On-line monitoring and chemometric modeling of *S.cerevisiae* fermentation processes with 2D spectrofluorometry

JONG IL RHEE[1,2,4], TAE-HYOUNG KANG[3,4], OK-JAE SOHN[1,4], SUN YONG KIM[1,4]
[1]BioProcess Technology Lab., [2]School of Applied Chemical Engineering,
[3]Department of Industrial engineering, [4]Research Center for Biophotonics,
Chonnam National University, YongBong-dong 300, 500-757 GwangJu, Republic of Korea

Abstract
During fermentation processes of *S. cerevisiae* 2D spectrofluorometry produces a large volume of spectral data, which can be analyzed using chemometric methods such as principal component analysis (PCA) and partial least square regression (PLS). PCA resulted in scores and loadings that were visualized in the score-loading plots and used to monitor the fermentation processes on-line. PLS was used to examine the correlation between the 2D fluorescence spectra and the process variables.

Key-Words: Chemometric modeling; Fermentation process; On-line monitoring; PCA; PLS; 2D fluorescence spectra

1 Introduction

A 2D spectrofluorometer has received considerable attention for the non-invasive monitoring and control of various biological processes among many spectroscopic methods. For example, it has been used to monitor the concentrations of the cell mass in *Saccharomyces cerevisiae* fermentations [1] and the level of the antibiotic polymyxin B production in *Bacillus polymyxa* cultivations [2].

The 2D spectrofluorometer allows many combinations of excitation and emission wavelengths to be scanned continuously, and the data that it produces can be handled using either an artificial neural network [3] or chemometric methods etc [4]. But artificial neural networks require a non-linear function such as a sigmoid equation to connect the nodes in order to establish some mathematical relationships for a bioprocess [5-6].

Chemometric methods such as principal component analysis (PCA) and partial least squares regression (PLS) are useful for the quantitative analysis of the spectroscopic data. PCA allows the entire spectrum to be analyzed quantitatively and provides a synthetic description of large data sets with a minimal loss of information. It has also been used to analyze the excitation-emission fluorescence matrices of olive oil [7] and to reduce the dimensions of the fluorescence spectra resulting from the monitoring of wastewater treatment processes [8]. The metabolic changes (e.g. FAD/FMN, pyridoxal-5-phosphate) in recombinant *E.coli* with time were analyzed qualitatively in a selective and nonselective environment using the PCA technique [9]. The PLS method has been used to analyze and model the spectral data. It is one of the most widely used multivariate calibration methods and has been applied extensively to the chemometric modeling of spectroscopic data, e.g. IR spectroscopic data. Chemometric models of 2D fluorescence spectra by PLS have been reported in some biotechnological processes. For example, PLS was used to establish mathematical relationships between the on-line collected fluorescence data and the off-line process data such as cell mass and polymyxin concentrations in *Bacillus polymyxa* cultivations [2]. The PLS models established using fluorescence measurement data were also used to predict the cell mass and substrates (glycerol, methanol) during non-induced and induced *Rhizopus oryzae* lipase production in *Pichia pastoris* fermentations [10], the protein and alkaloid concentrations during the fermentation of *Claviceps purpurea* [4] and CO_2 and O_2 compositions in exhaust gas in the cultivations of *Pseudomonas fluorescence* [11].

In this study, two fermentation processes with *Saccharomyces cerevisiae* were monitored on-line using a 2D spectrofluorometer. The monitoring of the fermentation process was interpreted with a reduction of the dimensions of the 2D fluorescence spectral data by PCA. The PLS-based calibration

models were established for a few fermentation state variables using the 2D fluorescence spectra collected and their prediction performance were evaluated.

2 Materials and methods

2.1 Fermentations with *S.cerevisiae*

Yeast *S. cerevisiae* ATCC7754 (American Type Cell Collections, USA) was employed for the production of intracellular glutathione (GSH). The fermentation was performed at pH 5.5 and 30 °C in the bioreactor with 1 vvm of the aeration rate and 350 rpm of the stirrer speed. The fermentation medium consisted of glucose, salts (NaCl, KH_2PO_4, KCl, $MgSO_4$), a trace element solution, vitamin solution, and three precursors (glutamic acid, L-cysteine, glycine) for GSH [12]. Some details about the analysis of GSH and L-cysteine, etc. were described in our previous paper [13].

