

Olli Haavisto and Heikki Hyötyniemi. 2009. Recursive multimodel partial least squares estimation of mineral flotation slurry contents using optical reflectance spectra. *Analytica Chimica Acta*, volume 642, numbers 1-2, pages 102-109.

© 2008 Elsevier Science

Reprinted with permission from Elsevier.



Recursive multimodel partial least squares estimation of mineral flotation slurry contents using optical reflectance spectra

Olli Haavisto*, Heikki Hyötyniemi

Department of Automation and Systems Technology, Helsinki University of Technology (TKK), P.O. Box 5500, FI-02015 TKK, Finland

ARTICLE INFO

Article history:

Received 9 July 2008

Received in revised form 9 October 2008

Accepted 6 November 2008

Available online 18 November 2008

Keywords:

Mineral flotation

Reflectance spectroscopy

Recursive partial least squares

Orthogonal signal correction

Local modeling

ABSTRACT

In mineral flotation the X-ray fluorescence (XRF) grade measurements of the slurries typically give the most important on-line information on the state of the flotation process. It has been shown that the visual and near-infrared (VNIR) reflectance spectrum measurements of certain mineral slurries can be used to complete the sparse XRF slurry content information. This study focuses on the chemometrical analysis of the VNIR spectrum of the slurries and presents a new partial least squares (PLS)-based recursive multimodel approach with local orthogonal signal correction (OSC) for predicting the slurry contents. The advantage of the presented approach is that it can recursively adapt to real process data variations in normal operating conditions, and is still able to remember the rare process failure situations with notably different content values.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Froth flotation is one of the most important concentration methods in mineral processing (see, e.g. [1]). Before flotation the ore is mixed with water and ground to small particle size in mills to liberate the minerals. Flotation chemicals are then added to the formed slurry in order to modify the surface chemical properties of the particles. Slurry is further processed in flotation cells where air is pumped to the bottom of the cell and heavy mixing is applied. Due to the chemicals the particles containing specific minerals attach to the rising air bubbles and are conveyed to the froth layer on the surface of the cell. Froth is then allowed to overflow or skimmed to obtain the concentrate rich in the specific minerals, whereas the other particles are removed in the tailing flow from the bottom of the cell.

In order to effectively monitor and control a flotation process consisting of several flotation cells and slurry lines connecting them, it is essential to be able to assay the main slurry lines in real time and on-line. The most common on-line assaying device in large concentration plants is an X-ray fluorescence (XRF) analyzer, which is capable of measuring the elemental contents of the slurry. However, a typical XRF analyzer contains only one analyzer probe which measures each slurry line in turns increasing the sampling inter-

val of a single line to over 10 min. This deteriorates the detection of sudden content changes like process failures or high-frequency oscillations occasionally present in flotation circuits.

The image and color analysis of flotation froths and slurries has been intensively studied during the last 15 years [2–5] as an alternative assaying method and for determining the state of the flotation cells. However, mainly ordinary cameras have been utilized also for measuring the color. In [6] it was shown that visual and near-infrared (VNIR, 400–1000 nm) diffuse reflectance spectroscopy can be successfully utilized as a supplemental on-line method for determining the elemental contents of zinc concentrate slurry along with the traditional XRF analyzer in a real process environment. The advantages of the presented VNIR spectral approach with respect to the XRF analysis are – as in many other applications – speed and simplicity: Spectra can be easily measured with a very small sampling interval (e.g. 10 s) without specific sample preparation. Using a PLS-based modeling approach, a practically continuous on-line assay of the slurry can be provided. The improved slurry content information can be further utilized by the automatic control system and the process operators to react faster to the unexpected changes in the process, thus improving the overall productivity.

Due to variations in the slurry properties (particle size distribution, solid content, mineralogy) a regular PLS model between the spectra and the slurry contents calculated using a fixed calibration set does not remain accurate for longer periods. This is why in [6] the PLS model was recursively adapted with new XRF and spectrum samples obtained during the normal operation of the process. The learning was achieved by a recursive partial least squares (rPLS)

* Corresponding author. Tel.: +358 9 451 6090; fax: +358 9 451 5208.

E-mail addresses: olli.haavisto@tkk.fi (O. Haavisto), heikki.hyotyniemi@tkk.fi (H. Hyötyniemi).

model (see, e.g. [7]) with nonlinear data preprocessing and rather fast exponential forgetting.

