TWO-STAGE ROBUST ESTIMATORS FOR PRODUCT CONCENTRATION IN A PENICILLIN FERMENTATION

A. W. Chow[1], A. Ahmad[2,*],

[1]Laboratory of Process Control, Department of Chemical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
[2]Department of Chemical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
*Corresponding author. Phone: +607-5535502, Email: arshad@fkkksa.utm.my

ABSTRACT

Two strategies are proposed for inferential estimations of penicillin concentration in a fed-batch fermentation process. The first scheme employs a partial least squares (PLS)-based estimator to directly estimate the intended product concentration. The second introduces a two-step estimation where biomass concentration is first estimated and used as the input to the final estimator along with other process variables. The results confirmed the strength of the two-stage approach. To further extend the capability of the estimation scheme, a wavelet-based filter is added to the estimation model. The improved performance obtained proved the efficacy of the wavelet filter.

Keywords: Inferential Estimation, PLS, Wavelet, Penicillin Fermentation.

1 INTRODUCTION

Despite the fast developments in industrial process technologies, there remain systems that are difficult to control due to the fact that their product qualities cannot be measured rapidly and reliably. This impacts the overall plant profitability and catalyzes the development of sensor technology. In addition to the advancement in measurement hardwares, soft-sensor has also received considerable interests. This inferential estimation strategy employs measurements of secondary process outputs that are readily available in industrial processes such as temperatures and flow rates to infer primary outputs, such as product quality (Joseph and Brosilow, 1978). The inferential estimator employs some forms of process models and if those models are accurate and robust, consistent and reliable estimation can be produced.

Since 1970s, significant efforts have been made and various approaches, including that of theory-based, black box, and statistical methods, have been introduced. However, in this paper, the focus is on partial least square regression (PLS) model, which is one of the statistical methods. Compared to classical multiple linear regressions (MLR) and principal component regression (PCR), PLS has been widely studied and applied recently due to its robustness (Geladi and Kowalski, 1986). In addition, PLS was also proven to give good prediction when dealing with large numbers of highly correlated measured variables without over-fitting (Kresta et al., 1994). This method has also been proven to be more robust to missing data and sensor failure.

2 PRELIMINARIES

2.1 PARTIAL LEAST SQUARES REGRESSION

Partial least squares (PLS) regression is a multivariate method ideal for studying the variation in large numbers of highly correlated process variables (X) and relating them to a set of output variables (Y). It handles this by projecting the input-output data down into a latent space, extracting a number of principal factors with an orthogonal structure, while

capturing most of the variance in the original data (Wold, 1985). The schematic diagram of the PLS model is shown in Figure 1.
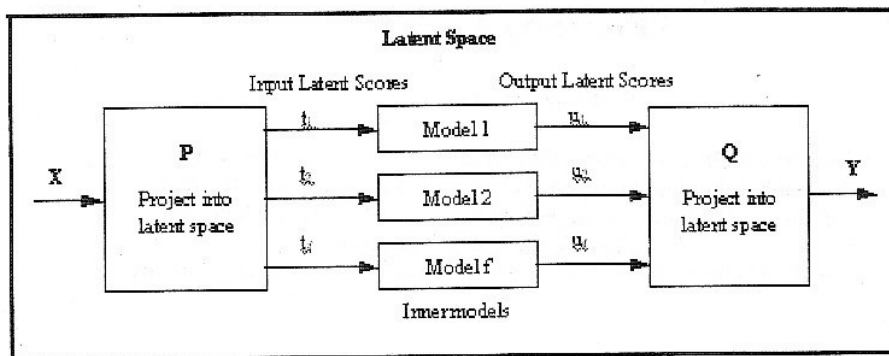

FIGURE 1. A schematic diagram of the PLS model (Adebiyi and Corripio, 2003)

The PLS model consists of outer relations (X and Y block individually) and an inner relation (linking both blocks). These relations are in matrix form, where X and Y represents the matrixes of independent and dependent variables, respectively. The input X is projected into the latent space by the input-loading factor, P to obtain the input scores, T. In a similar way, the output scores, U is obtained by projecting the output Y into latent space through the output-loading factor, Q. These relations, which are called outer relations, are shown in Equation (1) and (2).

