Proceedings of the 9th International Symposium on
Dynamics and Control of Process Systems (DYCOPS 2010),
Leuven, Belgium, July 5-7, 2010
Mayuresh Kothare, Moses Tade, Alain Vande Wouwer, Ilse Smets (Eds.)

Data Fusion for Enhanced Fermentation Process Tracking

Shengnan Yu, Gary Montague, Elaine Martin*,

School of Chemical Engineering and Advanced Materials, Newcastle University,
Newcastle upon Tyne, NE1 7RU, UK
(Tel: +44 (0)191 222 6231; e-mail: e.b.martin@ ncl.ac.uk)

Abstract: Near-infrared spectroscopy along with process control variables, such as integral of airflow rate and the integral of alkali addition rate can be used as the basis for the monitoring of key analyte concentrations on a fermentation process. Within this paper, sequential data fusion modeling is applied first, embracing both physical and chemical information. Aiming to overcome the limitations of sequential modeling and to compare model accuracy, a novel data fusion methodology based on Partial Least Squares, weighted multivariate calibration, is introduced. The methodologies are applied to data from an industrial fermentation process and it is shown that the data fusion method results in a 50% improvement in the Root Mean Square Error of Cross Validation (RMSECV) compared to more traditional calibration approaches. An optimisation procedure was then considered in association with spectral window selection (SWS) to attain more accurate data fusion models.

Keywords: Data fusion; Partial least squares; Near-infrared spectroscopy; Fermentation process

1. INTRODUCTION

In industrial fermentation processes, achieving process consistency and reproducibility is of particular importance, in the manufacture of product of consistently high quality (Navratil et al., 2004). To reduce batch-to-batch variation, a number of statistical process monitoring approaches have been developed, which utilize on-line process measurements to monitor key analyte concentrations including near-infrared (NIR) spectroscopy. This is a rapid, reliable, and robust monitoring technique, and a powerful method for qualitative and quantitative analysis (Šašić and Ozaki, 2001; Hermida et al., 2001). Together with other process monitoring sensors, a number of on-line batch process monitoring schemes have been proposed based on the multivariate projection techniques of Principal Component Analysis (PCA) and Partial Least Squares (PLS) (Nomikos and MacGregor, 1994,; Kourti et al., 1995; Wold et al., 1998).

Data fusion, a multivariate statistical analysis method is concerned with the use of methodologies that combine data from a number of sources. The concept of data fusion originates from marketing studies (Baker et al., 1989). The hypothesis is that data fusion can extract more information than that is achievable from a single source. The fusion of data from NIR spectroscopy and electronic noses (EN) has been successfully used for the monitoring of yogurt fermentation in a laboratory (Cimander et al., 2002). In other data fusion studies, the integration of process data (physical state) from a bioreactor, with on-line signals including spectroscopic data (chemical state) has also been shown to result in a significant improvement in bioprocess monitoring through the application of multivariate statistical data modelling and analysis tools (Trygg and Wold, 1998; Gurden et al. 2002).

In this paper, calibration models, based on partial least squares (PLS) are calculated for an industrial data set consisting of spectroscopic and process data. The goal is to assess whether the performance of individual calibration models is enhanced when either the spectroscopic and process variables or individual models are combined using a number of data fusion techniques.

One approach that has previously been shown to be successful was sequential data fusion (Triadaphillou et al, 2007). The basis of this approach is that a PLS model is built on the spectroscopic data to predict the analyte concentration of interest and the residuals from this model are then predicted by the process data using a second PLS model. A final calibration model is then formed by combining the two PLS models. The results showed that the accuracy of the calibration model was improved. A potential limitation of sequential data fusion is that no consideration is given to the weightings of the two PLS models. There is thus a research challenge to identify the appropriate weights. To address this issue a weighted data fusion approach is proposed where the PLS models are built individually and the weighting combination that gives the smallest RMSECV (root mean square error of cross validation) is determined.

The second approach considered is based on the optimisation of the weighting of the individual variables as opposed to the models. In summary, it is observed that the optimisation approach gives the model with the lowest RMSECV for this data set.

