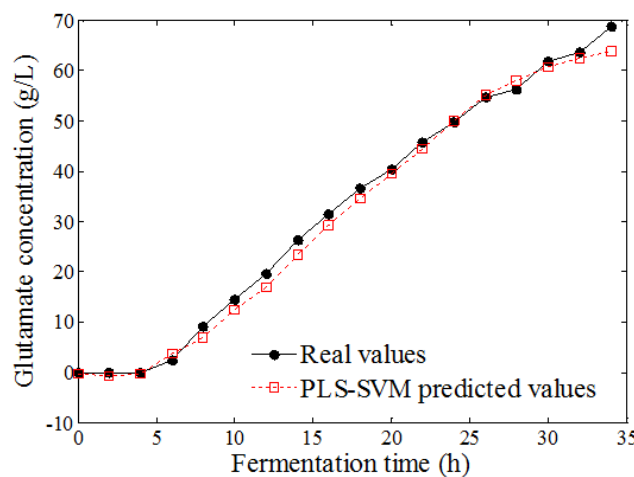


Fig. 7. Prediction results by PLS-SVM for glutamate concentration

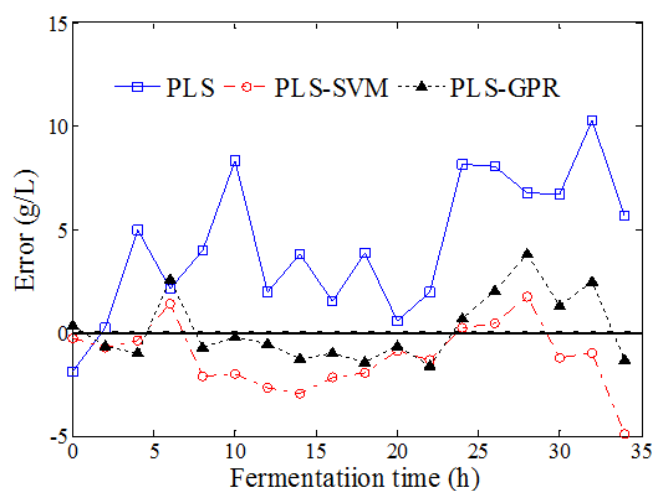


Fig. 8. Prediction error comparison of three different methods



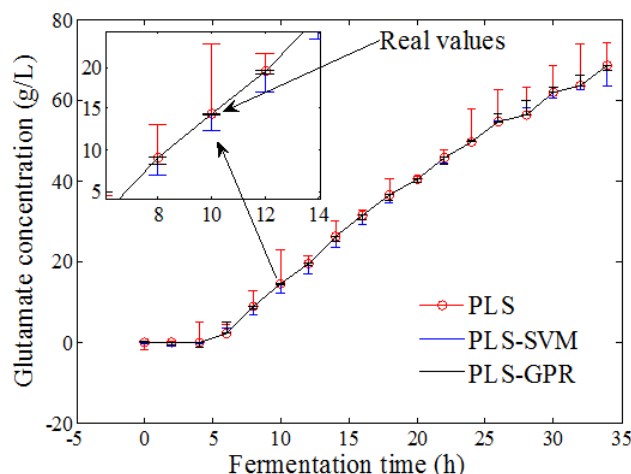


Fig. 9. Prediction errorbar comparison of three different methods

Table 1. The prediction results of all soft sensors

Models	RMSE	$R^2$
PLS	7.98	0.912
SVM	3.65	0.989
GPR	1.32	0.997
PLS-SVM	1.95	0.996
PLS-GPR	1.59	0.996

## Conclusion

Glutamate fermentation process is a kind of complicated batch process which is severely nonlinear and time-varying. In this study, a soft sensor prediction model based on Gaussian Process Regression (GPR) for glutamate fermentation was proposed to predict production concentration.

The prediction performance of the soft sensor is mainly impacted by the input secondary variables of GPR model. First, the correlation analyses of the input variables are carried out by PLS to reduce model complexity and dimensionality of input variables.