$$X = TP^T + E_f \tag{1}$$

$$Y = UQ^T + F_f \tag{2}$$

Here, the matrixes E_f and F_f are residuals of X and Y, respectively. The residuals can be calculated from Equation (3) and (4). Initially, $X = E_0$ and $Y = F_0$.

$$E_f = E_{f-1} - T_f P_f^T \tag{3}$$

$$F_f = F_{f-1} - U_f Q_f^T \tag{4}$$

The inner relation that captures the relationship between the input and output latent scores is written in Equation (5).

$$U = BT \tag{5}$$

The procedure of determining the scores and loading vectors is carried out sequentially from the first factor to the fth factor. Scores and loading vectors for each factor are calculated from the previous residuals (outer relationship). This is performed until the number of required factors is extracted or the residual is below certain threshold (Adebiyi and Corripio, 2003).

2.2 TWO-STAGE ESTIMATION

The two-stage estimation model is constructed where X_{est} is first estimated and used as the input to the final estimator along with other process variables (X_2, X_3) in order to predict the output, Y. The schematic illustration of the two-stage estimation model is depicted in Figure 2.
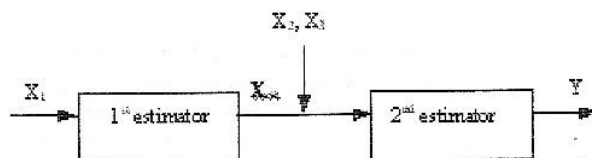
FIGURE 2. The schematic illustration of a two-stage estimation model

2.3 WAVELETS

Wavelet transform is a frequency domain technique that can be applied to both continuous and discrete time signals. It is more flexible compared to Fourier methods as it enables localization of signals both in time and frequency (Alsberg *et al.*, 1998). The time-frequency decomposition is done in such a way that the frequency resolution is matched to the corresponding time scale. Another property of the wavelet transform is that it can use a variety of different basis functions with different properties. Wavelets are therefore good in signal analysis and compression.

In this paper, wavelet transform is used to filter signal noise with minimum loss of process information. The procedure begins with noisy process data transformed into wavelet domain. Then the wavelet coefficients (except at the coarsest level) are subjected to thresholding. This is followed by inverse transform which of the diagonal filter to produce a noised-reduced version of the measurement data

3 PROBLEM DEFINITION

This paper aims at formulating a robust inferential estimator using two-stage estimation strategy, assisted by a wavelet-based filter to cope with external disturbances as well as changes in operating conditions.

3.1 PROCESS DESCRIPTION

The case study considered here is the penicillin fermentation process. This process is characterized by two distinct regimes. The initial phase is to grow cells in a batch culture. It is also known as rapid-growth phase. Then it is followed by a fed-batch operation to promote the biosynthesis of penicillin, i.e. the slow-growth phase. For the batch process, there is no continuous input of substrate and cells grow in the culture medium that contains the necessary nutrients. After a period of time, when the concentration of substrates in the fermentor drops to a certain value, the process is switched to a fed-batch mode and the fermentor is fed with a controlled supply of fresh substrates.

The process is sensitive to temperature and pH change since it involves living microorganisms. To facilitate growth the operating temperature should be within the range of 23 °C to 25 °C with the pH maintained about 5. In this work, temperature and pH in the fermentor are kept constant at 25 °C and 5.0, respectively. Temperature is maintained by manipulating the cooling water and steam flow rates, while pH is controlled through manipulation of the input of concentrated acid and base solutions flows into the system.

3.2 PROCESS SIMULATION

The simulation study conducted here employed a benchmark simulator called PENSIM. Details of the mathematical model as well as the mathematical tools used can be found in the work of (Birol et al. 2002). Analyses of dynamic behavior were performed to investigate the sensitivity and the dynamics of the process variables. These were accomplished by imposing step changes to various process inputs. The responses of process outputs were examined and used as a guide for variables selections for the model development stage.

4 DEVELOPMENT OF PROCESS ESTIMATOR

For the purpose of model development, a few sets of data were generated using PENSIM simulator. These data sets were used for testing, training and validating the estimation model. The model was intended for estimating the penicillin concentration in the culture broth. The performance of the estimators was evaluated based on mean squared error of prediction (MSE). The purpose of the evaluation was to investigate the robustness of the models.