2. CASE STUDY

2.1 Process Description

The dataset, comprising five batches, was generated from an industrial fermentation process where each batch is of approximately 10 days in duration. Within this study, model building is based on two calculated process variables, the integral of airflow rate and the integral of alkali addition rate.

These variables are indicative of the behaviour of the culture as both dissolved oxygen and pH are under control. The process measurements are recorded online with a sampling rate of 5 s. The NIR spectra are also monitored with a sampling rate of approximately every 1 min 50 s. The offline measurement of interest in this paper is glucose concentration (g/l). It was measured once or twice a day.

The first step in the analysis was to address the difference in sampling rates. The datasets were merged and data alignment was carried out using the nearest point method, i.e. the data for modelling was selected based on the data/time stamp generated by each data logging system. Critical to this approach was to ensure that the two computer data collection systems were synchronized prior to data acquisition. In summary the more frequently sampled sources (the process and spectroscopic data) were down-sampled to align with the off-line data. This resulted in approximately 12 samples per batch.

2.2 Pre-processing of Spectroscopic Data

Data pre-processing is an essential step irrespective of the objectives of model development. The following procedures were carried out using MATLAB software with the PLS toolbox (Eigenvector Research, INC, Manson, WA).

Prior to performing mathematical transformations on the data, the first step was to remove the areas of instrument saturation from the spectroscopic data. From the raw NIR spectra plot (Fig. 1) two areas of saturation can be observed, 4500 ~ 5300 cm-1 and 6500 ~ 7200cm^{-1}. These sections were removed from the data set for all batches.

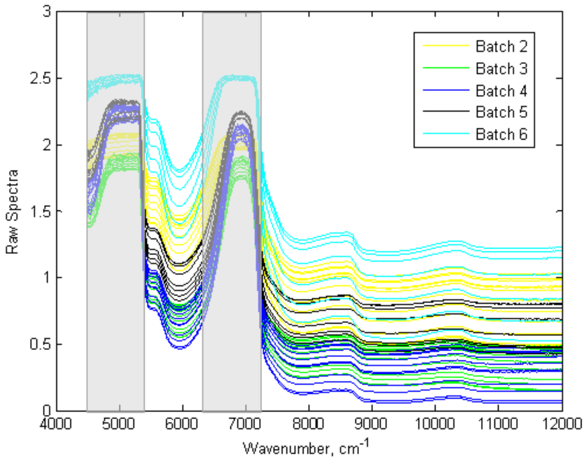

Fig. 1. Raw NIR data plot

The next stage was to apply MSC (Multiplicative Scattering Correction) to each batch, to remove the scattering effect due to the process not being a homogenous mixture and the viscosity of the product changing over time thereby impacting on the NIR spectra as can be observed from Fig.1. In the application of MSC, the reference spectrum was selected as the mean spectra of each batch. Savitzky Golay smoothing and first derivatives were then applied to the MSC corrected data to smooth the signal and remove the water background and baseline offsets. The first derivative was calculated using Savitsky Golay smoothing with a 15 point window and a second order polynomial. These settings were found to be appropriate to achieve noise reduction whilst maintaining signal information content.

The final stage was to select regions of the spectra based on knowledge of the chemical structure of the product. The wavenumbers of interest were 5997.5 ~ 5634.9 cm^{-1}, the first overtone of the CH band and 7258.7 ~ 7243.3 cm^{-1}, the second overtone of the CH band.

