International Journal of Control and Automation
Vol.10, No.9 (2017), pp.27-36
http://dx.doi.org/10.14257/ijca.2017.10.9.03

Prediction from Uncertain Inputs for Partial Least Squares Regression

Zehui Chen[1], Weiguo Zhang[1], Jianghong Ren*[2] and Shaowei Gong[1]

[1]*School of Economics and Business Administration, Chongqing University, Chongqing, Chongqing 400044, PR China*
[2]*School of Automation, Chongqing University, Chongqing, Chongqing 400044, PR China*
[1]*zhhchen@cqu.edu.cn,* [2]*wgzhang@cqu.edu.cn,*[3] *rjh@cqu.edu.cn,*
[4]*gongshaowei@gmail.com*

Abstract

This paper presents two approaches to calculating the predictions for a partial least squares (PLS) regression model in the presence of uncertain inputs (predictors, covariates). Input uncertainty is a common issue associated with the practical application of PLS, such as in the use of multi-way PLS for the modeling of a batch process and the development of multi-step ahead predictions in time series. By assuming a multivariate normal distribution for the input variables, two methods are considered to accommodate the input uncertainty when calculating the mean and prediction intervals for a PLS model: Monte Carlo approximation and an analytical method. The proposed approaches are demonstrated for the prediction of the end-of-batch quality of a simulated batch polymerization process.

1. Introduction

Partial least squares (PLS) regression is one of the most widely applied regression techniques in the chemometric community, with a wide variety of applications including spectroscopic calibration, empirical modelling and the monitoring of the performance of manufacturing processes (Geladi and Kowalski[1]; Nelson *et al.*[2]; Nomikos and MacGregor[3]; Mehmood *et al.*[4]; Liu[5]; Dong[6]). Following the calculation of a PLS model, prediction intervals can be obtained (Denham[7]; Faber and Kowalski[8]; Serneels *et al.*[9]; Lu *et al.*[10]) to quantify the predictive uncertainty within the regression model.

More recently, it has been identified that in some application areas, the regression model must be capable of making predictions from uncertain input variables (predictors, covariates). For example, when either single or multiple inputs are not available due to data recording issues, values for these are typically estimated from the observed data, consequently input uncertainty materializes as a result of the estimation procedure (Nelson *et al.*[2][11]; Nomikos and MacGregor[3]; Dijkstra *et al.*[12]). Adopting the 'mean-replacement' approach (Nelson *et al.*[2]; Nomikos and MacGregor[3]; Mehmood *et al.*[4]), the missing data is replaced by its estimated mean value, and thus the impact of input uncertainty on the predictions is ignored. More recently, Nelson *et al.* [11] recognised this issue and they proposed calculating the predictive variance due to the presence of missing data. Missing data is not the only source of input uncertainty. Uncertainty is also present in the input variables in time series when utilising an auto-regressive model for recursive multi-step ahead prediction. This is a consequence of the fact that the one-step ahead output that is predicted at the current time point is uncertain, and this serves as an input in the recursive calculation of the multi-step ahead output

Received (June 21, 2016), Review Result (August 25, 2016), Accepted (December 6, 2016)

ISSN: 2005-4297

predictions (Girard *et al.*[13]; Liu[5]). In Girard *et al.*[13] the focus was to develop a method to incorporate input uncertainty for a specific non-parametric regression model, the Gaussian process.

This paper extends the algorithm specifically designed for handling missing data (Nelson *et al.*[11]). A more general situation is considered where input uncertainty can arise from different sources, including missing data and multi-step ahead predictions. The uncertainty of the input variables is characterized by a multivariate normal distribution and this is taken into account when making predictions with a PLS model. In this sense, input uncertainty is only accounted for in the prediction stage. The addressing of the uncertainty in the training data when building a PLS model falls into the research area of errors-in-variables (EIV) (Faber and Kowalski[8]) and is outwith the scope of this paper. Two approaches are presented, Monte Carlo approximation and an analytical method, to calculate the predictive mean and intervals for a PLS model where input uncertainty is taken into consideration. The proposed approaches are demonstrated in terms of their ability to predict the end-of-batch quality, number average molecular weight, of a simulated polymerization process. Although the two approaches give similar predictions, the analytical method is recommended for practical application due to its computational efficiency.