4.1 PLS ESTIMATOR

The PLS estimator was developed in MATLAB environment. The NIPALS algorithm of PLS (as shown in Table 1) was used in this work.

TABLE 1. The NIPALS algorithm of PSL (Geladi and Kowalski, 1986; Gonzalez et al., 2003; Kresta *et al.*, 1994)

Step	Procedure	
0	Mean centre and scale X and Y	
1	Set u equal to a column of Y	
2	Calculate the input weights w by regressing X on u	$w^T u = u^T . X / u^T . u$
3	Normalize w to unit length	$w^T_{new} = w^T_{old}/\|w^T_{old}\|$
4	Calculate the X matrix scores	$t = X . w / w^T . w$
5	Calculate the output loadings q by regressing Y on t	$q^T = t^T . Y / t^T . t$
6	Normalize q to unit length	$q^T_{new} = q^T_{old}/\|q^T_{old}\|$
7	Calculate the Y matrix scores	$u = Y . q / q^T . q$
8	Check convergence on u: if yes, go to step 9; if no, go to step 2	
9	Compute the input loadings p by regressing X on t	$p^T = t^T . X / t^T . t$
10	Find the regression coefficient b for the linear relation	$b = u^T . t / t^T . t$
11	Calculate the residual matrixes	$E = X - t . p^T$ $F = Y - b . t . q^T$
12	If additional PLS dimensions are necessary then replace X and Y by E and F and repeat steps 1 to 9.	

4.2 TWO-STAGE ESTIMATOR

The 2-stage estimator proposed here was built by constructing two PLS models in series. The output of the first estimator was the biomass concentration, and was used as the input to the second estimator along with substrate concentration and fermentor volume.

4.3 WAVELET FILTER

In order to improve the robustness of the estimator, a wavelet-based filter was added to the model. This was accomplished using both Wavelet Toolbox and MATLAB programming language. Here wavelet *sym4* function provided by the toolbox was employed.

5 RESULTS AND DISCUSSIONS

The models were tested on data with different operating conditions as follows:
- Data A - single batch (normal operating condition)
- Data B - combination of several batches of data sets (with small changes to process variables)
- Data C - combination of several batches of data sets (with larger changes to process variables)

The Mean Squared Error (MSE) of the single stage estimator, two-stage estimator, and two-stage estimator with a wavelet-based filter, are summarized in Table 2. Graphical illustrations are also provided by Figure 3, 4 and 5.

Table 2. Summary of the results for PLS, two-stage estimator, and two-stage estimator with a wavelet-based filter

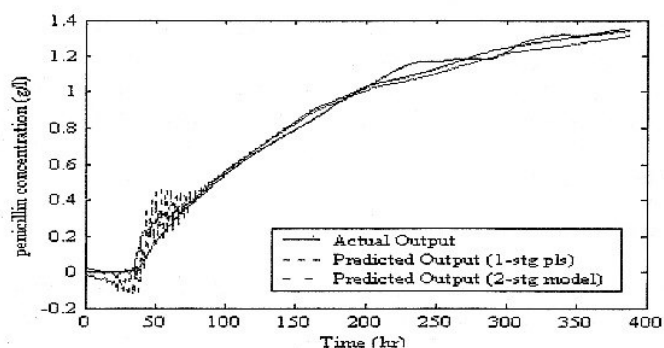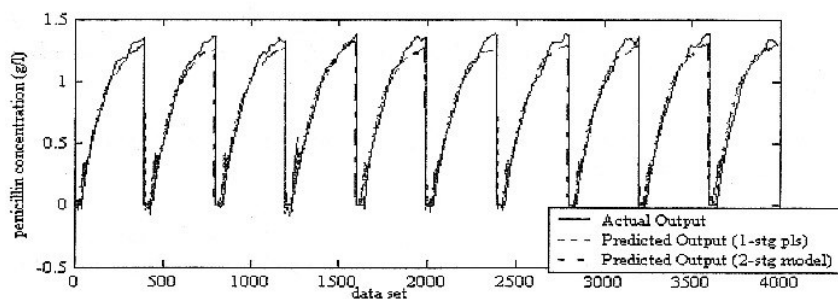| Model	MSE			
	Data A	Data A + noise	Data B	Data C
PLS	3.44 E-03	3.48 E-03	7.84 E-03	1.91 E-02
2-stage estimator	9.93 E-04	1.42 E-03	7.58 E-03	1.51 E-02
2-stage estimator (with a wavelet-based filter)	9.38 E-04	8.94 E-04	7.53 E-03	1.50 E-02

