LOCAL LINEAR AND NON-LINEAR MULTI-WAY PARTIAL LEAST SQUARES BATCH MODELLING

N.M. Fletcher, A.J. Morris and E. B. Martin

Centre for Process Analytics and Control Technology,
University of Newcastle, Newcastle upon Tyne, England, NE1 7RU, UK

Abstract: An industrial fed-batch fermentation process forms the basis of a study that illustrates the development and use of local models for the construction of a performance monitoring scheme. Linear models are initially built for different operating regions within the batch using Dynamic Partial Least Squares (DPLS) to capture the underlying process dynamics. The local models are then pieced together and a global model developed based on the latent variable scores. A comparison with a global non-linear quadratic DPLS model is then made prior to using the latent variable scores from the local models to build a nominal process representation. Copyright 2002 IFAC.

Keywords: Dynamic Modelling, Local structures, Batch Control

1. INTRODUCTION

It is recognised that through the development of parsimonious models better understanding and control of a process can be achieved resulting in enhanced production and increased product yield. Traditionally batch control has been based on detailed process scheduling. Complementary to this approach are those technologies that are based on the multivariate statistical techniques of Principal Component Analysis (PCA) and Partial Least Squares (PLS). Such techniques aid not only the prediction of product quality but also the identification of the onset of changes in process operation and process faults. These elements are key to the achievement of consistent high quality production.

The linear multivariate statistical batch monitoring techniques of Multi-way PCA (MPCA) and Multi-way PLS (MPLS) (Nomikos and MacGregor, 1994, 1995) have been used for the modelling and monitoring of processes that give rise to collinear and/or highly correlated variables. However, in practice batch processes are synonymous with non-linear, time-variant behaviour. These properties challenge such monitoring approaches and question the applicability of the multi-way techniques for the modelling and monitoring of batch processes. One solution is through the extension of PLS to encapsulate the dynamic non-linear behaviour. Kaspar and Ray (1993) extended PLS to dynamic PLS whilst Baffi et al. (2000) developed a non-linear dynamic PLS algorithm.

An alternative approach to modelling dynamic non-linear systems is to sub-divide the batch process variable trajectories into a number of distinct operating regions. In this situation a local linear model can be fitted to each separate region. These individual models can then be pieced together, thereby providing an overall non-linear global model (e.g. Foss et al., 1995). This piecing together of the local models can be achieved through a number of means including validity functions and fuzzy membership functions (e.g. Murray-Smith and Johansen, 1997). These functions are then used to determine which local model is the most suitable for modelling a particular observation in the batch trajectory. Such a local model based structure realises a novel approach to batch process
monitoring. The methodology is demonstrated by application to an industrial fed-batch process.

2. PARTIAL LEAST SQUARES (PLS)

PLS is a multivariate regression method that maximises the correlation between the process variables ($X$) and the quality variables ($Y$). (Geladi and Kowalski, 1986). The objective of linear PLS is to project the data down onto a number of latent variables, say $t_j$ and $u_j$, and to then develop a regression model, termed the inner relationship, between the latent variables:

$$ u_j = b_j t_j + e_j \quad (j = 1, \ldots, A) \quad (1) $$

where $b_j$ are the coefficients of the model and $A$ is the number of latent variables defined, for example, from cross-validation (Wold, 1978). The matrices $X$ and $Y$ are decomposed as the sum of the outer products of the latent variables, $t_j$, and the loadings, $p_j$, and the prediction $\hat{u}_j$ of $u_j$, where $\hat{u}_j = \hat{b}_j t_j$, and the loadings, $q_j$, respectively:

$$ X = \sum_{j=1}^A t_j p_j + E \quad (2) $$

$$ Y = \sum_{j=1}^A \hat{u}_j q_j + F \quad (3) $$

where $E$ and $F$ are the residual matrices for the matrix decomposition of $X$ and $Y$.

However, when modelling complex chemical and biochemical systems, the underlying behaviour may be non-linear and hence linear PLS cannot reliably be used to model the underlying structure. A number of algorithms have been proposed to integrate non-linear features within the linear PLS framework. Wold et al. (1989) proposed a polynomial (quadratic) PLS algorithm that retains the framework of linear PLS but which modifies the inner relationship:

$$ u_j = c_{0j} + c_{1j} t_j + c_{2j} t_j^2 + e_j \quad (4) $$

This seminal work of Wold et al. (1989) has led to a variety of non-linear PLS algorithms that attempt to capture the non-linear features by using a non-linear mapping in the inner relationship. Baffi et al. (1999) provided an overview of the different approaches and proposed a modified algorithm for the updating of the inner and outer weights, error-based PLS. It is this algorithm that forms the basis of this study.