Effectively a recursive modeling algorithm gives the same result as would the re-calibration of a batch-type regression model, assuming that the re-calibration was performed with all the previous calibration samples every time a new sample is measured. The exponential forgetting means that the calibration samples are taken into account in the modeling using exponentially decreasing weights: the older the sample is, the smaller effect it has in the final model. The difference is that in a recursive algorithm the re-calibration is realized by incrementally adapting the existing model instead of storing and re-using all the previous calibration samples, thus saving both memory and computations. Exponential forgetting is conveniently obtained by including a forgetting factor λ in the adaptation formula. With a suitable forgetting factor the adapted model mainly describes the recent data and gradually forgets the old data that has become irrelevant. Exponential forgetting is a common method, e.g. in recursive system identification algorithms [8].

It is a common requirement for adaptive systems [9] that the data should be persistently exiting. This means that a rPLS model in slurry VNIR spectrum analysis works well only for those slurry lines and elements where the content values vary continuously over the whole data range, so that despite the forgetting the calibration model remains robust. However, in flotation circuits some content values are in normal operation very small, but in the (rare) case of a process failure they can suddenly increase considerably for a short period of time. As demonstrated in this study, the regular rPLS model presented in [6] cannot predict well these process failures, especially if a long time has passed since the previous such failure. However, because these increases in the normally small contents are an indicator of an exceptional process behaviour, they should be detected as soon as possible.

This study focuses on the detection of the increased slurry content levels related to process failures. In Section 2 a recursive multimodel PLS algorithm (rMM) is derived. It divides the set of all possible spectra and XRF samples among several local models, each covering only a subset of samples located near the center point of that model in the predictor and response domain. Correspondingly, each local model is calibrated mostly with the calibration samples located in its subset. The final prediction of the total model is calculated as a weighted combination of the local model predictions. In the prediction phase, each local model performs its own orthogonal signal correction (OSC) preprocessing [10] to improve the evaluation of the validity of that local model for the current predictor sample. The proposed algorithm utilizes the kernel versions of OSC and PLS algorithms.

The performance of the rMM algorithm is demonstrated by applying it to a large VNIR spectrum and XRF sample data set collected from a real zinc flotation circuit and comparing the copper content modeling results to the results given by normal PLS and recursive PLS models.

2. Theory

2.1. Problem formulation

Due to the differences in the sampling frequencies of the VNIR spectrophotometer and the XRF analyzer, new predictor samples x (spectra) are continuously received with a high sampling frequency, and response samples y (XRF assays) with a low sampling frequency. Assuming that the VNIR spectrophotometer and the XRF analyzer measure the same slurry sample at the same time (or with a small constant time difference), for each XRF assay a correspond-

ing spectrum can be found, thus forming a set of predictor-response pairs suitable for data-based model calibration and validation. The rest of the VNIR spectra are used to predict the element contents when no XRF measurement is available.

The aim of the modeling is to form an on-line mapping from the predictors to the responses with the following additional restrictions: (1) Since the modeling approach is to be applied continuously to real process data, there exists variation in the relationship between the predictors and responses; i.e. the mapping from predictors to responses is not constant in time. (2) The variation in data does not cover the whole data range evenly in time; instead, for long periods the data variation is small (normal operation), and only seldom large peaks are received (process failures).

2.2. Approach

A classical way to tackle with the first restriction is to assume that the changes in the process are slow and smooth, so that it is possible to continuously adapt the model with new data to better predict the response variable values for the future predictors. For this purpose, a rPLS algorithm with exponential weighing of old data can be used. The existing rPLS approaches are typically based on the recursively updated data covariance matrices and kernel-type PLS calculation [11,7], even though some other variants have also been presented [12,13].

To prevent the model from forgetting the rare process failures, a local modeling approach is proposed in this study. The modeling data space – both in predictor and response domains – is divided among several local rPLS models with exponential forgetting instead of using one global rPLS model. The assumption is that if only the data near the center of each local model is utilized in calibration of that model, the model remains representative.

Traditionally, local models are typically used to deal with nonlinearities in the data as an alternative to global nonlinear models. In the field of near-infrared spectroscopy (NIRS) local modeling has been utilized to exploit spectral libraries: a local model is fitted to the subset of library spectra that are similar to an unknown spectrum to be predicted. This is repeated for each new unknown spectrum. There exists several different strategies for subset selection and model calibration: locally weighted regression (LWR) was originally suggested in [14] as a data smoothing method and has since then been further developed to a localized version of principal component regression [15,16]; CARNAC initially [17] utilized Fourier transformation in subset selection, but also other versions exist; the LOCAL algorithm [18] uses local PLS models; and locally-biased regression [19], that utilizes both PLS and OSC. The application of these methods in NIRS is recently reviewed in [20], showing that local methods indeed are a common tool for nonlinear spectrum data analysis among the variety of artificial neural networks.