2.2 Fermentation system with a 2D spectrofluorometer

The fermentation system consisted of a 2.5 liter stainless steel stirred tank reactor (working volume: 5.0 liter, KoBiotech Co., Korea), a pH sensor and a DO-meter (Mettler-Toledo Co., USA), an O_2-/CO_2-analyzer (Lokas Co., Korea), as well as temperature-, stirrer speed-, antifoam- and pH controllers. A 2D spectrofluorometer (Model F-4500, Hitachi Co., Japan) was connected to a quartz window in a 19-mm electrode port of the stainless steel reactor. It was operated in the same conditions as shown in our previous work [13]

2.3 Chemometric methods

2.3.1 Principal component analysis (PCA)

The whole 2D fluorescence spectral data gathered during fermentation can be structured in the form of the fermentation time and combinations of the excitation and emission wavelengths. PCA decomposes a given fluorescence spectral data matrix (**X**) as the sum of the outer product of vectors $\mathbf{q_a}$ and $\mathbf{p_a}$ plus a residual matrix **E**, as shown in the following equation.

$$\mathbf{X} = \sum_{a=1}^{n} \mathbf{q_a p_a^T} + \mathbf{E} = \mathbf{QP}^T + \mathbf{E} \qquad (1)$$

where **Q** is known as the score matrix and contains information on the relationship between samples. **P** denotes the loading matrix and includes information on the relationship between variables. In PCA, the first principal component (q_1 and p_1 pair or PC1) captures the largest amount of variances in the data. Each subsequent PC (PC2, PC3 …) captures the largest possible amount of variance remaining at that step.

2.3.2 Partial least squares regression (PLS)

PLS attempts to identify the factors (called latent variables, LVs) that not only capture the largest amount of variance in the fluorescence spectra, but also allows a linear correlation to be obtained between the spectral data and process variables through an inner relation. Details of the PLS regression technique are described elsewhere [14].

The above-mentioned chemometric methods were applied using the MatLab 6.2 (The MathWorks, Inc., Natick, USA) program with the statistical toolbox [15].

2.3.3 Input data and model evaluation

A total of 1558 (number of combinations of excitation and emission wavelengths, CWL) x 600 (scan numbers) spectral data were collected in the case where the fluorescence sensor was operated with a step size of 10 nm in the excitation wavelength range of 250-650 nm and the emission wavelength range of 280-650 nm, with a full spectral scan being taken every 5 min for 50 fermentation hours[16]. After filtering some of the light scattering data, the total CWLs could be reduced to 493 CWLs. The spectral data concerning the scan numbers, i.e. fermentation time, can be selected randomly and split into 70% (training data) and 30 % (prediction data) portions using the cross-validation technique. The training data was subdivided into calibration data (70%) and validation data (30%) using the above-mentioned cross-validation technique. After a multivariate calibration model had been established using the calibration data, it was tested using the validation data. The prediction data was also used to test the prediction capability of the calibration models. The on- and off-line measurement data were also interpolated or extrapolated and compared with the model data.

The performance of a chemometric model can be evaluated by estimating the 'average' deviation of the model from the data. This means that an evaluation of the calibration (RMSEC), cross-validation (RMSECV) and prediction power

(RMSEP) of a chemometric model can be performed using the root mean squared errors (RMSE) for the calibration data, validation data and prediction data. The RMSE is given as follows [17]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_{i,\,\mathrm{model}} - y_i)^2}{n}} \qquad (2)$$

where $\hat{y}_{i,\mathrm{model}}$ is the value of the property of interest calibrated, validated or predicted by the model for object i, and y_i is the known value (i.e. on- & off-line data) of the property of interest for object i, where n is the number of sample data points.