2.3 Computational Methods

In spectral calibration, the most commonly applied method is PLS, (Geladi and Kowalski, 1986). The goal of PLS regression is to predict the response **Y** (glucose concentration) from the descriptors **X** (spectral absorbances). In PLS modelling, it can be considered that there are two relationships, the outer relationship, which can explain **X** and **Y** individually and the inner relationship, which links the two blocks. The outer relationship is given by:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{j=1}^{h} \mathbf{t}_j \mathbf{p}_j^T + \mathbf{E} \tag{1}$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}^* = \sum_{j=1}^{h} \mathbf{u}_j \mathbf{q}_j^T + \mathbf{F}^* \tag{2}$$

From these two orthogonal decompositions, it can be observed that linear PLS projects **X** and **Y** onto a number of latent variables (LVs), $\mathbf{t}_j$ and $\mathbf{u}_j$, respectively. The loadings for each block **X** and **Y** are also attained, $\mathbf{p}_j$ and $\mathbf{q}_j$. The ultimate aim of PLS modeling is to minimise $\mathbf{F}^*$ (the residual of the **Y** decomposition) whilst at the same time calculating the inner relation between **X** and **Y**:

$$\hat{\mathbf{u}}_j = \mathbf{b}_j \mathbf{t}_j \text{ where } j = 1,2,\ldots,h \tag{3}$$

where $\mathbf{b}_j = \mathbf{u}_j^T \mathbf{t}_j / \mathbf{t}_j^T \mathbf{t}_j$, and $\mathbf{b}_j$ is equivalent to a regression coefficient. A key step in PLS is the determination of the number of LVs (h) that are retained in the model. This can be determined using cross validation. The final model is given by:

$$\mathbf{Y} = \mathbf{TBQ}^T + \mathbf{F}^* \tag{4}$$

2.4 Spectral Window Selection (SWS)

In spectral analysis, one of the issues is whether to apply the analysis to all the wavenumbers or only specific regions. More specifically when developing a calibration model from the complete spectra for quantitative analysis, the prediction results can be affected by those wavenumbers that do not provide predictive information about the analyte of interest (Triadaphillou et al., 2007). Consequently wavenumber selection can be used to address this issue by removing noise or other variations that may degrade calibration model performance.

One approach to wavenumber selection is through knowledge of the analyte chemical structure as described previously. However it is not always possible to select specific regions due to a lack of knowledge of the chemical composition of the variable of interest. Thus it is important to consider alternative statistical based approaches. One methodology that will be considered is spectral window selection (SWS). This is a more flexible method for seeking regions that are more appropriate for model construction. The philosophy and algorithm of SWS was discussed in Hinchliffe et al. (2003) where it was termed “the binning method”. Triadaphillou et al. (2007) extended the concept to spectroscopic data. The underlying algorithm is as follows:

1. Mean centre the pre-processed spectroscopic data and the process data;

2. Set up the initial specification of the number of windows to be selected, (only one window is considered in this case study), and randomly select the starting wavenumber (s_1) and width (w_1) for the first window;

3. The wavenumbers selected in the window are extracted and where more than one window is being considered the overlapping regions are removed;

4. A PLS calibration model is built on the data generated in Step 3 and the root mean square (RMS_1) is calculated;

5. Increments Δs_1 and Δw_1 are generated from a uniform distribution to randomly change the starting wavenumber and width of window. ($s_2=s_1+\Delta s_1$; $w_2=w_1+\Delta w_1$). A new PLS model is built on the new window and RMS_2 is calculated;

6. RMS_2 is compared with RMS_1 and if RMS_2 is smaller then the model is improved and hence the incrementsΔs_1 and Δw_1 are retained ($\Delta s_2=\Delta s_1$; $\Delta w_2=\Delta w_1$) and used to search for the next window ($s_3=s_2+\Delta s_2$; $w_3=w_2+\Delta w_2$). If RMS_1 is smaller repeat steps 5 and 6, randomly selecting a new Δs_1 and Δw_1.

7. This search procedure is repeated until the specified number of windows has been retained and the final calibration model obtained or a time limit/number of iterations is reached for the seeking of a model that gives the smallest RMS.

3. DATA FUSION METHOLODGIES

3.1 Sequential Data Fusion

Sequential data fusion modelling was first proposed by Triadaphillou et al (2007). A schematic of the methodology is given in Fig. 2. More specifically the steps in the algorithm for the data set being analysed are as follows:

(1) The first step is to fit a PLS calibration model, model A, to the glucose concentration using the spectroscopic data as the input variables. The residuals of this model are then calculated.