2. PLS Regression and Prediction Intervals

PLS is a technique to estimate the regression vector, $\mathbf{b}$, in a general linear regression model:

$$y = \mathbf{b}^{\mathrm{T}}\mathbf{x} + \varepsilon \tag{1}$$

where $\mathbf{x}$ is a vector of *p* input variables, *y* is the output and ε is a residual term. Whilst PLS regression can be applied to multivariate output variables, this paper considers the univariate case to ensure the derivation of the predictive distribution for a particular input, associated with uncertainty, does not appear to be more complicated than in practice due to the complexity of the notation. In addition the data is assumed to be mean centred.

A major advantage of PLS over ordinary least squares is that it can provide a reliable and stable estimate of **b** in the presence of multi-collinearity in the input variables, or when *p* is smaller than the number of training data points, *N*. This is achieved by extracting a small number of *latent variables* that are a linear combination of the original input variables. The latent variables are then related to the output. The calculation of the latent variables and the regression vector **b** can be realized through either the nonlinear iterative partial least squares (NIPALS) algorithm (Geladi and Kowalski[1]) or the Helland algorithm (Helland[14]).

Through local linearization, the predictive distribution, $p(y^* \mid \mathbf{b}, \mathbf{x}^*)$, for a new input vector, $\mathbf{x}^*$, can be approximated by a normal distribution with mean and variance (Denham[7]):

$$E(y^*) = \mathbf{b}^{\mathrm{T}}\mathbf{x}^* \tag{2}$$

and

$$\mathrm{var}(y^*) = \sigma^2\left[1 + \frac{1}{N} + \mathbf{x}^{*\mathrm{T}}\frac{\partial \mathbf{b}}{\partial \mathbf{y}}\left(\frac{\partial \mathbf{b}}{\partial \mathbf{y}}\right)^{\mathrm{T}}\mathbf{x}^*\right] \tag{3}$$

where σ^2 is the variance of the regression residuals. The $100(1-\alpha)\%$ prediction interval for PLS is then given by:

$$E(y^*) \pm t_{\alpha/2,\mathrm{df}}\sqrt{\mathrm{var}(y^*)} \tag{4}$$

where $t_{\alpha/2,\mathrm{df}}$ is the critical value obtained from a Student *t*-distribution with degrees of freedom (df). The calculation of σ^2, $\partial\mathbf{b}/\partial\mathbf{y}$ and the degrees of freedom are given in (Denham[7]; Serneels *et al.*[9]).

3. Prediction with Input Uncertainty

Now assume a prediction is to be made at an uncertain input $\mathbf{x}^*$ with normal distribution: $\mathbf{x}^* \sim N(\mathbf{m},\mathbf{S})$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$. Incorporating input uncertainty, the resulting predictive mean and variance are given by (Girard *et al.*[13]):

$$m(\mathbf{m},\mathbf{S}) = E_{x^*}\left[E(y^*)\right] = \int E(y^*)p(\mathbf{x}^* \mid \mathbf{m},\mathbf{S})d\mathbf{x}^* \tag{5}$$

and

$$\begin{aligned} v(\mathbf{m},\mathbf{S}) &= E_{x^*}\left[\mathrm{var}(y^*)\right] + \mathrm{var}_{x^*}\left[E(y^*)\right] = E_{x^*}\left[\sigma^2(y^*)\right] + E_{x^*}\left[E(y^*)^2\right] - E^2_{x^*}\left[E(y^*)\right] \\ &= \int \mathrm{var}(y^*)p(\mathbf{x}^* \mid \mathbf{m},\mathbf{S})d\mathbf{x} + \int E(y^*)^2 p(\mathbf{x}^* \mid \mathbf{m},\mathbf{S})d\mathbf{x} - \left(\int E(y^*)p(\mathbf{x}^* \mid \mathbf{m},\mathbf{S})d\mathbf{x}\right)^2 \end{aligned} \tag{6}$$

The subsequent two sub-sections describe how the integrals in Eqs. (5) and (6) can be calculated using either a Monte Carlo approximation or an analytical method.