Instead of calculating a new local model for each new predictor, it is also possible to precalculate a set of local models which cover the whole data range [21]. The final response prediction is then obtained, e.g. as a weighted sum of the local predictions. This approach has a couple of benefits: only the models have to be stored in the memory, not the whole data set; and the calculation of new predictions is less computationally intensive, since no model calibration is required at the prediction step.

Recursive adaptation of new data with exponential forgetting can also be utilized in model structures with several precalculated local models [22]. This approach has specially been developed for robot control [23,24], where strong nonlinearities typically exist. The same approach has further been expanded to work also in high dimensions [25] by exploiting PLS.

In this study, the local models are utilized for storing the regression information over time, not so much for compensating the nonlinearities of the data. In the earlier phase of this work [26], a memory-based LWR-type algorithm was used to deal with the model degeneration issue. However, the structure consisting of a fixed number of local models was preferred here since it is less memory intensive solution and suits better for recursive updating.

Evaluation of modeling performance for data-based methods is typically conducted by cross-validation. Because of the adaptive nature of the modeling method in this study, it is the one-step prediction performance that is the key measure of the model quality. This can be easily evaluated by predicting each new response y with the existing model before using the sample for adaptation, and then comparing the estimate \hat{y} with the measured value. Technically this procedure is comparable to the traditional cross-validation, since the model is validated with data not (yet) used in the model calculation.

2.3. Proposed modeling algorithm

The recursive multimodel OSC and PLS (rMM) modeling approach proposed in this study is based on a set of N local PLS models with OSC preprocessing. The centers of the local models are located in fixed points μ_y^i , $i = 1, \dots, N$ in the response space. Each local model is recursively calibrated as the data are gathered, thus enabling the on-line use of the rMM model. To maintain the localization, Gaussian weighing functions are used in the calibration (ρ_y^i) and prediction (ρ_x^i). This results in a situation where a subset of predictor space (spectra) centered in μ_x^i is modeled by the i th local model to a subset in the response (elemental contents) space centered in μ_y^i .

In the experimental part of this study, a simple case of the model structure with two local models and one-dimensional response space is utilized as shown in Fig. 1. The first local model ($i = 1$) is centered in the normal operating range of the process in the response space, whereas the second model ($i = 2$) handles the large response values related to process failures. Because of the localized adaptation, the first model continuously adapts to the normal operation of the process, and the second model contains the latest available information related to process failures. In the following

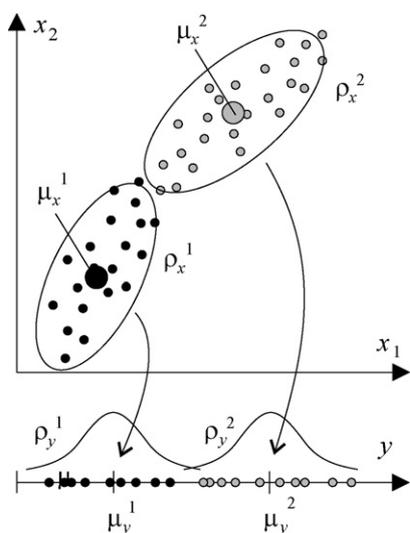


Fig. 1. Multimodel structure with two local models. Only two components of the predictor x are shown.

the rMM approach is derived in its general form with N local models and several response variables (i.e. y is a vector).

2.3.1. Model adaptation

Every time a new XRF measurement $y(k)$ is obtained, the corresponding VNIR spectrum $x(k)$ is picked and the local model structure is adapted. Adaptation in each local model i is based on the recursive updating of the old data covariance matrix estimates $\hat{R}_{xx}^i(k-1)$ and $\hat{R}_{xy}^i(k-1)$ with the new calibration sample pair $\{x(k), y(k)\}$:

$$\hat{R}_{xx}^i(k) = \lambda^i(k)\hat{R}_{xx}^i(k-1) + (1 - \lambda^i(k))x(k)x^T(k), \quad (1)$$

$$\hat{R}_{xy}^i(k) = \lambda^i(k)\hat{R}_{xy}^i(k-1) + (1 - \lambda^i(k))x(k)y^T(k), \quad (2)$$

where $0 \leq \lambda^i(k) \leq 1$ is the exponential forgetting factor. Additionally, the predictor mean value estimate for each model is updated correspondingly:

$$\hat{\mu}_x^i(k) = \lambda^i(k)\hat{\mu}_x^i(k-1) + (1 - \lambda^i(k))x(k). \quad (3)$$