(2) The resulting residuals are then modelled using the process data (integral of airflow rate and the integral of alkali addition rate) by a second PLS calibration model, model B. The residuals in this case are termed the innovations.

(3) The last step comprises calculating the final model that is formed from the predictions of the two models.

$$\begin{aligned} &\mathit{Model\ A}: \mathbf{Y} = \mathbf{X}_A\mathbf{B}_A + \mathbf{E}_A \\ &\mathit{Model\ B}: \mathbf{E}_A = \mathbf{X}_B\mathbf{B}_B + \mathbf{E}_B \\ &\mathit{Final\ Model}: \mathbf{Y} = \mathbf{X}_A\mathbf{B}_A + \mathbf{E}_A + \mathbf{E}_B \end{aligned} \tag{5}$$

This sequential modelling methodology is one approach to data fusion, and offers the opportunity to take account of both chemical and process information. The rationale for utilising the NIR data is that it provides specific information on glucose concentration whilst the process data provides insight into the operation of the process.

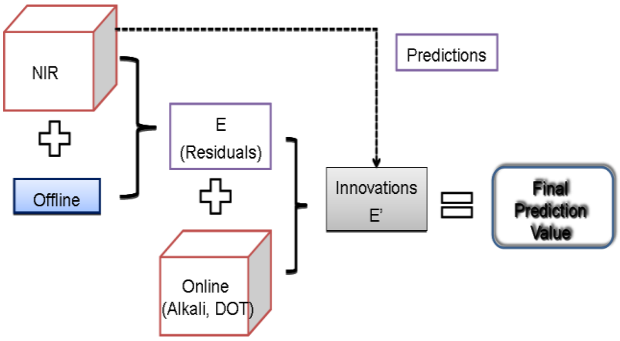

Fig 2. Sequential Data Fusion Strategy

3.2 Weighted Multivariate Calibration

One limitation of the sequential modeling approach is that for this specific dataset, the first PLS model explains in excess of 90% of the glucose concentration in the training data consequently the inclusion of the process data is theoretically unnecessary. To address this imbalance with respect to the weightings, Liu, et al. (2009) proposed a weighted multiscale regression methodology. The weighting was based on the prediction residual error sum of squares (PRESS):

$$w_i = \frac{\left(1/\mathrm{PRESS}_i^2\right)}{\sum_{i=1}^{n}\left(1/\mathrm{PRESS}_i^2\right)} \tag{6}$$

where the PRESS is given by:

$$\mathrm{PRESS}_i = \frac{1}{N}\sum_{j=1}^{N}\left(\sum_{k=1}^{M}\left(y_{jk} - \hat{y}_{jk}\right)^2\right) \tag{7}$$

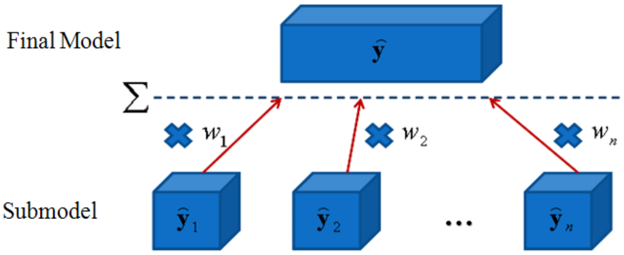

Fig 3. Weighted Multivariate Calibration Model Structure

To investigate the influence of the weightings on the model, weighted multivariate calibration was studied. PLS models were built individually for the NIR data and the process variables. These were then combined utilising pre-defined weightings (w_1, w_2, …, w_n, where $\sum_{i=1}^{n} w_i = 1$). The final model is given by:

$$\mathit{Model}: \hat{\mathbf{y}} = \hat{\mathbf{y}}_1 w_1 + \hat{\mathbf{y}}_2 w_2 + \ldots + \hat{\mathbf{y}}_n w_n \tag{8}$$

where n is the number of blocks or submodels. A schematic of the model structure is given in Fig 3.