3.1. Monte Carlo Approximation

The basic idea of Monte Carlo simulation is to draw *K* random samples from the input distribution: $\mathbf{x}_i^* \sim p(\mathbf{x}^* \mid \mathbf{m},\mathbf{S}), i = 1,\ldots,K$, and then to approximate the integrals in Eqs. (5) and (6) as follows. For the mean:

$$m(\mathbf{m},\mathbf{S}) \approx \frac{1}{K}\sum_{i=1}^{K}\mathbf{b}^{\mathrm{T}}\mathbf{x}_i^* \tag{7}$$

With the variance is given by:

$$v(\mathbf{m},\mathbf{S}) \approx \frac{1}{K}\sum_{k=1}^{K}\sigma^2\left[1 + \frac{1}{N} + \mathbf{x}_i^{*\mathrm{T}}\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\mathbf{x}_i^*\right] + \frac{1}{K}\sum_{k=1}^{K}(\mathbf{b}^{\mathrm{T}}\mathbf{x}_i^*)^2 - \left(\frac{1}{K}\sum_{i=1}^{K}\mathbf{b}^{\mathrm{T}}\mathbf{x}_i^*\right)^2 \tag{8}$$

The generation of Monte Carlo samples from a multivariate normal distribution is straightforward and available in many statistical software packages including Matlab and S-Plus. However Monte Carlo simulation requires a large number of samples to be drawn from the underlying distribution to approximate the integrals accurately, and thus incurs significant computational costs.

3.2. Analytical Method

Alternatively an analytical method can be derived to calculate the integrals in Eqs. (5) and (6). The predictive mean can be calculated as:

$$m(\mathbf{m},\mathbf{S}) = \mathbf{b}^{\mathrm{T}}\int \mathbf{x}^* p(\mathbf{x}^* \mid \mathbf{m},\mathbf{S})d\mathbf{x}^* = \mathbf{b}^{\mathrm{T}}\mathbf{m} \tag{9}$$

It can be seen that the predictive mean is calculated simply by replacing the input $\mathbf{x}^*$ with its mean vector $\mathbf{m}$.

The derivation of the predictive variance is more complex. The three integration terms in Eq. (6) are derived one by one. Firstly,

$$\int \mathrm{var}(y^*)p(\mathbf{x}^*\mid\mathbf{m},\mathbf{S})d\mathbf{x} = \int \sigma^2\left[1+\frac{1}{N}+\mathbf{x}^{*\mathrm{T}}\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\mathbf{x}^*\right]p(\mathbf{x}^*\mid\mathbf{m},\mathbf{S})d\mathbf{x}^*$$
$$= \sigma^2\left[1+\frac{1}{N}+\int \mathbf{x}^{*\mathrm{T}}\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\mathbf{x}^* p(\mathbf{x}^*\mid\mathbf{m},\mathbf{S})d\mathbf{x}^*\right] \tag{10}$$

By noting that $\mathbf{x}^{*\mathrm{T}}(\partial\mathbf{b}/\partial\mathbf{y})(\partial\mathbf{b}/\partial\mathbf{y})^{\mathrm{T}}\mathbf{x}^*$ is a scalar and thus is equal to $\mathrm{Tr}((\partial\mathbf{b}/\partial\mathbf{y})^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{*\mathrm{T}}(\partial\mathbf{b}/\partial\mathbf{y}))$, where Tr(.) is the trace of a square matrix, Eq. (10) can be simplified to give:

$$\int \mathrm{var}(y^*)p(\mathbf{x}^*\mid\mathbf{m},\mathbf{S})d\mathbf{x} = \sigma^2\left[1+\frac{1}{N}+\int \mathrm{Tr}\left(\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\mathbf{x}^*\mathbf{x}^{*\mathrm{T}}\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)p(\mathbf{x}^*\mid\mathbf{m},\mathbf{S})d\mathbf{x}^*\right]$$
$$= \sigma^2\left[1+\frac{1}{N}+\mathrm{Tr}\left(\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\int \mathbf{x}^*\mathbf{x}^{*\mathrm{T}}p(\mathbf{x}^*\mid\mathbf{m},\mathbf{S})d\mathbf{x}^*\cdot\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)\right]$$
$$= \sigma^2\left[1+\frac{1}{N}+\mathrm{Tr}\left(\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}(\mathbf{S}+\mathbf{m}\mathbf{m}^{\mathrm{T}})\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)\right] \tag{11}$$

The other two terms in Eq. (6) can be obtained as follows:

$$\int E\left(y^*\right)^2 p\left(\mathbf{x}^*\mid\mathbf{m},\mathbf{S}\right)dx = \int \mathbf{b}^T\mathbf{x}^*\mathbf{x}^{*T}\mathbf{b}\; p\left(\mathbf{x}^*\mid\mathbf{m},\mathbf{S}\right)dx^*$$
$$= \mathbf{b}^T\int \mathbf{x}^*\mathbf{x}^{*T}\; p\left(\mathbf{x}^*\mid\mathbf{m},\mathbf{S}\right)dx^*\mathbf{b} = \mathbf{b}^T\left(\mathbf{S}+\mathbf{m}\mathbf{m}^{\mathbf{T}}\right)\mathbf{b} \tag{12}$$

and

$$\left(\int E\left(y^*\right)p\left(\mathbf{x}^*\mid\mathbf{m},\mathbf{S}\right)dx\right)^2 = \left(\mathbf{b}^T\mathbf{m}\right)^2 = \mathbf{b}^T\mathbf{m}\mathbf{m}^T\mathbf{b} \tag{13}$$

Therefore by combining Eqs. (11), (12) and (13) the predictive variance is given by:

$$v(\mathbf{m},\mathbf{S}) = \sigma^2\left[1+\frac{1}{N}+\mathbf{m}^{\mathrm{T}}\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\mathbf{m}\right]+\left[\sigma^2\mathrm{tr}\left(\left(\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)^{\mathrm{T}}\mathbf{S}\frac{\partial\mathbf{b}}{\partial\mathbf{y}}\right)+\mathbf{b}^{\mathrm{T}}\mathbf{S}\mathbf{b}\right] \tag{14}$$

Eq. (14) shows that the predictive variance can be decomposed into two parts. The first part is the variance associated with model uncertainty and the second term relates to the uncertainty associated with the inputs.

4. Case Study: The Prediction of An End-Of-Batch Quality Variable

The proposed approaches are demonstrated in terms of the on-line prediction of the final quality variables of a batch process utilising a PLS model. This application scenario was discussed in (Nomikos and MacGregor [3]) under the framework of missing data, and serves as a test-bed to show how the general approaches developed in the paper can be applied to this specific problem. The on-line inference of final product quality is critical to ensure the consistent production of high quality chemicals, and forms the foundation of process performance monitoring and control schemes (Nomikos and MacGregor [3]; Pan and Lee [15]).

4.1. Missing Data

In the model development stage, a multi-way PLS model is developed by unfolding the three-way batch data ($N_{\mathrm{batch}}\times N_{\mathrm{variable}}\times N_{\mathrm{time}}$) into a large two-dimensional matrix

($N_{\text{batch}} \times N_{\text{variable}} N_{\text{time}}$) according to Nomikos and MacGregor[3]. The difficulty with the on-line inference of the final value of the quality variable is that the prediction must be made before the batch is complete. Consequently on-line measurements are only available up to the current time point, and the data from the next time step to the batch finishing time is missing. The conditional distribution of the missing data, given the available measurement, can be obtained as follows (Nelson *et al.* [11]).