PLS can be extended to handle three-dimensional data, Multi-way PLS (MPLS). Consider the case where the data consists of $I$ batches, where $J$ process variables and $L$ quality variables are measured over $K$ and $M$ time intervals, respectively, throughout the duration of the batch. This information can be arranged into two three-way data matrices $X$ ($I \times J \times K$) and $Y$ ($I \times L \times M$). Adopting a bi-linear approach to the analysis, the data arrays are unfolded into two-way arrays. There are three possible ways to unfold three-way matrices. In this study, the batch observation level approach of Wold, et al. (1998) is adopted. The batches are stacked in a vertical manner, $X$ ($I \times J$) $\times K$ and $Y$ ($I \times L$) $\times M$ and these data matrices are then decomposed into a series of latent variables comprising score matrices and loading vectors, plus residual matrices $E$ and $F$ (Eq. 2 and 3). These decompositions are consistent with the principles of PLS, with each score vector corresponding to the evolution of a batch with time.

2.1 Dynamic PLS for Multi-way Analysis.

Data collected on chemical and biochemical processes typically exhibits both serial and cross correlation. This is as a consequence of both the dynamic nature of the process and its inherent state interactions. One approach to modelling serially correlated data is through time series representations such as Finite Impulse Response (FIR) or Auto-Regressive with eXogenous inputs (ARX) models. An ARX model is represented by:

$$ y(k) = \sum_{i=1}^{n_y} A_i y(k-i) + \sum_{j=1}^{n_u} B_j u(k-1) + w(k) \quad (5) $$

where $u(k)$ and $y(k)$ are the process input and output variables at time point $k$, respectively, $n_y$ and $n_u$ are the number of process input and output variables respectively and $w(k)$ is a noise vector. $A$ and $B$ are coefficient matrices calculated using least squares. Adopting an ARX structure through the inclusion of time-lagged process measurements and quality variables results in the dynamic features of the process being encapsulated within the model.

The main drawback of a least squares approach to dynamic empirical modelling is related to the sensitivity of the least squares solution to data matrices incorporating collinear or highly correlated variables. To address this issue, dynamic PLS (DPLS) algorithms have been proposed. This is achieved through the incorporation of the ARX structure defined in Equation (5) within Equation (2). The data matrix $X$ in Equation (2) is replaced by:

$$ X = \begin{bmatrix} y_1(k-1) & u_1(k-1) & \cdots & y_1(k-n) & u_1(k-n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_n(k-1) & u_n(k-1) & \cdots & y_n(k-n) & u_n(k-n) \end{bmatrix} \quad (6) $$

This matrix $X$ is then used in the DPLS algorithm to form the basis of the analysis.
where \( y \) is a process output, \( u \) is a process input, \( n \) is the number of lags, \( k \) is the observation number and \( m \) is the number of time points in the batch. This dynamic structure can also be used as the basis of a dynamic non-linear PLS algorithm (Baffi et al. 2000).

3. LOCAL MODELS

By partitioning the batch variable trajectories into a number of operating regions where linear approximations apply, a non-linear global model can be approximated through the piecing together of the local models. The models developed for the local operating regions are valid when the process operates under specific conditions, gradually becoming invalid as the process moves outside of that region. A validity function vector, \( \rho_{op} \) (0 ≤ \( \rho \) ≤ 1) defines the weight for a specific operating region, \( op \), at each time point, \( \phi \), throughout the duration of the batch. In this initial study, linear interpolation was used to define the transition between the operating regions (Foss et al., 1995):

\[
\omega_{op}(\phi) = \frac{\rho_{op}(\phi)}{\sum_{op=1}^{N} \rho_{op}(\phi)}
\]

(7)

where \( \sum_{op=1}^{N} \omega_{op}(\phi) = 1 \) \( \forall \phi \)

\( N \) is the number of operating regions, \( op \), within the batch process. A nominal PLS model is then built for each local region using cross-validation to select the number of latent variables for each local model. The scores and output predictions calculated for each local model are then joined using the interpolation function, \( \omega_{op} \), to create a global model.