4. RESULTS AND DISCUSSION

Three sets of variables are considered in this paper, NIR spectra, integral of airflow rate and the integral of alkali addition rate. The latter two are process related variables. They can be considered individually as two separate data blocks or treated as one block comprising two columns. Both situations are investigated in this paper.

4.1 Two Dimensional Model

Two blocks are analysed in this approach, the NIR spectra and the two process variables. The NIR data block comprise the preprocessed data with the wavenumbers associated with the first overtone, first overtone combination and second overtone wavenumbers of CH band. The model for this two dimensional model is:

$$\text{2D Model}: \hat{\mathbf{y}} = \hat{\mathbf{y}}_1 w_1 + \hat{\mathbf{y}}_2(1-w_1) \qquad (9)$$

where $\hat{\mathbf{y}}$ is the final model prediction; $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ are the predictions from the PLS model built on the NIR block and process data block respectively; and w_1 is the weighting of the NIR submodel. To attain appropriate weightings, all possible combinations were investigated and the set that gave the minimum RMSECV was selected, Fig. 4.

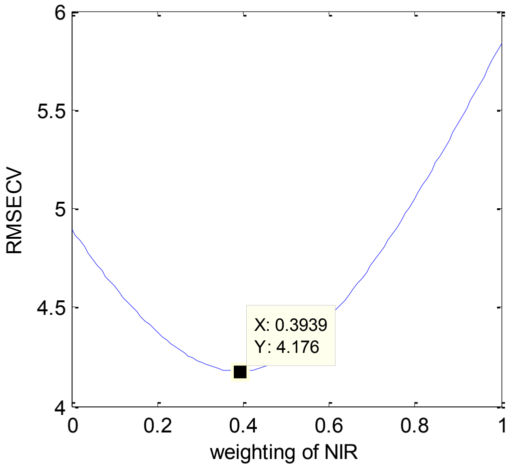

Fig 4. RMSECV for the Two Dimensional Model

The 2D model is given in equation (10), 39.4% of the model is explained by NIR ($\hat{\mathbf{y}}_1$), the remaining 60.6% explained by the process variables ($\hat{\mathbf{y}}_2$).

$$\hat{\mathbf{y}} = 0.394 \times \hat{\mathbf{y}}_1 + 0.606 \times \hat{\mathbf{y}}_2 \qquad (10)$$

The plot of observed versus fitted values is given in Fig. 5 and it can be concluded that the model is satisfactory.

The weighting approach proposed by Liu, et al. (2009) was also considered. The PRESS was calculated using leave-one-batch-out method since there were only 5 batches in the dataset. The following model was attained:

$$\hat{\mathbf{y}} = 0.344 \times \hat{\mathbf{y}}_1 + 0.656 \times \hat{\mathbf{y}}_2 \qquad (11)$$

The weightings differ slightly but in both cases the process data is more important with respect to capturing the level of glucose concentration.

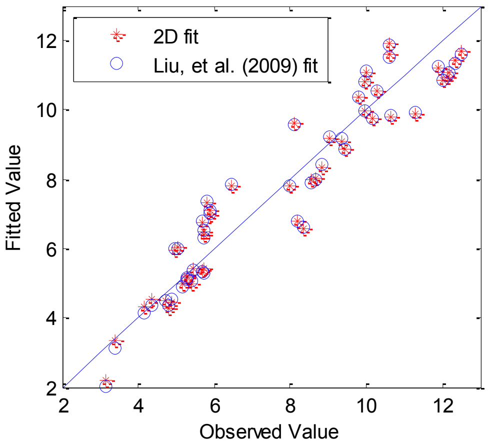

Fig 5. Fit of Two Dimensional Model

4.2 Three Dimensional Model

Three blocks NIR ($\hat{\mathbf{y}}_1$), integral of airflow rate ($\hat{\mathbf{y}}_2$) and the integral of alkali addition rate ($\hat{\mathbf{y}}_3$) were then considered. The underlying model can be written as:

$$\text{3D Model}: \hat{\mathbf{y}} = \hat{\mathbf{y}}_1 w_1 + \hat{\mathbf{y}}_2 w_2 + \hat{\mathbf{y}}_3(1-w_1-w_2) \qquad (12)$$

Fig 5 is the RMSECV plotted against the weights of the NIR block and the integral of airflow rate (DOT). By finding the minimum RMSECV, the appropriate weightings can be selected.