For ease of derivation, the process measurement can be split as $\mathbf{x}^* = [\mathbf{x}_1^{*\mathrm{T}}, \mathbf{x}_2^{*\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ are sub-vectors of the observed and missing data respectively. Assuming $\mathbf{x}^*$ is normally distributed with the mean, m, and the covariance matrix, S, being estimated from historical batch data. Likewise the mean and covariance are also divided into two blocks pertaining to the observed and missing data:

$$\mathbf{m} = [\mathbf{m}_1^{\mathrm{T}}, \mathbf{m}_2^{\mathrm{T}}]^{\mathrm{T}}; \qquad \mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \tag{14}$$

According to the properties of the normal distribution, when $\mathbf{x}_1^*$ is observed (and thus is not random), the conditional distribution of the missing data $\mathbf{x}_2^*$ given $\mathbf{x}_1^*$, $p(\mathbf{x}_2^* \mid \mathbf{x}_1^*)$, is also a normal distribution with mean given by $E(\mathbf{x}_2^* \mid \mathbf{x}_1^*) = \mathbf{m}_2 + \mathbf{S}_{21}\mathbf{S}_{11}^{-1}(\mathbf{x}_1^* - \mathbf{m}_1)$. The corresponding covariance is $\mathrm{cov}(\mathbf{x}_2^* \mid \mathbf{x}_1^*) = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$.

In summary the input variables have a normal distribution with mean $[\mathbf{x}_1^{*\mathrm{T}}, E(\mathbf{x}_2^* \mid \mathbf{x}_1^*)^{\mathrm{T}}]^{\mathrm{T}}$ and covariance matrix:

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{cov}(\mathbf{x}_2^* \mid \mathbf{x}_1^*) \end{bmatrix} \tag{15}$$

where since $\mathbf{x}_1^*$ is an observed non-random vector, its auto-covariance and cross-covariance with $\mathbf{x}_2^*$ are zero sub-matrices. Based on the distribution of the input variables, the prediction for the PLS model can be obtained using either a Monte Carlo simulation or the analytical method, as described in Section 3.

4.2. Batch Polymerization Process

The process shown in Figure 1 is a free-radical batch polymerization of methyl-methacrylate (MMA) with a water solvent and benzoyl peroxide initiator (Chen *et al.* [16]; Kiparissides *et al.*[17]). The polymerization temperature is maintained .by a cascade control system through manipulating the flow rates of the hot and cold water streams. A detailed mathematical model of the process is described in (Kiparissides *et al.*[17]), and it forms the basis for the development of a simulation program. The detailed settings for the simulation, such as process kinetic parameters, physical properties, reactor operating conditions, and controller configurations, can be found in (Kiparissides *et al.* [17]). In this study, it is assumed that on-line measurements are available every two minutes for monomer conversion, jacket inlet and outlet temperature, reactor temperature and coolant flow rate. The measurements are corrupted by white noise with a signal-to-noise ratio of 20 db. The batch duration is 120 minutes.

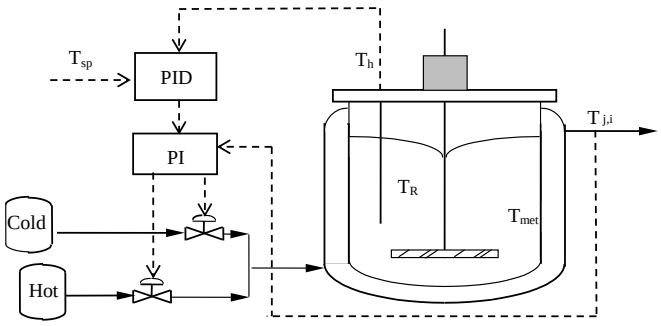


Figure 1. Batch Polymerization Reactor

Batch variations were simulated through introducing a disturbance into the initial condition of the batch, the initial initiator weight (I_0):

$$I_0 = 2.5 + \tau \tag{16}$$

where 2.5 (g) is the canonical value of I_0, and τ is a zero mean normal random noise with standard deviation of 0.05.

The objective of this study is to infer an important end-batch quality variable, number average molecular weight (Mn), from online process measurements. A multi-way PLS model is developed based on 50 training batches. Additional data from one unseen batch was utilized to test the on-line prediction of the end-of-batch number average molecular weight.

