CI&CEQ
Chemical Industry & Chemical Engineering Quarterly

Available on line at
Association of the Chemical Engineers of Serbia AChE
www.ache.org.rs/CICEQ

Chem. Ind. Chem. Eng. Q. 22 (2) 127–135 (2016) CI&CEQ

CONGLI MEI
MING YANG
DONGXIN SHU
HUI JIANG
GUOHAI LIU
ZHILING LIAO

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China

SCIENTIFIC PAPER

UDC 004:681.586:615.33:663.14

DOI 10.2298/CICEQ150125026M

SOFT SENSOR BASED ON GAUSSIAN PROCESS REGRESSION AND ITS APPLICATION IN ERYTHROMYCIN FERMENTATION PROCESS

Article Highlights

- A systematic soft sensor modeling method based on GPR and PCA is proposed
- The variance of the predicted output was designed on the output uncertainty of the GPR model
- Practical applications show the superiority of the proposed soft sensor method

Abstract

Erythromycin fermentation is a typical microbial fermentation process. Soft sensors can be used to estimate the biomass of Erythromycin fermentation process due to their relative low cost, simple development, and ability to predict difficult-to-measure variables. However, traditional soft sensors, e.g. artificial neural network (ANN) soft sensors, support vector machine (SVM) soft sensors, etc., cannot represent the uncertainty (measurement precision) of outputs, which results in difficulties in practice. Gaussian process regression (GPR) provides a novel framework to solve regression problems. The output uncertainty of a GPR model follows Gaussian distribution, expressed in terms of mean and variance. The mean represents the predicted output. The variance can be viewed as the measure of confidence in the predicted output that distinguishes the GPR from NN and SVM soft sensor models. We propose a systematic approach based on GPR and principal component analysis (PCA) to establish a soft sensor to estimate biomass of erythromycin fermentation process. Simulations on industrial data from an erythromycin fermentation process show the proposed GPR soft sensor has high performance of modeling the uncertainty of estimates.

Keywords: fermentation process; soft sensor; uncertainly; Gaussian process regression; principle component analysis.

Online control and optimization methods of fermentation processes depend on reliable and accurate measurements of key variables. However, not all important variables are available in real-time to allow timely action. Lack of key measurements can be contributed to two factors: 1) insufficient automation of some sensitive analysis; 2) even if real-time measurement is possible, the cost of installing an additional sensor may not be economically attractive. Advanced control algorithms cannot always be used in microbial fermentation processes due to the lack of measurements of important biological parameters. Thus, people usually rely on past experience to operate microbial fermentation process. This situation increases costs of production and operation in the fermentation industry. In order to achieve effective monitoring and control of the fermentation process, on-line measurement problems of the key parameters, such as biomass concentration, etc., should be resolved [1].

Recently, the soft sensor technique has seen wide applications in bioprocesses. Compared to traditional measurement methods, it is an innovative detection method developed to estimate unmeasurable variables or difficult-to-measure variables through online and measurable secondary variables. Some typical methods have been recognized to have strong

Correspondence: C. Mei, School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, P.R. China.
E-mail: clmei@ujs.edu.cn
Paper received: 25 January, 2015
Paper revised and accepted: 10 July, 2015

potential in bioprocesses, such as mechanism models [2], adaptive observers [3], PLS (partial least squares) [4], filtering techniques (*e.g.*, Kalman filter, extended Kalman filter, etc.) [5,6]. In addition, data-driven methods have been also widely investigated in the field of microbial fermentation, *e.g.*, SVM (support vector machine) [7–10] and ANN (artificial neural network) [11–13]. The root mean square error (*RMSE*) is usually used as a measure to evaluate soft sensors [10]. According to the measurement principle, measurements are meaningless without precision information (uncertainty). For measuring instruments, the uncertainty of measurements is easy to be estimated by the precision of instruments. However, for the above-mentioned soft sensors, the uncertainty of outputs is difficult to determine. This is an obvious drawback of existing soft sensor methods.

Gaussian process regression (GPR) is nonparametric and flexible, meaning that the complexity of the model grows as more data points are received and predictions can be obtained without giving the unknown function $y(x)$ an explicit parameterization. As a Bayesian probabilistic model, GPR is able to model the uncertainty inherent in noisy data, and give full probabilistic predictions or error bars. They naturally grow in regions away from training data where there is high uncertainty about the prediction model. The Gaussian process was first used by O'Hagan [14] as an alternative approach to the artificial neural network approach. Williams and Rasmussen [15] first described GPR in a machine learning context. With new knowledge in machine learning, Gaussian process has recently seen explosive growth. However, only a limited number of soft sensing applications based on GPR have been reported [16–19].