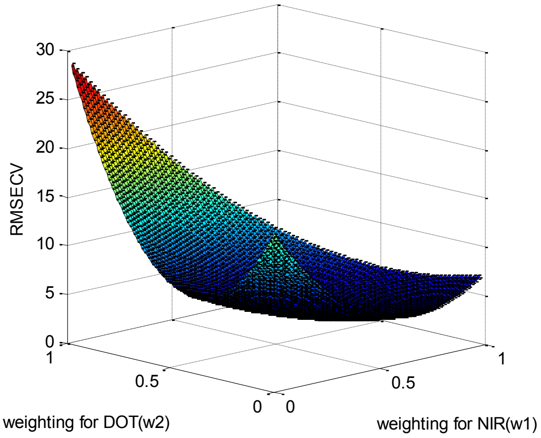

Fig 6. RMSECV in Three Dimensional Model

The final model is given in equation (11), in which 55.6% of the model is explained by NIR, 10.1% is explained by DOT and the remaining 34.3% is explained by alkali addition rate.

$$\hat{\mathbf{y}} = 0.556 \times \hat{\mathbf{y}}_1 + 0.101 \times \hat{\mathbf{y}}_2 + 0.343 \times \hat{\mathbf{y}}_3 \qquad (13)$$

The performance of this 3D model is shown in Fig 7, where the results are again satisfactory.

The weighting approach of Liu, et al. (2009) was also considered and the following model was attained:

$$\hat{\mathbf{y}} = 0.800\times\hat{\mathbf{y}}_1 + 0.042\times\hat{\mathbf{y}}_2 + 0.158\times\hat{\mathbf{y}}_3 \qquad (14)$$

Compared with the proposed 3D approach, model (14) does not give as good fit. In the approach of Liu et al (2009), the weighting is based on the individual samples as opposed to the individual models and thus in this case the NIR data dominates the calculation.

Comparing the proposed 2D model with the proposed 3D model (equations 10 and 13 respectively) the process component is dominant. The rationale for this is that the combined effect of integral of airflow rate and the integral of alkali addition rate impact on glucose concentration. This correlation structure is lost when treating the variables independently.

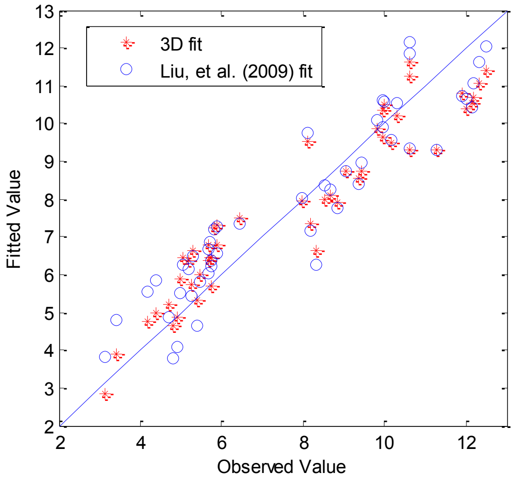

Fig 7. Fit of the Three Dimensional Model

4.3 Comparison Between Data Fusion Methods

The root mean square error of cross validation (RMSECV) is calculated for the different data fusion methods. However as there are only five batches in the data set, leave-one-batch-out method was applied, i.e. each batch was omitted from the training data set and served as a validation batch. The results are summarised in Table 1.