4.3. Results and Discussions

Figure 2 gives the prediction results of the end-of-batch number average molecular weight, Mn, at different time points throughout the duration of the batch. Prediction intervals calculated using both Monte Carlo approximation (K=1000 random samples) and the analytical method are presented. In the initial stage of the batch, *i.e.* up to time 40 minutes, the predictive mean is not satisfactory, and the prediction intervals are relatively large, reflecting a significant amount of uncertainty in the prediction. This predictive uncertainty is due to the limited data available initially. However the prediction intervals tend to decrease over time when more on-line measurements are available. Notably after 40 minutes, the predictive mean appears to converge to the target average molecular weight, Mn, and the magnitude of the prediction intervals tends to stabilize. Figure 2 also verifies that the prediction intervals obtained from the analytical method are approximately the same as those calculated using the Monte Carlo approximation approach. The analytical method requires significantly less computational time (47 seconds, Matlab implementation with Intel i5 3.0 GHz computer running under Windows 7) than the Monte Carlo approximation (2703 seconds), and thus the analytical method should be adopted in practice.

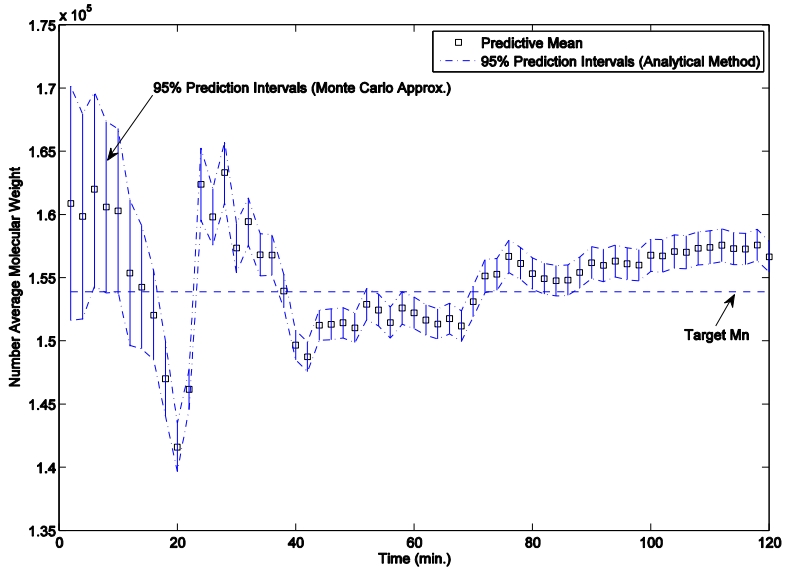


Figure 2 Prediction of the End-of-Batch Number Average Molecular Weight (Mn). The 95% Prediction Intervals Calculated Using Monte Carlo Approximation are Illustrated by Vertical Lines

Figure 3 illustrates the predictive variance throughout the duration of the batch. More specifically the variance is decomposed into input uncertainty and model uncertainty according to Eq. (14). Initially the variance associated with the input uncertainty is dominant, since a considerable proportion of future measurements have to be estimated (Eq. (15)). These contribute a significant amount of uncertainty to the PLS prediction. As the batch processing proceeds over time and more on-line data becomes available, fewer future measurements require to be estimated, and thus the input uncertainty decreases. According to Figure 3, after 20 minutes the input uncertainty is negligible when compared with model uncertainty. This phenomenon reflects the fact that when estimating missing data $\mathbf{x}_2^*$ from $\mathbf{x}_1^*$, the covariance matrix, $\text{cov}(\mathbf{x}_2^* \mid \mathbf{x}_1^*)$, is close to zero. In other words, the available on-line data are capable of determining the future batch measurements with high precision. High precision, a measure of predictive uncertainty, should not be confused with high accuracy, a measure of predictive performance.