Modern industry processes are always loaded with many sensors. It is well known that it is not generally possible to use all available sensor variables as soft sensor inputs, because measurement redundancy generally makes the calibration of the regression model troublesome. The satisfactory performance of soft sensors is likely to be achieved if only those sensor variables, named secondary variables, that are most sensitive to the primary variables are employed. Several methods were proposed to select best variables. Simple statistical analyses have been conducted for process variables to identify a subset of measurements for use in the soft sensors [20]. Ma *et al.* used a conventional stepwise variable selection method to develop a soft sensor [21]. Alternative Methods on principal component analysis (PCA) [22] and singular value decomposition (SVD) [23] were also investigated. For the effectiveness and simplicity, the PCA based method was widely used in the field of soft sensor modeling to select secondary variables [22,24].

To the best of authors' knowledge, little information in literature has been devoted to soft sensor development based on GPR combined PCA. In this paper, a systematic soft sensor modeling approach using GPR and PCA for the Erythromycin fermentation process was presented. The focus of this method was modeling the uncertainty using GPR in bioprocesses. To simplify the complexity of the GPR based soft sensor, the PCA-based variable selection method was used to select input variables of the soft sensor. Finally, the proposed soft sensor was evaluated on industrial data from an actual Erythromycin fermentation process.

MATERIAL AND METHODS

Erythromycin fermentation process

Erythromycin is a macrolide antibiotic. In terms of structure, this macrocyclic compound contains a 14-membered lactone ring with ten asymmetric centers 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 [25]. The flow chart of fermentation reaction is shown in Figure 1. Also, the simplified structure and parameters of reactor are shown in Figure 2.

Generally, the process parameters of microbial fermentation processes include physical parameters, chemical parameters and biological parameters. Physical parameters typically include temperature (T) and pressure (P) of the fermenter, air flow (F_A), flow of cooling water (F_W), water temperatures of the inlet and the outlet (T_1, T_2), rotate speed of the stirring motor (*RMP*), etc. Chemical parameters include pH, dissolved oxygen content (*DO*), etc. The physical and chemical parameters can be automatically measured by instruments. However, biological parameters, such as cell concentration, metabolite concentration, substrate concentration and growing rate of the cell, etc., cannot be measured by sophisticated real-time measuring instruments. Therefore, soft sensors were proposed to estimate biological parameters in bioprocesses.

For an erythromycin fermentation process, biomass concentration plays a decisive role in the final product (erythromycin) concentration. Thus, the primary way of ensuring product quality of Erythromycin is to control biomass concentration, which can be affected by many process factors. In this process,

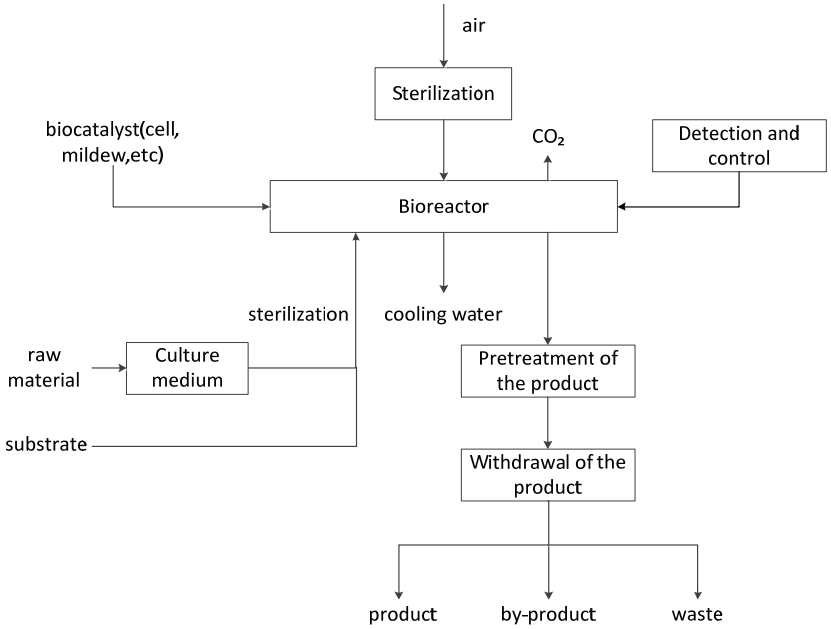

Figure 1. The flow chart of fermentation reaction.

secondary variables can be determined by statistical methods on process variables including time, dissolved oxygen, pH, dextrin flow, soybean oil flow, isopropanol flow, water flow, volume of dextrin, volume of soybean, volume of isopropanol, volume of water, temperature, air pressure, stirring speed, air flow, etc.