FIGURE 3. Estimation Results of Data A

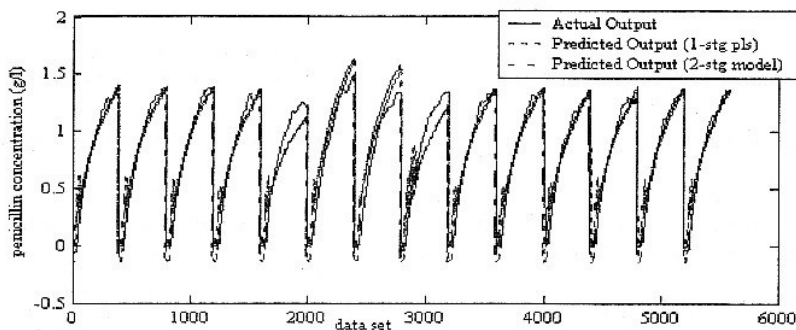FIGURE 4. Estimation Results of Data B

FIGURE 5. Estimation Results of Data C

In all cases, reasonable estimation qualities were obtained. The single stage PLS model was tested on four data sets, the results of prediction were good and acceptable. The predicted product concentration was close to the actual value for Data A and B, as depicted in Figure 3 and 4. But for Data C, the accuracy of prediction was decreased due to the increasing percentage of variation in the process variables. This was shown in Table 2 and Figure 5.

Compared to the single stage estimator, the two-stage estimator performed better in changing operating conditions and also noisy environment. Referring to Table 2, the MSE was reduced by utilizing the two-stage estimator. For data A, data A with noise, data B, and data C, the MSE has decreased 71.1%, 59.2%, 3.3% and 20.9%, respectively. However, similar to the PLS model, the MSE of the model increased with the increase of the variation in the process variables. Despite that, the predicted trends still followed the actual ones. Referring to the overall results, the two-stage estimator with a wavelet-based filter provided the best results. This was shown by its lowest MSE (as illustrated in Table 2).

6 CONCLUSION

This paper has described the application of PLS-based inferential estimator for the product concentration of a penicillin process. The results have proven the capability of the various models proposed. While estimation accuracy was improved by introducing the two-stage PLS estimation scheme, robustness was enhanced by the wavelet-based filter added to estimation model. The results presented here are some preliminary findings and further fine-tunings are currently on-going.

ACKNOWLEDGEMENTS

REFERENCES

Adebiyi, O.A. and Corripio, A.B. (2003). "Dynamic Neural Networks Partial Least Squares (DNNPLS) Identification of Multivariable Processes." Computers & Chemical Engineering. 27. 143-155.

Alsberg, B.K., Woodward, A.M., Winson, M.K., Rowland, J.J. and Kell, D.B. (1998) Analytica Chimica Acta. 368. 29-44.

Birol, G., Ündey, C. and Cinar, A. (2002). "A Modular Simulation Package for Fed-batch Fermentation: Penicillin Production." Computers and Chemical Engineering, 26, 1553-1565.

Geladi, P. and Kowalski, B.R. (1986). "Partial Least-Squares Regression: A Tutorial." Analytica Chimica Acta. 185. 1-17.

Gonzalez, G.D., Orchard, M., Cerda, J.L., Casali, A. and Vallebuona, G. (2003). "Local Models for Soft-sensors in A Rough Flotation Bank." Minerals Engineering. 16. 441-453.

Joseph, B. and Brosilow, C.B. (1978). "Inferential Control of Processes." AIChE Journal. 24(3). 485-5508.

Kresta, J.V., Marlin, T.E. and MacGregor, J.F. (1994). "Development of Inferential Process Models Using PLS." Computers & Chemical Engineering. 18(7). 597-611.

Wold, H. (1985). "Partial Least Squares." In Encyclopedia of Statistical Sciences. New York: Wiley. Vol 6. 581-591.

Zhang, H. and Lennox, B. (2004). "Integrated Condition of Monitoring and Control of Fed-batch Fermentation Processes." Journal of Process Control. 14. 41-50.