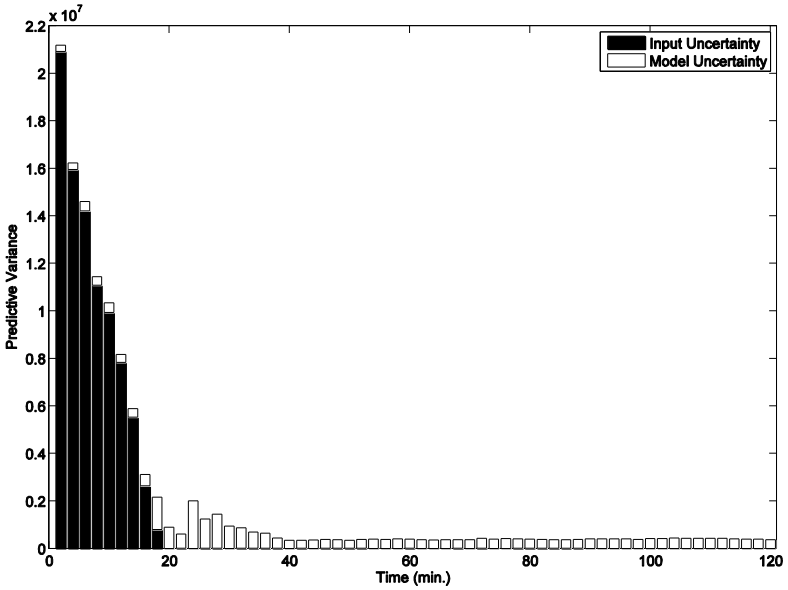


Figure 3. Predictive Variance Pertaining to Input and Model Uncertainty over Time

Finally, Figure 3 shows that after approximately 40 minutes, the predictive variance associated with model uncertainty stabilises, indicating that convergence of the prediction is achieved.

5. Conclusions

This paper presented two methods to calculate the PLS predictions in the presence of uncertain inputs: Monte Carlo approximation and an analytical method. The proposed approaches were demonstrated for the prediction of the end-of-batch quality parameter, number average molecular weight, of a simulated polymerization process. The results from Monte Carlo approximation and the analytical method are consistent. In practice the analytical method is more attractive since it incurs significantly lower computational costs than the Monte Carlo approximation approach. However, the Monte Carlo method is easier to implement, and thus it can be utilized to verify the validity of the analytical approach. More importantly, if the input vector does not conform to a normal distribution, it may be difficult to calculate the predictive distribution analytically. In this situation, Monte Carlo approximation may be a good option.

The proposed algorithm can be applied to a wide variety of problems where the uncertainty of the input variables requires to be addressed in the modelling procedure. Two examples are where predictions require to be made in the presence of missing data and in the implementation of recursive multi-step ahead predictions in time series modelling.

Acknowledgments

The works supported by Research Fund for the Doctoral Program of Higher Education of China(RFDP) 20130191120058.