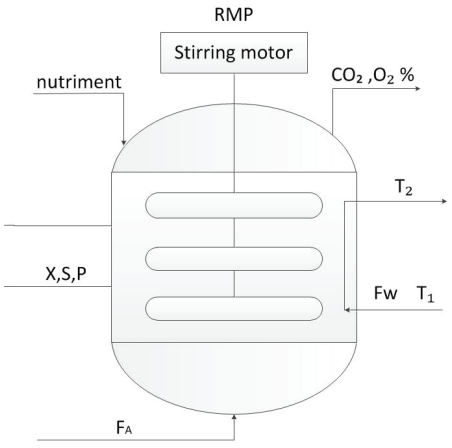

Figure 2. The diagram of erythromycin fermentation process.

The microbial growth curve is nonlinear and can be divided into four different phases: A) lag phase, B) log phase or exponential phase, C) stationary phase, and D) death phase. 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 fact, even in batch culture, the four phases are not always well defined. Cells do not reproduce in synchrony and their growth rate during exponential phase is often not a constant, instead of a slowly decaying rate [26], so it is difficult to accurately predict the growth of microorganisms, especially during the exponential phase.

Data acquisition

During this research, training and test data sets were collected from Zhenjiang Medicine Co., Ltd, China. The whole Erythromycin fermentation process lasted about seven days. In this work, 180 group points have been collected by physical sensors during each batch fermentation process, and the biomass concentration was obtained by off-line laboratory analysis every hour synchronously. Nine batches ran under similar environment and initial conditions. The sample data set was divided into two parts: training set and test set. The batches 1-6 were used as training data sets, and another batches were used as test data sets. Fourteen process variables were measured in the process besides the time variable. The curves of all process variables are shown in Figure 3. The legend numbers 1-9 denote the 1st to the 9th batch fermentation process. The subfigures a-n represent different process variables.

Gaussian Process Regression (GPR)

In this section, the basic GPR method [27,28] is introduced.

Suppose a training set:

$$D = \{(\mathbf{x}_i, y_i)\}, \ (i = 1,\ldots,l) \tag{1}$$

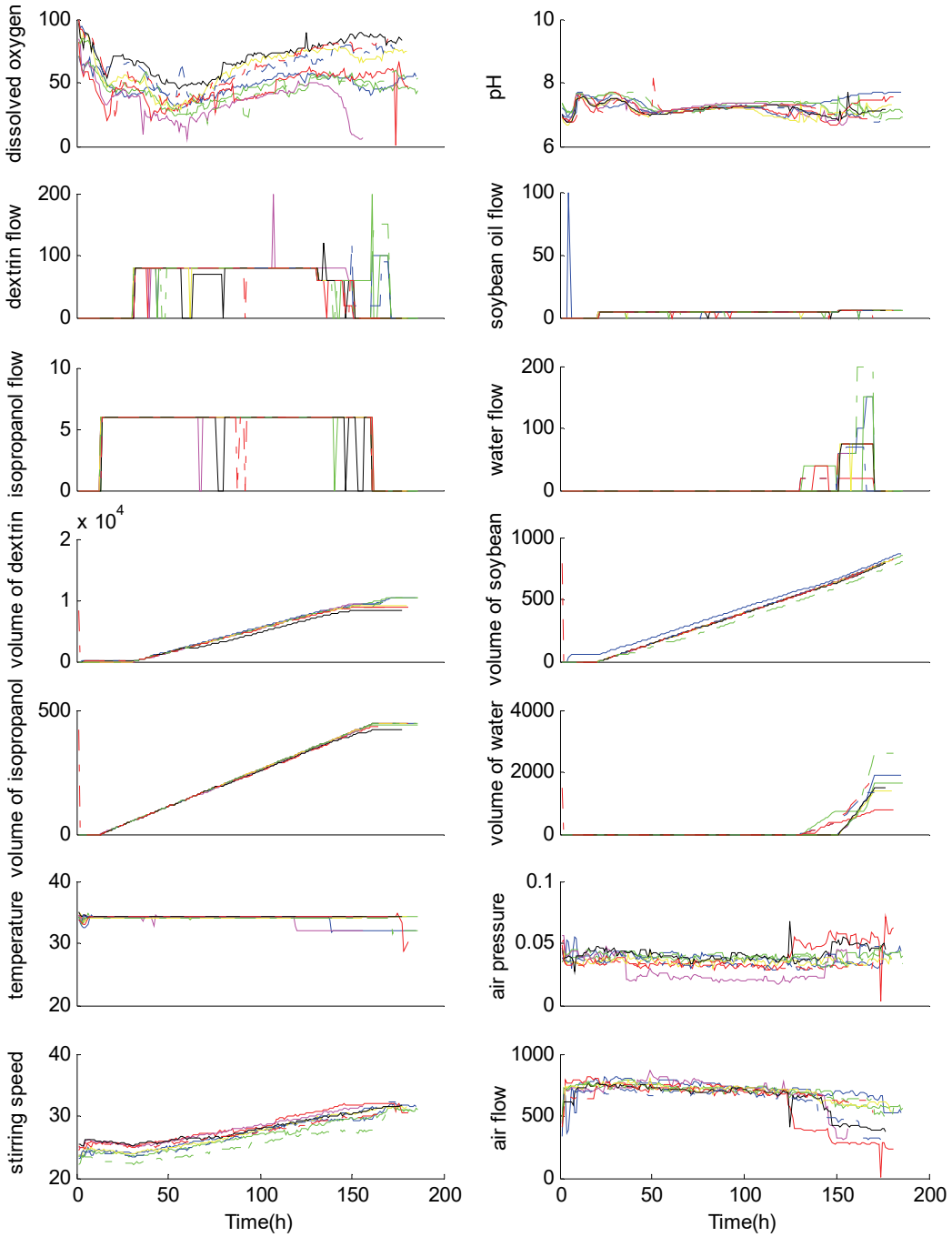

Figure 3. Variation curves of process variables.

where $\mathbf{x}_i \in \mathbf{R}^n$ denotes the n-dimensional input (observation) vector, $y_i \in \mathbf{R}$ denotes a scalar output or target, and I is the number of training samples. The task of a common regression is to estimate the conditional expectation of the dependent variable given the independent variables, *i.e.*, learning the mapping relationship between input $\mathbf{x}$ and output y in terms of training set, and predict the most likely value of the test data $\mathbf{x}_*$. A Gaussian process is commonly specified by its mean and covariance function. To make calculation easy, we generally set a zero mean GP model on the function variables as:

$$f(\mathbf{x}) \sim N(0, k(\mathbf{x}, \mathbf{x}_*)) \tag{2}$$

For the regression problems, considering the following model:

$$y = f(\mathbf{x}) + \varepsilon\ ,\ \varepsilon \sim N(0, \sigma_n^2) \tag{3}$$

where ε is the noise, and σ_n^2 is the variance of the noise. In this paper, considering fermentation is a highly complex process and is susceptible to interference, we should add the noise to the prior (*i.e.*, observed value), such that:

$$\mathbf{y} \sim N(0, K(X,X) + \sigma_n^2 I) \tag{4}$$

where $K(X,X) = \left(k(\mathbf{x}_i, \mathbf{x}_j)\right)$ is a symmetric and non-negative definite covariance matrix, I is the $n \times n$ dimensional identity matrix. Then we can write the joint distribution of the observed target values and the function values at the test point under the prior as:

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim N\left(0, \begin{pmatrix} K(X,X) + \sigma_n^2 I & K(X,\mathbf{x}_*) \\ K(\mathbf{x}_*, X) & k(\mathbf{x}_*, \mathbf{x}_*) \end{pmatrix}\right) \tag{5}$$

where f_* is the GP posterior (predicted value, here we consider only 1-dimensional output), $K(X,\mathbf{x}_*)$ is the $n \times 1$ covariance matrix between training and test cases, $K(\mathbf{x}_*, X)$ is the transposition of $K(X,\mathbf{x}_*)$, $k(\mathbf{x}_*, \mathbf{x}_*)$ is the covariance function evaluated at $\mathbf{x}_*$ and itself. Then, we can get the predictive distribution of GPR as:

$$f_* \mid X, \mathbf{y}, \mathbf{x}_* \sim N\left(\overline{f}_*, \mathrm{cov}(f_*)\right) \tag{6}$$

where:

$$\overline{f}_* = K(\mathbf{x}_*, X)[K(X,X) + \sigma_n^2 I]^{-1}\mathbf{y} \tag{7}$$

$$\mathrm{cov}(f_*) = k(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, X)[K(X,X) + \sigma_n^2 I]^{-1} K(X, \mathbf{x}_*) \tag{8}$$

Obviously, the predicted mean and variance of the test point $\mathbf{x}_*$ can be respectively written as Eq. (7) and Eq. (8).