4. INDUSTRIAL FERMENTATION PROCESS

Data from a fed-batch fermentation process is used to demonstrate the local model approach to process modelling and process monitoring. The process is well established and has been running for many years. Although it is operated under highly scheduled operating conditions that have been finely tuned, fluctuations can still be observed in the yield of the product. A large number of process variables are recorded regularly during each batch run. Following earlier work, Fletcher et al. (2001), a reduced number of process variables were selected for the analysis, Table 1. The output variable is Potency. Fifteen batches were used to build the nominal models with five unseen batches being used for validation. All the batches were representative of standard batch runs with some fluctuation in yield but with no large deviations from the optimum operating conditions.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Potency</td>
<td>y1</td>
</tr>
<tr>
<td>pH</td>
<td>x1</td>
</tr>
<tr>
<td>Dissolved oxygen</td>
<td>x2</td>
</tr>
<tr>
<td>Power</td>
<td>x3</td>
</tr>
<tr>
<td>( O_2 ) Respiration rate</td>
<td>x4</td>
</tr>
<tr>
<td>( CO_2 ) evolution rate</td>
<td>x5</td>
</tr>
<tr>
<td>Substrate 1 addition rate</td>
<td>x6</td>
</tr>
<tr>
<td>Chemical A addition rate</td>
<td>x7</td>
</tr>
<tr>
<td>Residual substrate 1</td>
<td>x8</td>
</tr>
<tr>
<td>Residual chemical B</td>
<td>x9</td>
</tr>
<tr>
<td>Residual chemical A</td>
<td>x10</td>
</tr>
</tbody>
</table>

4.1 Operating Region Specification

The operating regions are defined based on knowledge of the different biological phases that occur during the fermentation, and the operating conditions that are applied to control the process. A precursor is added to the fermenter to initialise the reactions. This marks the transition between the first stage, initialisation, and the second stage, initial growth rate. Maximum growth rate and product completion form the third and fourth stages respectively. Three variable trajectories are shown in Fig. 1 with the operating regions identified. As can be seen, the transition from one operating region to another tends to be rather blurred. Furthermore as a result of the variability within individual fermentations, considerable overlap between regions can occur.

4.2 Dynamic Specification

The next stage is to select the number of lags for each variable, within each operating region model. In a complex fermentation, the dynamics may change between the various operating regions, consequently the number of lags included within the augmented data set will differ. To determine the number of lags for each variable, an ARX model was constructed and used to identify the weighting of each lagged variable. Fig 2 shows the ARX model variable weightings for \( CO_2 \) Evolution Rate (CER). At least 4
lags should be included in the augmented data matrix since after time lag 3 the weightings are seen to become significant in comparison to lower order time lags.

Fig. 2. Lagged variable weighting for CER.

A principal component loadings plot, Fig 3, shows the significance of each lagged variable within the PLS framework for operating region 1, using latent variables 1 and 2. The loading variables, 12 to 15, relating to the current and 3 lagged values for the variable, CO₂ evolution rate, can be seen to be significant, as the bivariate loadings do not lie close to the origin. This is further evidence to retain them in the analysis.

Fig. 3. Bivariate loadings of latent variables 1and 2.

Linear interpolation was then performed between the scores from each local dynamic PLS model, using the interpolation function, \( \omega_{op} \), to give a global non-linear model.

5. COMPARISON OF PREDICTION METHODS

The global model constructed from the local dynamic models was compared with a global non-linear dynamic PLS model.

5.1 Prediction using Local DPLS Models

Cross validation, using a random block approach, was used to select the number of latent variables to include in each local dynamic PLS model, Fig. 4. This value was confirmed by analysing the regression coefficients of the latent variables and their associated standard errors:

\[
se_{t_i} = \pm \frac{s_u}{\sqrt{\sum_{i=1}^{n} (t_i - \bar{t})^2}}
\]

where \( s_u \) is the standard deviation of the scores for each latent variable, \( t_i \) is the score at each observation point and \( \bar{t} \) is the mean of the scores. Fig. 5 shows the regression coefficients \( \pm \) the standard error. Zero is included in the interval of the lower order latent variables indicating that these may be removed from the analysis. From cross validation, nine latent variables were selected (Fig. 4) for inclusion in operating region three of the local DPLS model. This was confirmed from examination of the error bars of the regression coefficients, i.e. the first nine latent variables do not include zero (Fig. 5).

Fig. 4. Cross validation of nominal batches.

Fig. 5. Regression coefficients with error bars.

Using the interpolation function to move between the different operating regions as the batch progresses with time, the nominal model for each local operating region was used to predict the potency variable for each of the validation batches. The Residual Sum of Squares (RSS) for the prediction of potency was then used as a measure of prediction capability (Table 2).