3 Results and discussion

3.1 Fermentation processes with *S. cerevisiae*

During the fermentation of *S. cerevisiae*, cysteine was added to the processes at 11.0 h, whereas glutamic acid and glycine were added to the bioreactor at the beginning of the fermentation in CultPro1 and at 11.0 h in CultPro2, respectively [12]. The cell growth and GSH production in CultPro1 and CultPro2 are shown in Fig. 1. After 10.5 h, the DCW in CultPro1 was higher than that in CultPro2, because a higher concentration of glucose (20 g/L) was added in CultPro1 than in CultPro2 (5 g/L). The glucose concentration in CultPro1 decreased very rapidly and its concentration reached zero after 10.5 h. However, a lower concentration of cysteine (8 mM) was introduced into the CultPro1 than into the CultPro2 (16 mM). Higher amounts of intracellular GSH were produced in CultPro1 than in CultPro2, but its maximum concentration was reached at 22.0 h in both processes. The concentrations of CO_2 and O_2 in the exhaust gas of the two processes were maximal and minimal at 10.5 h, respectively, although their concentrations were quite different.

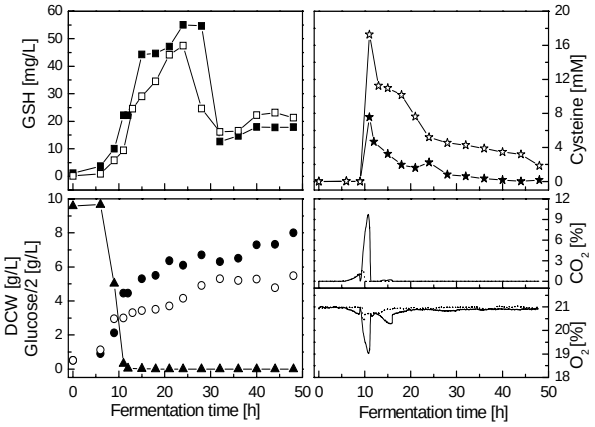

Fig. 1. On- and off-line measurement data of two fermentation processes with *S. cerevisiae*: CultPro1 data for DCW (●), GSH (■), glucose (▲), CO_2 (—), O_2 (—), and cysteine (ξ); CultPro2 data for DCW (○), GSH (□), CO_2 (---), O_2 (---), and cysteine (ψ).

3.2 On-line Monitoring of the fermentation processes with analysis of 2D fluorescence spectra by PCA

The score and loading plots produced by the PCA do not only help to understand the relationship between each fluorescence spectrum and the cellular states, but also provide some qualitative information on the fermentation process.

The large volume of 2D fluorescence spectra produced during fermentation can be reduced in dimension by PCA. After filtering out some of the light scattering data, the spectral data can be reduced and used as the column of the fluorescence spectral matrix (**X**) in Eq. (1). The whole spectral matrix can then be decomposed into the score and loading data matrices. The number of columns in the fluorescence spectral matrix, i.e. 378 CWLs in this study, can affect the computation time of the data matrix and the amount of information. Therefore, a total of 98 CWLs, which were selected by using SOM algorithm [13], was employed to calculate the variance which can be captured by each PC. In the case of CultPro1, PC1, PC2, PC3, PC4 and PC5 captured 61.8%, 4.71%, 1.52%, 1.10% and 1.04% of the total variances in the entire fluorescence spectra, respectively, while each score value for CultPro2 was 57.1%, 3.10%, 1.98%, 1.32%, and 1.23%.

The score plots of the PCs in CultPro1 and CultPro2 are shown in Fig. 2. The cell growth, as well as the difference in the time of addition of the two amino acids (glutamic acid and glycine) to the process, can

be interpreted by comparing the score plots of the two processes. From the score plots, the trends in the score values of the PCs at the beginning of fermentation and at 11.0 h in CultPro1 were different from those in CultPro2. Glutamic acid and glycine were added to CultPro1 at the beginning of fermentation, whereas they were added to CultPro2 at 11.0 h. This difference can be observed in the change of the scores of PC1 and PC2 in the score plots. The score values of PC1 and PC3 reflected the addition of cysteine to both processes at 11.0 h. In the score plots of PC1 and PC3, the increase in the score values of PC3 starting from 11.0 h might also result from the change from an oxidative to an oxidoreductive metabolism [8], for example the conversion of glucose to ethanol. The increase in the scores of PC3 starting from 35.0 h may represent the production of other metabolites or the degradation of GSH within the cells, as shown in Fig. 1.