References

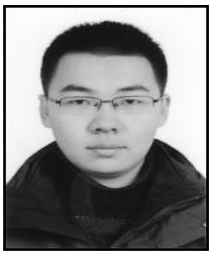
[1] P. Geladi and B. Kowalski, "Partial least-squares regression: a tutorial", Analytica Chimica Acta, vol. 185, no. 1, **(1986)**, pp 1-17.
[2] P. R. C. Nelson, P. A. Taylor and J. F. MacGregor, "Missing data methods in PCA and PLS: score calculations with incomplete observations", Chemometrics and Intelligent Laboratory Systems, vol. 35, no. 1, **(1996)**, pp. 45-65.
[3] P. Nomikos and J. F. MacGregor, "Multi-way partial least squares in monitoring batch processes", Chemometrics and Intelligent Laboratory Systems, vol. 30, no. 1, **(1995)**, pp. 97-108.
[4] T. Mehmood, K. H. Liland, L. Snipen and S. Sæbø, "A review of variable selection methods in Partial Least Squares Regression", Chemometrics and Intelligent Laboratory Systems, vol. 118, no. 1, **(2012)**, pp. 62-69.
[5] J. Liu, "Developing a soft sensor based on sparse partial least squares with variable selection", Journal of Process Control, vol. 24, no. 7, **(2014)**, pp. 1046-1056.
[6] J. Dong, K. Zhang, Y. Huang, G. Li and K. Peng, "Adaptive total PLS based quality-relevant process monitoring with application to the Tennessee Eastman process", Neurocomputing vol. 154, **(2015)**, pp. 77-85.
[7] M. C. Denham, "Prediction intervals in partial least squares", Journal of Chemometrics, vol. 11, no. 1, **(1997)**, pp. 39-52.
[8] K. Faber and B. R. Kowalski, "Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares", Journal of Chemometrics, vol. 11, no. 3, **(1997)**, pp. 181-238.
[9] S. Serneels, P. Lemberge and P. J. van Espen, "Calculation of PLS prediction intervals using efficient recursive relations for the Jacobian matrix", Journal of Chemometrics, vol. 18, no. 2, **(2004)**, pp 76-80.
[10] B. Lu, I. Castillo, L. Chiang and T. F. Edgar, "Industrial PLS model variable selection using moving window variable importance in projection", Chemometrics and Intelligent Laboratory Systems, vol. 135, **(2014)**, pp. 90-109.
[11] P. R. C. Nelson, J. F. MacGregor and P. A. Taylor, "The impact of missing measurements on PCA and PLS prediction and monitoring applications", Chemometrics and Intelligent Laboratory Systems, vol. 80, no. 1, **(2006)**, pp. 1-12.
[12] T. K. Dijkstra and J. Henseler, "Consistent and asymptotically normal PLS estimators for linear structural equations", Computational Statistics & Data Analysis, vol. 81, no. 1, **(2015)**, pp. 10-23.
[13] A. Girard, C. E. Rasmussen, J. Quiñonero-Candela and R. Murray-Smith, "Multiple-step ahead prediction for non-linear dynamic systems - a Gaussian process treatment with propagation of the uncertainty", in Advances in Neural Information Processing Systems 15 (Becker, S., S. Thrun and K. Obermayer (Ed)), MIT Press, Cambridge, Massachusetts, **(2003)**, pp. 529-536.
[14] I. S. Helland, "On the structure of partial least squares regression", Communications in statistics -- Simulation and computation, vol. 17, no. 2, **(1988)**, pp. 581-607.
[15] Y. Pan, and J. H. Lee, "Recursive data-based prediction and control of product quality for a PMMA batch process", Chemical Engineering Science, vol. 58, no. 14, **(2003)**, pp. 3215-3221.
[16] T. Chen, J. Morris and E. Martin, "Particle filters for state and parameter estimation in batch processes", Journal of Process Control, vol. 15, no. 6, **(2005)**, pp. 665-673.
[17] C. Kiparissides, P. Seferlis, G. Mourikas and A. J. Morris, "Online optimizing control of molecular weight properties in batch free-radical polymerization reactors", Industrial and Engineering Chemistry Research, vol. 41, no. 24, **(2002)**, pp. 6120-6131.

Authors

Zehui Chen, born in 1969, received the master's degree on Technological Economy from School of Economics and Business Administration, Chongqing University, Chongqing, China. His research interests are in the areas of the management in energy system, in particular for energy management based on analytics.

Weiguo Zhang, born in 1965, received the doctor's degree on enterprise management & technological economy from School of Economics and Business Administration, Chongqing University, Chongqing, China. Currently, he is a professor at School of Economics and Business Administration, Chongqing University, Chongqing, China.

Jianghong Ren, received the B.E., M.E. and Ph.D. degree in 1999, 2004 and 2011 from College of Automation, Chongqing University, Chongqing, China. He is teaching at College of Automation, Chongqing University, Chongqing, China. His research interests are in the areas of the analytics in energy system, in particular for energy conservation in building and industrial process.



Shaowei Gong, born in 1989, received the Master of Communication from School of Journalism and Communication, Chongqing University, Chongqing, China. His research interests are in Enterprise management and innovation policy.