Table 1 RMSECV Comparison between Models

Validated Batch	NIR Raw	Pre-processed NIR	Sequential	2D	3D
B2	1.906	1.060	1.130	1.052	1.529
B3	4.827	1.515	0.892	0.611	0.597
B4	1.172	0.583	0.804	0.640	0.720
B5	2.552	1.448	1.363	0.856	0.443
B6	6.214	1.233	1.065	1.016	1.108
Sum	16.671	5.838	5.253	4.176	4.397

In Table 1, five models are presented. The first is built solely from the raw NIR spectra; the second is based on the pre-processed NIR data. With the third built from the sequential modelling approach. The fourth and fifth models are the two-dimensional and three-dimensional weighted multivariate calibration models discussed in the previous sections. Although not reported the sum RMSECV from the Liu et al. (2009) approach are 4.188 and 5.322 for the 2D and 3D models respectively.

Comparing the single source models with the data fusion based models it is evident that data fusion enhanced the results in terms of the RMSECV. Comparing across the three data fusion models the performance of the 2D and 3D models is better than sequential modelling, indicating the weighting of each submodel in data fusion is critical. These results are presented in Fig 8. In this figure the error for the individual batches is also presented.

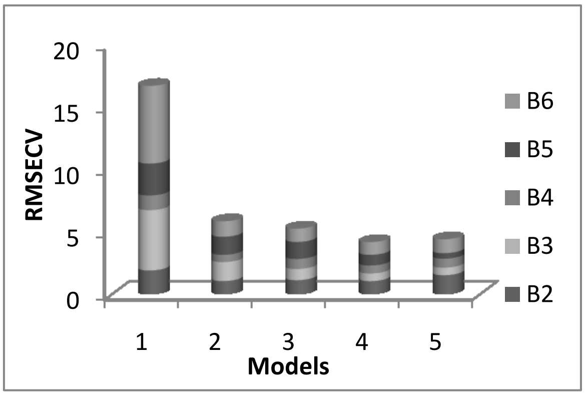

Fig 8. RMSECV Comparison Bar Chart (Model 1 – NIR raw data; Model 2 – Preprocessed NIR data; Model 3 – Sequential Model; Model 4 – 2D Model; 5 – 3D Model)

4.4 Optimisation of Weighted Multivariate Calibration

After the data fusion methodology was found to exhibit better performance, an alternative approach was considered based on the optimisation of the weightings of the individual variables. The general expression is given by:

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1 + \mathbf{X}_2\mathbf{B}_2 + \mathbf{E} \qquad (15)$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are the regression coefficients of data block $\mathbf{X}_1$ (NIR) and $\mathbf{X}_2$ (Process) respectively, and $\mathbf{E}$ is the error. This model can be rewritten as:

$$\mathbf{E} = \mathit{function}(\mathbf{B}_1, \mathbf{B}_2) \qquad (16)$$

By applying this optimisation function, the minimum error can be found.

Table 2 SWS Search Result

Search No.	Start point (cm^{-1})	End Point (cm^{-1})	RMSEC
1	11991	11987	4.406
2	11987	11980	4.256
3	11976	11956	4.180
4	9492	9419	4.079
5	7687	7332	3.833

Additional to the optimisation of the weights, the selection of the wavenumbers through application of the SWS algorithm to the pre-processed NIR data was considered. It is conjectured that by refining the selection of wavenumbers an enhanced model will materialise. The results following the application of SWS are given in Table 2. Five windows were selected from the optimisation process, and the root mean square error of calibration (RMSEC) was recorded for each window. Based on the RMSEC, window 5 was selected for the subsequent analysis, and validation was then performed. The results are given in Table 3 and Fig. 9.

Table 3 RMSECV Comparison after Optimisation

Validated Batch	2D	Optimisation without SWS	Optimisation with SWS
	Model 1	Model 2	Model 3
B2	1.052	1.072	0.994
B3	0.611	0.494	0.516
B4	0.640	0.753	0.714
B5	0.856	0.649	0.626
B6	1.016	0.899	0.773
RMSECV	4.176	3.867	3.622