The covariance function is the crucial ingredient in a Gaussian process predictor, as it encodes our assumptions about the function which we want to get [27]. To be a valid covariance function, it must be positive semidefinite. There are some covariance functions that can be used. A suitable covariance function is:

$$k(\mathbf{x}_i, \mathbf{x}_j) = v_0 \exp\left[-\frac{1}{2}\sum_{t=1}^{n} w_t (x_i^t - x_j^t)^2\right] + v_1 \delta_{ij} \tag{9}$$

where $\Theta = [w_1, \ldots, w_n, v_0, v_1]^T$ are the hyperparameters of the covariance functions. v_0 represents a overall measure of the prior knowledge, or put simply, it controls the typical amplitude of covariation. v_1 represents the variation of noise and it follows Gaussian distribution. δ_{ij} is Kronecker delta. w_t can be seen as the weight of the t-th dimension, *i.e.*, how many information contained in the corresponding input dimension t. The first term of Eq. (9) expresses the belief that inputs that are close to each other give rise to outputs that are close to each other or that are highly correlated. The second term of Eq. (9) involving the hyperparameter $v_1 = \sigma^2$, *i.e.*, the variance of the noise model for the outputs and therefore only occurs when $i = j$.

The test value can be predicted if the covariance function had been given. However, the property of the GP model cannot be guaranteed if the parameters of covariance function have not been optimized. The GPR is in the framework of Bayesian rule [27,28]. Thus, the marginal likelihood function of Bayesian formalism, typically negative log marginal likelihood function, was chosen as an optimization objective function of hyperparameters, which can be written as:

$$L(\Theta) = -\frac{1}{2}\lg(|C|) - \frac{1}{2}\mathbf{y}^T C^{-1}\mathbf{y} - \frac{I}{2}\log(2\pi) \tag{10}$$

Then the optimized hyperparameters can be obtained by solving the partial derivatives of the marginal likelihood function as:

$$\frac{\partial L(\Theta)}{\partial \Theta_i} = \frac{1}{2}\mathrm{tr}\left((\boldsymbol{\alpha}\boldsymbol{\alpha}^T - C^{-1})\frac{\partial C}{\partial \Theta_i}\right) \tag{11}$$

where $\boldsymbol{\alpha} = C^{-1}\mathbf{y}$. Then, by making the partial derivatives be equal to zero, the optimal solution of hyperparameters can be obtained.

RESULTS AND DISCUSSION

Results of data processing

The correspondence between labels and process variables is shown in Table 1.

Table 1. Variable labels

Label	Meaning
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

To avoid attributes in greater numeric dominating those in smaller numeric ranges, the collected data was normalized by the following expression:

$$\mathbf{x}_i^{'} = -1 + \frac{\mathbf{x}_i - \mathbf{x}_{i,\min}}{\mathbf{x}_{i,\max} - \mathbf{x}_{i,\min}} \times 2 \quad (12)$$

where $\mathbf{x}_i$ is the i^{th} original data, $\mathbf{x}_i^{'}$ is the normalized value of $\mathbf{x}_i$, $\mathbf{x}_{i,\max}$ and $\mathbf{x}_{i,\min}$ are respectively the maximum and minimum value of $\mathbf{x}_i$, i represents the dimension of data. Another advantage of the normalization method is to effectively reduce numerical difficulties during the calculation. Furthermore, we employed the PCA method to reduce the dimensions (see in [29], concrete methods see in [30]). The results of PCA are shown in Table 2.

Table 2. The contribution of the tenth principle component in PCA algorithm

k-th principal component	Contribution rate %	Cumulative contribution rate, %
1	55.3062	55.3062
2	21.5165	76.8227
3	7.9241	84.7468
4	5.7096	90.4564
5	3.2777	93.7341
6	2.4002	96.1343
7	1.4697	97.6040
8	1.2109	98.8149
9	0.4661	99.2810
10	0.3592	99.6402

Table 2 shows that the first five principal components (PCs) accounted for 93.73% cumulative contribution, after adding the sixth one, which reaches 96.13%. Hence, the first six PCs were selected to ensure the cumulative contribution rate higher than 95%.

The PCs, also called latent variables, can be used to model processes. However, the PCs have no obvious physical meaning. The PCs represent the most sensitive directions to the primary variable. The cumulative contribution rate of the first six PCs was over 95%, so the portfolio coefficients of only the first 6 PCs are given in Table 3. The coefficients can be interpreted as measures of the contribution of variables to the PCs. To reduce the dimensions of input variables, a specific approach was used to select the variables with the greatest contribution to the high sensitive PCs as inputs of soft sensors: the coefficient with the largest absolute value corresponding process variable in every PC is selected as a secondary variable.

Then, in Table 3, process variables 8, 4, 4, 2, 11 and 3, corresponding to the V_1 to V_6 respectively, were selected as secondary variables. Finally, process variable 2, 3, 4, 8 and 11 were determined as input variables of a soft sensor those denote dissolved oxygen, pH, dextrin flow, volume of dextrin and volume of water. The output variable is biomass concentration.

Results of the GPR soft sensor

The proposed GPR soft sensor was built on training data set after variable selection. Then, process data of three batches was used as test data set to evaluate the proposed GPR soft sensor. Figure 4 shows the test results.