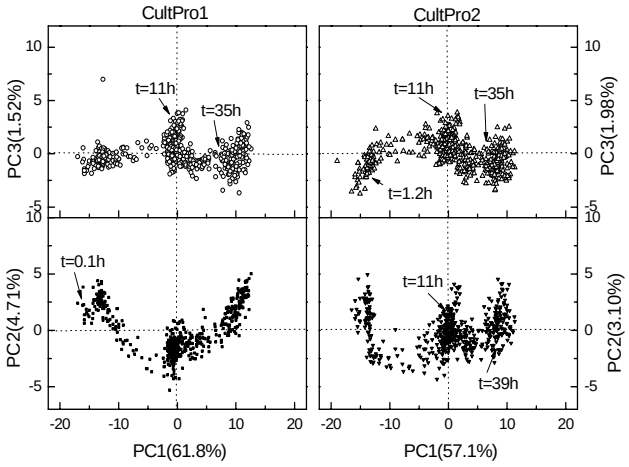

Fig. 2. Score plots of PC1, PC2 and PC3 for CultPro1 and CultPro2

3.3 Chemometric modeling of 2D fluorescence spectra

The 2D fluorescence spectra collected during fermentation have been used to establish PLS models. The calibration and prediction performance of the PLS models with the optimum number of LV have been compared for a few process parameters.

3.3.1 Determining optimum numbers of LV

The number of LVs in the PLS model influences the outcome of a chemometric model. The optimum number of LVs was determined by correlating some of the prediction values derived from the chemometric model with a few on- and off-line measurement data points, i.e. by computing the minimum value of the RMSECV. In a process, CultPro1, the optimum number of LVs for DCW and cysteine were determined by calculating the RMSECV values using 154 validation data points (30 % of the training data). Fig.3 shows a plot of the values of the RMSECV for DCW and cysteine as a function of the number of LVs. The RMSECV values of cysteine were relatively low at 3 LVs for the PLS model. However, DCW has low RMSECV values at 4 LVs for the PLS model. This means that 3 or 4 LVs may be chosen to construct each calibration model with 70 % of the training data. The optimal number of LV in the PLS were calculated for the process parameters in the fermentation processes.

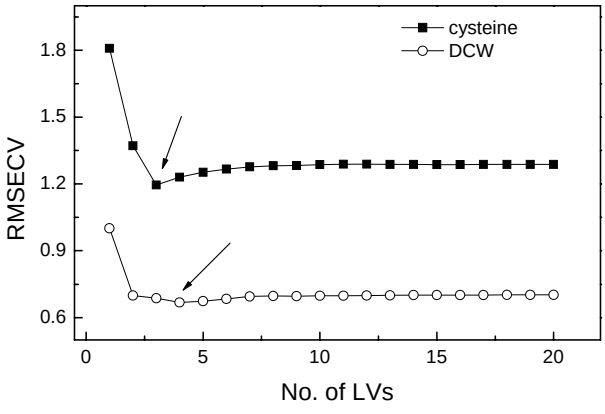

Fig. 3 RMSECV values of DCW and cysteine according to number of LVs for the PLS model in CultPro1

3.3.2 Calibration and prediction performance of the PLS models

The calibration power of the PLS models using the 2D fluorescence spectra obtained in a fermentation process was calculated, and the RMSEC value for a process parameter was evaluated. The RMSEC value usually provides some information regarding the fit of the calibration model to the measurement data of a particular process parameter.

The results of each calibration model were also associated with the on- and off-line measurement data. Table 1 shows the RMSEC values and the correlation

coefficients (R^2) between the PLS-calibrated DCW and cysteine and off-line measured DCW and cysteine in CultPro1 and CultPro2

Table 1 RMSEC values and correlation coefficients (R^2) for DCW and cysteine in the CultPro1 and CultPro2.