<table>
<thead>
<tr>
<th>Model</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local DPLS</td>
<td>0.47</td>
</tr>
<tr>
<td>Global NL Quadratic-DPLS</td>
<td>48.31</td>
</tr>
</tbody>
</table>

The plot of actual and predicted values of potency (Fig. 6) indicates that the local models for regions 1, 3 and 4 predict well. Model 2 appears the poorest with some noise seen in the predictions. This is a consequence of the phase changes that take place, during this period, as the fermentation progresses. The residual versus predicted plot, Fig. 7, indicates the presence of structure. However, within each local operating region, the residuals, represented by different symbols, can be seen to be unstructured for models 1, 2 and 4 with some structure remaining in the residuals of model 3. It is concluded that the process non-linearities and dynamics have been adequately captured through the local models.
5.2 Prediction using the NL Quadratic DPLS Model

Non-linear (quadratic) dynamic PLS was then applied to the fifteen nominal batches to model the global trajectory. The number of lags for each variable were selected for the dynamic global model using the ARX model approach, as described in section 5.1. Four latent variables were selected using block cross validation and a plot of the standard error. The model was then applied to the five validation batches. Fig. 8 shows the prediction of potency for one of the validation batches. It is clear that the performance of this model is poor in comparison with the local DPLS model (Fig. 6). The highly structured nature of the plot of residuals versus predicted values of potency (Fig. 9) shows that the dynamic non-linear structure of the process has not been captured by the global model. This is confirmed from the residuals sum of squares, Table 2 that is seen to be significantly higher than when a local model method is used.

6. MONITORING TOOLS

Comparison of the two modelling approaches showed that the local model based method was more able to capture the process dynamics and non-linear behaviour and hence to predict potency. The next step was to use the nominal scores calculated from the local models as the basis of a process performance monitoring tool. Latent variable scores plots, the Squared Prediction Error (SPE) and Hotelling’s $T^2$ trajectory plots were developed from each of the local DPLS models.

Based on the batch scores, action limits were calculated using ±3 standard deviations of the nominal scores at each observation point and then interpolating between the different limits using $\omega_{op}$. Fig. 10 shows the scores monitoring chart for the five validation batches for latent variable 1. Validation batch 1 can be seen to move outside of the limits near the beginning of process operation, between observations 21 and 62, prior to moving back within the limits for the remaining duration of the batch.

An SPE plot can also be used for monitoring purposes. Fig. 11 illustrates the SPE trajectory for each of the validation batches. Action limits were calculated for each local model with linear interpolation being used to piece together the local intervals. In this case the limits are seen to differ for each model. For most of the duration of the batches, the SPE for the five validation batches was seen to lie within the limits of the model. However at some observation points, the SPE moves outside the limits. For example, batch 1, represented by ‘+’, is seen to be out of the limits at time points 24, 32 and 33, similar to those for the latent variable scores plot (Fig. 10). The SPE is then moves back inside.

Hotelling’s $T^2$ statistic was then calculated for each nominal batch and the 99% limit calculated for each observation point in each local model. Linear interpolation was then performed between the operating regions. The $T^2$ metric was calculated for each validation batch (Fig. 12). Each batch is identified by a different symbol as it moves outside.
the limits. Batch 1 was observed to move out of the limits at similar observation points as in the latent variable scores plot and the SPE plot.

Fig. 12. $T^2$ monitoring chart.

The monitoring charts exhibit some similarities in terms of the identification of non-conforming behaviour. The $T^2$ monitoring identifies the largest number of observations outside the action limits, indicative of a drift in process operating conditions. The SPE plot detects individual points outside the limits highlighting outlying observations.

It can also be observed from the SPE plot, Fig. 11, that validation Batch 2, moves outside of the limits at observation point 325. The source of the problem can be investigated through a contribution plot. Fig. 13 illustrates the multivariate nature of the outlying observation for batch 2 for point 325. Examination of the trajectories for variables 2, 7, 13 and 19 revealed a drop in power (variable 13) and reduced chemical A addition rate (variable 19) at observation point 325, leading to lower dissolved oxygen (variable 7) and pH (variable 2). This combined set of changes had a detrimental effect on product growth.

Fig. 13. Variable contribution to SPE with 99% limits.

7. CONCLUSIONS

A local model based multivariate process modelling and monitoring approach has been demonstrated by application to data from a large scale fed-batch industrial fermentation process. The use of a local modelling approach leads to an improvement in prediction capability and a reduction in the structure of the residuals, in comparison to an overall global model based on a quadratic non-linear dynamic PLS. By splitting the process into local operating regions, each local model included dynamics specific to that region of the process and was more capable of modelling the process. The global dynamic model was unable to capture the underlying batch behaviour. This can be seen clearly from the results of RSS. The monitoring tools illustrate the potential of the method for an at-line or on-line process performance monitoring scheme. Future work will compare the local modelling approach with a wider range of non-linear dynamic global models including neural network dynamic PLS.

ACKNOWLEDGEMENTS

Miss N. Fletcher acknowledges the EPSRC and CPACT for financial support of her PhD.

REFERENCES