Figure 4a-c gives the output dynamics of the GPR soft sensor and true data synchronously. The predictions are expressed as mean (solid line) with 2*std (std, standard deviation) error bars (dotted

Table 3. Portfolio coefficients of PCA

	1	2	3	4	5	6	7
V_1	-0.3394	-0.0660	-0.1784	0.0259	-0.0047	0.2299	-0.1412
V_2	0.1776	-0.0980	-0.2148	0.5741	-0.0112	0.5619	-0.0330
V_3	0.0653	0.1314	0.0602	-0.7042	-0.0118	0.5692	-0.2591
V_4	-0.0239	-0.5572	0.4876	0.0145	-0.0368	0.2819	0.3608
V_5	-0.1043	0.2963	-0.1224	0.1188	0.0532	-0.0411	0.3835
V_6	-0.0367	0.5091	0.6137	0.0817	-0.1681	-0.0503	0.0541
	8	9	10	11	12	13	14	15
V_1	-0.4106	-0.3434	-0.3907	-0.2595	0.3550	0.1204	-0.3368	0.1206
V_2	0.2090	0.1527	0.2262	-0.2325	0.2093	-0.1132	0.1325	0.1406
V_3	0.1067	0.0547	0.0873	-0.2103	-0.1181	-0.0721	0.0468	0.0074
V_4	-0.1382	-0.0464	-0.0500	0.1250	-0.2879	-0.0797	-0.1259	0.3091
V_5	-0.0205	-0.1057	-0.0245	-0.5674	-0.5248	-0.2695	-0.1603	-0.1135
V_6	0.0135	-0.0250	-0.0384	-0.0856	0.3870	-0.3177	0.0639	0.2409

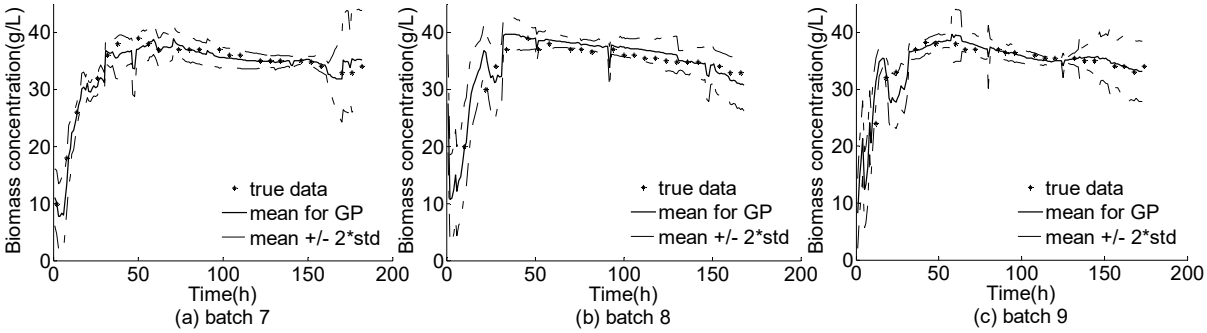

Figure 4. Variation curves of biomass prediction by the GPR soft sensor.

lines). The regions between the two dotted lines depict the confidence intervals.

To investigate the output uncertainty of the GPR soft sensor with different conditions, we set an abnormal condition in batch 9. We assumed that the DO sensor in batch 9 was abnormal (Figure 5). Figure 6 shows the output of the GPR soft sensor with the abnormal DO sensor.

Discussions

Figure 4 shows that the predicted microbial growth (output, *i.e.*, mean curves of the GPR soft sensor) approximately fits the microbial growth curve. The predicted biomass concentration curves, shown in Figure 4a-c, can also be divided into four stages: lag phase (in the beginning, though it is not obvious), log phase (during 10^{th}-50^{th} h), stationary phase (during 50^{th}-150^{th} h), and death phase (after the 150^{th} h). We can find that the predicted area of uncertainty, between the mean ±2std curves, in the log phase is bigger than that in other phases. That can be interpreted by fermentation mechanisms. Actually, biomass varies greatly and is with high uncertainty in the exponential phase. We can also found that only a few

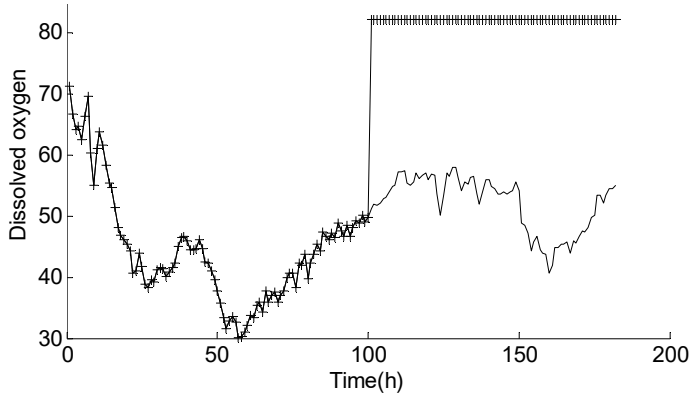

Figure 5. Variation curve of the abnormal DO sensor.