	CultPro1		CultPro2	
	DCW	Cysteine	DCW	Cysteine
RMSEC	0.50481	0.7862	0.38334	1.6046
R^2	0.994	0.985	0.991	0.981

The lower RMSEC values and higher correlation coefficients of the PLS model indicate that the PLS model using the 2D fluorescence spectra provides a good fit to the on- and off-line measurement data and can be used to predict the results for other processes.

The prediction power of the PLS models to new process data was investigated using 30 % of the total spectral data (i.e. prediction data), which was not included in the computation of the calibration model for the process parameters. Some process parameters were predicted by the PLS models and compared with the on- and off-line measurement data.

In Fig. 4 the data predicted using the PLS models are presented along with interpolated measurement data of the DCW and cysteine in CultPro2.

From Fig. 4 there was a slight difference between the measured data and the data predicted by the DCW model. The RMSEP value for DCW was 0.49546 g/L, while it was 1.9058 mM for cysteine. The correlation coefficients (R^2) between the prediction and measurement data were 0.9493 for DCW, and 0.9416 for cysteine, respectively. The high correlation coefficients between the prediction data and the measurement data and the low RMSEP values highlight the good statistical power of the chemometric model based on the 2D fluorescence spectra to predict new data.

The RMSEP values and correlation coefficients between the predicted and measurement data were calculated for the other parameters of the fermentation processes.

Chemometric methods provide a rich tool set for process monitoring and modeling because they not only handle a large number of process variables but provide data compression and a great deal of process diagnostic information. Furthermore, the development of process modeling is straightforward because they are deterministic and the computational requirements are relatively low.

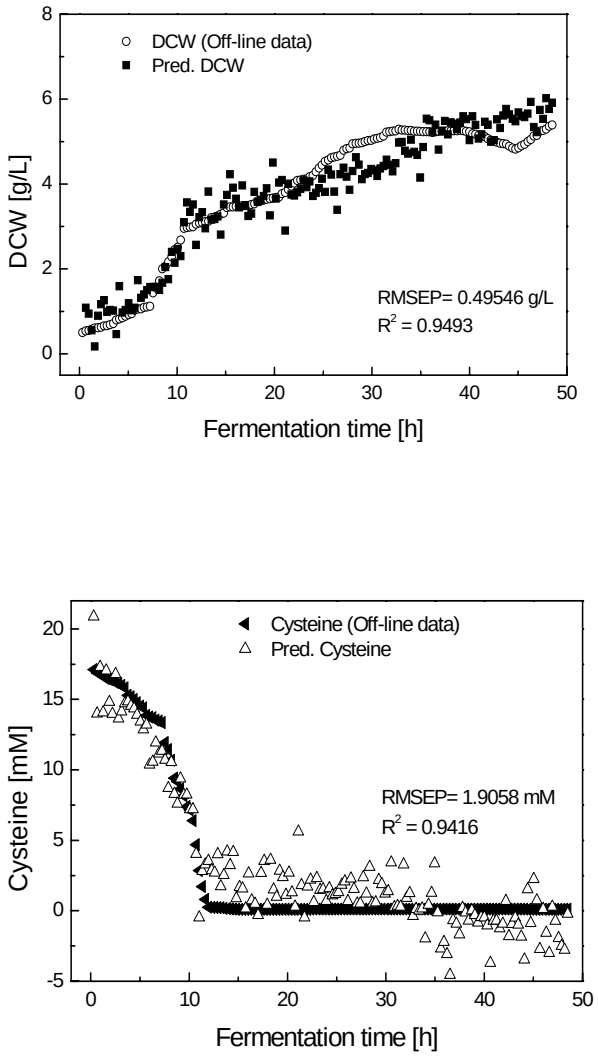

Fig. 4 Correlation between the PLS-predicted DCW and cysteine data to off-line measured DCW and cysteine data in CultPro2

4 Conclusion

In this study, we presented the monitoring of fermentation processes of S. cerevisiae using PCA as well as the process modeling with the PLS methods. A large amount of 2D fluorescence spectral data were collected in the fermentation processes, and the dimension of the data was reduced using PCA. The score and loading plots were used to describe the qualitative tendency of the

fermentation processes. The PLS models were used to analyze the 2D fluorescence spectra. The PLS model showed good calibration and prediction power.