The performance comparison among GPR, PLS and SVM method is implemented. Simulation results show that the GPR based soft sensor performs better than other traditional soft sensors, PLS-GPR exhibits better tracking performance, it can provide effective operation guidance for control and optimization of glutamate fermentation process.

## Funding Information

This work was supported by the National Natural Science Foundation of China under Grant No.61273131, the Graduate Innovation of Jiangsu Province (CXZZ12\_0741) and the Fundamental Research Funds for Central Universities (JUDCF12034).

## Author's Contributions

**Rongjian Zheng:** Conceived of the research, designed the research plan, drafted and revised the manuscript.

**Feng Pan:** Involved in acquisition and analysis of data, provided technical guidance and funding support.

## Ethics

The authors declare that they have no conflict of interest. All authors have read and approved the manuscript and no ethical issues involved.

## References

- Acuña, G., C. Ramirez and M. Curilem, 2014. Software sensors for biomass concentration in a SSC process using artificial neural networks and support vector machine. *Bioprocess Biosyst. Eng.*, 37: 27-36.
- Ažman, K. and J. Kocijan, 2007. Application of Gaussian processes for black-box modelling of biosystems. *ISA Trans.*, 46: 443-457. DOI: 10.1016/j.isatra.2007.04.001
- Baffi, G., E.B. Martin and A.J. Morris, 1999. Non-linear projection to latent structures revisited: The quadratic PLS algorithm. *Comput. Chem. Eng.*, 23: 395-411. DOI: 10.1016/S0098-1354(98)00283-X
- Cao, Y., E. Mpofu and Z.P. Shi, 2013. A novel metabolic model incorporating directed signal flow diagram with enzymatic activities data for evaluating glutamate yield in glutamate fermentation. *Biochem. Eng. J.*, 77: 136-146. DOI: 10.1016/j.bej.2013.05.015
- Chen, J.Y., J. Yu and Y.L. Zhang, 2014. Multivariate video analysis and Gaussian process regression model based soft sensor for online estimation and prediction of nickel pellet size distributions. *Comput. Chem. Eng.*, 64: 13-23. DOI: 10.1016/j.compchemeng.2014.01.010