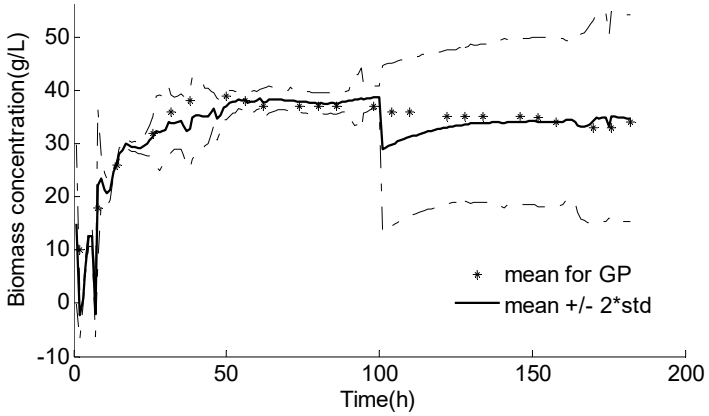

Figure 6. Output uncertainty of the GPR soft sensor with the abnormal DO sensor.

true measured points were out of the range of the area of uncertainty (see Figure 4a). This shows the predicted uncertainty reflects well the distribution of outputs. Therefore, we can conclude that the GPR soft sensor is suitable to model the uncertainty in fermentation processes for its probabilistic characteristic.

The performance of the GPR soft sensor with an abnormal condition was also investigated. As Figure 6 shows, when the DO sensor is abnormal, uncertainty output of the soft sensor is significantly greater than that in Figure 4c. That shows the uncertainty is sensitive to the distribution of input variables. The main reason is that the output uncertainty (*i.e.*, prediction distribution) of a GPR model depends on the distribution of input variables [27]. From the view of soft sensor maintenance, the uncertainty output can be used to design indexes of abnormal instruments or process states.

CONCLUSION

Microbial fermentation process is complex and nonlinear. The growth rates of microorganisms in different phases vary greatly, resulting in significant biomass concentration changes. For fermentation processes, dramatic changes usually exist in the exponential phase. Considering the uncertainty of microorganism growth, GPR combined PCA were used to build a soft sensor to measure biomass concentration on line in Erythromycin fermentation process. The proposed soft sensor not only gives the predicted values of biomass, but also gives the uncertainty of outputs synchronously. The adequate performance of the proposed soft sensor of modeling the uncertainty of estimates was evaluated with data obtained off-line of an industrial Erythromycin fermentation process. The output uncertainty can be regarded as measurement precision. Moreover, it is a useful reliability index of soft sensor or process condition in real applications.

Acknowledgements

The authors gratefully acknowledge the financial support provided by Natural Science Foundation of Jiangsu Province of China (grant no. BK20130531), the Priority Academic Program Development of Jiangsu Higher Education Institutions (grant no. PAPD 6) and Science & Technology Innovation Foundation of Ministry of Science & Technology for Small-Medium Enterprises, China (grant no. 12C26213202207).

REFERENCES

[1] J.R. Leigh, Modelling and Control of Fermentation Processes, Peter Peregrinus Ltd., London, 1987

[2] Y. Zhao, Studies on modeling and control of continuous biotechnical processes, Ph.D. Thesis, Norwegian University of Science and Technology, 1996

[3] G. Bastin, D. Dochain, O.L. Estimation, Adaptive control of bioreactors, Elsevier Science, Amsterdam, 1990

[4] H. Zhang, B Lennox, J. Process. Contr. **14** (2004) 41-50

[5] A. Jahanmiri, H. Rasooli, Biochem. Eng. J. **24** (2005) 115-123

[6] D.A. Gee, W.F. Ramirez, Biotechnol. Prog. **12** (1996) 132-140

[7] G. Liu, D. Zhou, H. Xu, C. Mei, Expert Syst. Appl. **37** (2010) 2708-2713

[8] L. Tian, Q.Q. Wang, A.Z. Cao, Adv. Mater. Res. **734** (2013) 2998-3002

[9] X. Dai, W. Wang, Y. Ding, Z. Sun, Comput. Chem. Eng. **30** (2006) 1203-1225

[10] K. Desai, Y. Badhe, S.S. Tambe, B.D. Kulkarni, Biochem. Eng. J. **27** (2006) 225-239