Acknowledgements
This work was supported by grant No. RTI04-03-03 from the Regional Technology Innovation Program of the Ministry of Commerce, Industry and Energy (MOCIE), and also in part by BK21 program from the Ministry of Education & Human Resources Development, Republic of Korea

References:

[1] M.B. Haack, A. Eliasson, L. Olsson, On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. Journ. Biotech. 114, 2004, 199-208

[2] A.E. Lantz, P. Jorgensen, E. Poulsen, C. Lindemann, L. Olsson, Determination of cell mass and polymyxin using multi-wavelength fluorescence. Journ. Biotechnol. 121, 2006, 544-554

[3] G. Wolf, J.S. Almeida, C. Pinheiro, V. Correia, C. Rodrigues, M.A.M. Reis, J.G. Crespo, Two-dimensional fluorometry coupled with artificial neural networks: a novel method for on-line monitoring of complex biological processes. Biotech. Bioeng. 72, 2001, 297-306

[4] D. Boehl, D. Solle, B. Hitzmann, T. Scheper, Chemometric modeling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. Journ. Biotech. 105, 2003, 179-188

[5] I.A. Basheer, M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application. Journ. Microbiol. Meth. 43, 2000, 3-31

[6] K.I. Lee, Y.S. Yim, S.W. Chung, J. Wei, J.I. Rhee, Application of artificial neural networks to the analysis of 2D fluorescence spectra in recombinant *E.coli* fermentation processes. Journ. Chem. Tech. Biotech. 80, 2005, 1036-1045

[7] F. Guimet, J. Ferre, R. Boque, F.X. Rius, Application of unfold principal component analysis and parallel factor analysis to the extrapolatory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy. Anal. Chim. Acta 515, 2004, 75-85

[8] B. Tartakovsky, L.A. Lishman, R.L. Legge, Application of multi-wavelength fluorometry for monitoring wastewater treatment process dynamics. Water Res. 30, 1996, 2941-2948

[9] L. Johansson, G. Liden, A study of long-term effects on plasmid-containing *Escherichia coli* in carbon-limited chemostat using 2D-fluorescence spectrofluorimetry. Biotechnol. Prog. 22, 4, 2006, 1132-1139

[10] A. Surribas, D. Geissler, A. Gierser, Th. Scheper, B. Hitzmann, J.L. Montesinos, F. Valero, State variables monitoring by in situ multiwavelength fluorescence spectroscopy in heterologous protein production in *Pichia pastori.* Journ. Biotech. 124, 2006, 412-419

[11] E. Skibsted, C. Lindemann, C. Roca, L. Olsson, On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration. Journ. Biotech. 88, 2001, 47-57

[12] H. Shimizu, K. Araki, S. Shioya, K.I. Suga, Optimal production of glutathione by controlling the specific growth rate of yeast in fed-batch culture, Biotechnol. Bioeng. 38, 1991, 196-205

[13] J.I. Rhee, K.-I., Lee, C.-K. Kim, Y.-S. Yim, S.-W. Chung, J. Wei, K.-H. Bellgardt, Classification of two-dimensional fluorescence spectra using self-organizing maps. Biochem. Eng. Journ. 22, 2005, 135-144

[14] P. Geladi, B.R. Kowalski, Partial least-squares regression: tutorial. Anal Chim. Acta 185, 1986, 1-17

[15] Maltlab manual, vers. 6.1, The Mathworks, Inc., USA, 2002

[16] S. Marose, C. Lindemann, T. Scheper, Two-dimensional fluorescence spectroscopy: A new tool for on-line bioprocess monitoring. Biotech. Progr. 14, 1998, 63-74

[17] R. Bo, Multivariate calibration. What is in chemometrics for the analytical chemist? Anal.Chim. Acta 500, 2003, 185-194