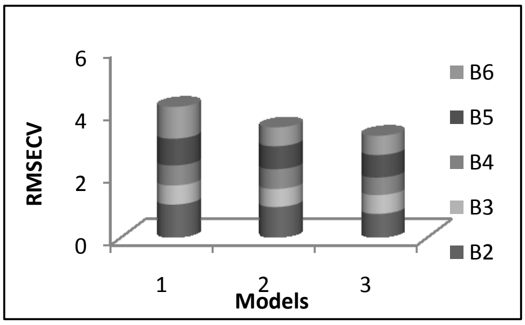

Fig 9. RMSECV Comparison after Optimisation

(Model 1 – 2D model; Model 2 – Optimization on fixed window; Model 3 – Optimization with SWS)

From both Table 3 and Figure 9, the benefits of optimisation can be observed. It can be seen that by refining the wavenumbers selected based on knowledge of the chemistry through the application of SWS the performance of the model was improved.

5. CONCLUSIONS

This paper has presented two approaches to data fusion, sequential modelling and weighted multivariate calibration. These two methods were applied to an industrial dataset. Both of these data fusion methods gave better performance than a single source data model with respect to the RMSECV, with approximately 50% improvement being observed thereby demonstrating that the combination of spectroscopic and process data provides improved calibration model accuracy and facilitates greater understanding of process behaviour. As a novel methodology, weighted multivariate calibration improves model performance by finding the best weighting balance of each variable. Optimisation was then applied to the weighted multivariate calibration approach alongside the refinement of the wavenumbers through the application of the SWS algorithm. Once again an improvement in model performance was observed.

ACKNOWLEDGEMENTS

This project is co-funded by the UK TSB Collaborative R& programme, following an open competition. The TSB is an executive body established by the Government to drive innovation. It promotes and invests in R&D and the exploitation of science, technology and new ideas for the benefit of business - increasing sustainable economic growth in the UK and improving quality of life. For more information visit www.innovateuk.org

REFERENCES

Baker, K., P. Harris, et al. (1989). "Data fusion: An appraisal and experimental evaluation." *Journal of the Market Research Society*, 31(2): 153-212.

Cimander, C., M. Carlsson, et al. (2002). "Sensor fusion for on-line monitoring of yoghurt fermentation." *Journal of Biotechnology*, 99(3): 237-248.

Geladi, P. and B. R. Kowalski (1986). "Partial least-squares regression: a tutorial." *Analytica Chimica Acta*, 185: 1-17.

Gurden, S. P., J. A. Westerhuis, et al. (2002). "Monitoring of batch processes using spectroscopy." *AIChE Journal*, 48(10): 2283-2297.

Hermida, M., J. M. Gonzalez, et al. (2001). "Moisture, solids-non-fat and fat analysis in butter by near infrared spectroscopy." *International Dairy Journal,* 11(1-2): 93-98.

Hinchliffe, M., G. Montague, et al. (2003). "Correlating polymer resin and end-use properties to molecular-weight distribution." *AIChE Journal*, 49(10): 2609-2618.

Kourti, T., P. Nomikos, et al. (1995). "Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS." *Journal of Process Control*, 5(4): 277-284.

Liu, Z., W. Cai, et al. (2009). "A weighted multiscale regression for multivariate calibration of near infrared spectra." *Analyst*, 134(2): 261-266.

Navratil, M., C. Cimander, et al. (2004). "On-line Multisensor Monitoring of Yogurt and Filmjölk Fermentations on Production Scale." *Journal of Agricultural and Food Chemistry*, 52(3): 415-420.

Nomikos, P. and J. F. MacGregor (1994). "Monitoring batch processes using multiway principal component analysis." *AIChE Journal*, 40(8): 1361-1373.

Šašić, S. and Ozaki, Y. (2001). "Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment." *Analytical Chemistry*, 73(1): 64-71.

Triadaphillou, S., E. Martin, et al. (2007). "Fermentation process tracking through enhanced spectral calibration modeling." *Biotech. & Bioeng.*, 97(3): 554-567.

Wold, S., N. Kettaneh, et al. (1998). "Modelling and diagnostics of batch processes and analogous kinetic experiments." *Chemometrics and Intelligent Laboratory Systems*, 44(1-2): 331-340.