[11] A. Borin, M.F. Ferrão, C. Mello, L. Cordi, L.C. Pataca, N. Durán, R.J. Poppi, Anal. Bioanal. Chem. **387** (2007) 1105-1112

[12] J. Peng, F. Meng, Y. Ai, Bioresour. Technol. **138** (2013) 345-352

[13] S.G. Kulkarni, A.K. Chaudhary, S. Nandi, S.S. Tambe, B.D. Kulkarni, Biochem. Eng. J. **18** (2004) 193-210

[14] A. O'Hagan, J. Kingman, J. R. Stat. Soc., B **40** (1978) 1--42

[15] C.K. Williams, C.E. Rasmussen, in: Advances in Neural Information Processing Systems 8, D. S. Touretzky, Ed.,, The MIT Press, Cambridge, 1996, p. 514

[16] R. Grbić, D Slišković, P Kadlec, Comput. Chem. Eng. **58** (2013), 84-97

[17] F. di Sciascio, A.N. Amicarelli, Comput. Chem. Eng. **32** (2008) 3264-3273

[18] X. Li, H. Su, J. Chu, Chinese. J. Chem. Eng. **17** (2009) 95-99

[19] X. Yuan, Z. Ge, Z. Song, Chemometr. Intell. Lab. **138** (2014) 97-109

[20] S. Bhartiya, J. R. Whiteley, ISA. T. **40** (2001) 307-323

[21] M.D. Ma, J.W. Ko, S.J. Wang, M.F. Wu, S.S. Jang, S.S. Shieh, D.S.H. Wong, Control. Eng. Pract. **17** (2009) 1026-1034

[22] E. Zamprogna, M. Barolo, D.E. Seborg, J. Process. Control **15** (2005) 39-52

[23] T. Mejdell, S. Skogestad, Ind. Eng. Chem. Res. **30** (1991) 2543-2555

[24] M. Shakil, M. Elshafei, M. A. Habib, F. Maleki, Comput. Electr. Eng. **35** (2009) 578-586

[25] C. Mims, H. Dockrell, R. Goering, I. Roitt, D. Wakelin, Medical Microbiology (3rd ed.), Mosby Ltd., London, 2004, p. 489

[26] A. Novick, Annu. Rev. Microbiol. **9** (1955) 97-110

[27] C.E. Rasmussen, C.K. Williams, Gaussian processes for machine learning, The MIT Press, Cambridge, 2006, p. 16

[28] M. Gibbs, D.J. MacKay, Efficient implementation of Gaussian processes, Unpublished manuscript, Cavendish Laboratory, Cambridge, 1997, p. 2

[29] I. Jolliffe, Principal component analysis, John Wiley & Sons, New York, 2005, p. 28

[30] S. Valle, W. Li, S.J. Qin, Ind. Eng. Chem. Res. **38** (1999) 4389-4401.

CONGLI MEI
MING YANG
DONGXIN SHU
HUI JIANG
GUOHAI LIU
ZHILING LIAO

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China

NAUČNI RAD

SOFT SENZOR ZASNOVAN NA GAUSOVOJ REGRESIJI I NJEGOVA PRIMENA U PROCESU FERMENTACIJE ERITROMICINA

Fermentacija eritromicina je tipičan mikrobni process. Za određivanje biomase u procesu fermentacije eritromicina mogu se koristiti soft senzori zbog njihove niske cene, jednostavne izrade i sposobnosti da predvide teško merljive promenljive. Međutim, tradicionalni soft senzori, kao npr. soft senzori zasnovani na veštačkim neuronskim mrežama (ANN), mašini sa vektorima podrške (SVM) i tako dalje, ne mogu da predstavljaju mernu nesigurnost (preciznost merenja), što dovodi do teškoća u praksi. Gausova regresija (GPR) omogućuje novi okvir za rešavanje regresijskih problema. Merna nesigurnosti GPR modela sledi Gausovu raspodelu, izraženu preko srednje vrednosti i varijanse. Srednja vrednost prestavlja predviđeni izlaz. Varijansa se može posmatrati kao mera pouzdanosti predviđenog izlaza koji razdvaja GPR od ANN i SVM modela soft senzora. Predlaže se sistematski pristup zasnovan na GPR i analizi glavnih komponenti (PCA), da bi se uspostavio soft senzor za procenu biomase u procesu fermentacije eritromicina. Simulacije sa industrijskim podacima iz procesa fermentacije eritromicina pokazuju da predloženi GPR model soft senzora ima visok stepen modelovanja procena merne nesiguurnosti.

Ključne reči: fermentacije, soft senzori, merna nesigurnost, Gausova regresija, analiza glavnih komponenata.