- Ding, J., Y. Cao, E. Mpofu and Z.P. Shi, 2012. A hybrid support vector machine and fuzzy reasoning based fault diagnosis and rescue system for stable glutamate fermentation. *Chem. Eng. Res. Design*, 90: 1197-1207. DOI: 10.1016/j.cherd.2012.01.004
- Dong, C.L., 2008. Study on fault diagnosis for glutamic acid fermentation process. MSc Thesis, Jiangnan University, Wuxi, Jiangsu, China.
- Fujiwara, K., H. Sawada and M. Kano, 2012. Input variable selection for PLS modeling using nearest correlation spectral clustering. *Chemometr. Intell. Laboratory Syst.*, 118: 109-119. DOI: 10.1016/j.chemolab.2012.08.007
- Ge, Z.Q., T. Chen and Z.H. Song, 2011. Quality prediction for polypropylene production process based on CLGPR model. *Control Eng. Practice*, 19: 423-432. DOI: 10.1016/j.conengprac.2011.01.002
- Geladi, P. and B.R. Kowalski, 1986. Partial least squares regression: A tutorial. *Anal. Chim. Acta*, 185: 1-17. DOI: 10.1016/0003-2670(86)80028-9
- Grbić, R., D. Sliškoivić and P. Kadlec, 2013. Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. *Comput. Chem. Eng.*, 58: 84-97. DOI: 10.1016/j.compchemeng.2013.06.014
- Gu, B.J. and F. Pan, 2015. A soft sensor modelling of biomass concentration during fermentation using accurate incremental online v-support vector regression learning algorithm. *Am. J. Biochem. Biotechnol.*, 11: 149-159. DOI: 10.3844/ajbbsp.2015.149.159
- Hasegawa, T., K. Hashimoto, H. Kawasaki and T. Nakamatsu, 2008. Changes in enzyme activities at the pyruvate node in glutamate-overproducing *Corynebacterium glutamicum*. *J. Biosci. Bioeng.*, 15: 12-19. DOI: 10.1263/jbb.105.12
- Hermann, T., 2003. Industrial production of amino acids by coryneform bacteria. *J. Biotechnol.*, 104: 155-172. DOI: 10.1016/s0168-1656(03)00149-4
- Höskuldsson, A., 1988. PLS regression methods. *J. Chemometr.*, 2: 211-228. DOI: 10.1002/cem.1180020306
- Kaneko, H. and K. Funatsu, 2014. Database monitoring index for adaptive soft sensors and the application to industrial process. *AIChE J.*, 60: 160-169. DOI: 10.1002/aic.14260
- Khan, N.S., I.M. Mishra, R.P. Singh and B. Prasad, 2005. Modeling the growth of *Corynebacterium glutamicum* under product inhibition in L-glutamic acid fermentation. *Biochem. Eng. J.*, 25: 173-178. DOI: 10.1016/j.bej.2005.01.025
- Liu, Y., T. Chen and J.H. Chen, 2015. Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions. *Ind. Eng. Chem. Res.*, 54: 5037-5047. DOI: 10.1021/ie504185j
- Pal, M. and S. Deswal, 2010. Modelling pile capacity using Gaussian process regression. *Comput. Geotechn.*, 37: 942-947. DOI: 10.1016/j.compgeo.2010.07.012
- Park, T.C., T.Y. Kim and Y.K. Yeo, 2010. Prediction of the melt flow index using partial least squares and support vector regression in High-Density Polyethylene (HDPE) process. *Korean J. Chem. Eng.*, 27: 1662-1668. DOI: 10.1007/s11814-010-0280-x
- Rasmussen, C.E. and C.K.I. Williams, 2006. *Gaussian Processes for Machine Learning*. 1st Edn., MIT Press, Cambridge, ISBN-10: 026218253X, pp: 248.
- Shi, Z.P. and F. Pan, 2010. Analysis, control and test technology of fermentation process. Chemical Industry Press, Beijing.
- Shokri, S., M.T. Sadeghi, M.A. Marvast and S. Narasimhan, 2015. Improvement of the prediction performance of a soft sensor model based on support vector regression for production of ultra-low sulfur diesel. *Pet. Sci.*, 12: 177-188. DOI: 10.1007/s12182-014-0010-9
- Wang, T., J.W. Sun, W.D. Zhang and J.Q. Yuan, 2014. Prediction of product formation in 2-keto-l-gulononic acid fermentation through Bayesian combination of multiple neural networks. *Process Biochem.*, 49: 188-194. DOI: 10.1016/j.procbio.2013.11.003
- Wang, X.F., J.D. Chen, C.B. Liu and F. Pan, 2010. Hybrid modeling of penicillin fermentation process based on least square support vector machine. *Chem. Eng. Res. Des.*, 88: 415-420. DOI: 10.1016/j.cherd.2009.08.010
- Xiao, J., Z.P. Shi, P. Gao, H.J. Feng and Z.Y. Duan *et al.*, 2006. On-line optimization of glutamate production based on balanced metabolic control by RQ. *Bioprocess Biosyst. Eng.*, 29: 109-117. DOI: 10.1007/s00449-006-0059-y
- Yu, J., 2012. Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach. *Chem. Eng. Sci.*, 82: 22-30. DOI: 10.1016/j.ces.2012.07.018
- Zamprogna, E., M. Barolo and D.E. Seborg, 2005. Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis. *J. Process Control*, 15: 39-52. DOI: 10.1016/j.jprocont.2004.04.006
- Zhang, C.Y., Z.P. Shi, P. Gao, Z.Y. Duan and Z.G. Mao, 2005. On-line prediction of products concentrations in glutamate fermentation using metabolic network model and linear programming. *Biochem. Eng. J.*, 25: 99-108. DOI: 10.1006/j.bej.2005.03.012