The forgetting factor $\lambda^i(k)$ is determined according to the validity $\rho_y^i(k)$ of the calibration sample measured in the response (y) space:

$$\rho_y^i(k) = \exp[-(y(k) - \mu_y^i)^T \tau^{-1}(y(k) - \mu_y^i)], \quad (4)$$

$$\lambda^i(k) = \lambda_0^{\rho_y^i(k)}, \quad (5)$$

where τ is a scaling factor, μ_y^i is the fixed response mean of model i and λ_0 is the nominal forgetting factor. Here the validity function $\rho_y^i(k)$ (see, e.g. [21], Ch. 3) is selected as the unnormalized Gaussian function giving values between 0 and 1. Thus the forgetting factor varies between 1 (no updating when $y(k)$ is very far from μ_y^i) and λ_0 (maximal updating when $y(k) = \mu_y^i$).

Given the updated covariances (1) and (2), one has to calculate the OSC weights $W^i(k)$ and loadings $P^i(k)$ so that the OSC preprocessing can be performed. The OSC calculation is based on the kernel approach presented by Fearn [27]. However, since only the covariance matrices are available, the original data matrix inner products have to be replaced by the covariance matrix estimates: The f OSC weights

$$W^i(k) = [w_1^i(k), \dots, w_f^i(k), \dots, w_j^i(k)] \quad (6)$$

can be determined as the f most important eigenvectors of

$$M^i(k)\hat{R}_{xx}^i(k), \quad (7)$$

where

$$M^i(k) = I - \hat{R}_{xy}^i(k)(\hat{R}_{xy}^i(k)^T \hat{R}_{xy}^i(k))^{-1} \hat{R}_{xy}^i(k)^T. \quad (8)$$

Furthermore, also the OSC loadings can be obtained using the covariance matrices:

$$p_j^i(k) = \frac{\hat{R}_{xx}^i(k)w_j^i(k)}{w_j^i(k)^T \hat{R}_{xx}^i(k)w_j^i(k)}, \quad (9)$$

and collected in the columns of the matrix

$$P^i(k) = [p_1^i(k), \dots, p_j^i(k), \dots, p_f^i(k)]. \quad (10)$$

The OSC deflation related to the model i of a predictor vector x is calculated by subtracting the f OSC factors:

$$x_0^i = x - P^i(k)W^i(k)^T x, \quad (11)$$

The PLS model $B^i(k)$ for the local model i is now calculated between the OSC deflated predictors x_0^i and the original responses y . Since the kernel algorithm is to be used, the covariance matrix

estimate $\hat{x}_o^i(k)$ of the deflated predictor x_o^i has to be known. From (11) the outer product is

$$\begin{aligned} x_o^i x_o^{iT} &= (x - P^i(k)W^i(k)^T x)(x - P^i(k)W^i(k)^T x)^T \\ &= (I - P^i(k)W^i(k)^T)xx^T(I - W^i(k)P^i(k)^T) \end{aligned} \quad (12)$$

which gives

$$\hat{R}_{xxo}^i(k) = (I - P^i(k)W^i(k)^T)\hat{R}_{xx}^i(k)(I - W^i(k)P^i(k)^T). \quad (13)$$

Because $\hat{R}_{xoy}^i(k) = \hat{R}_{xy}^i(k)$ [27], the PLS model $B^i(k)$ for each local model can be calculated using the matrices $\hat{R}_{xxo}^i(k)$ and $\hat{R}_{xy}^i(k)$ with a kernel PLS algorithm. In this study, the modified kernel algorithm #2 presented in [28] was utilized.

2.3.2. Prediction

After the adaptation step, each local model has the up-to-date OSC weights $W^i(k)$ and loadings $P^i(k)$, and PLS model $B^i(k)$. Until the next XRF measurement $y(k+1)$ is received, the rMM model is used to predict the responses for the unknown VNIR spectra to fill in the gaps between the sparse XRFs.

Whereas the amount of updating of each local model is based on the distance between the response sample value $y(k)$ and the model mean μ_y^i (see (4) and (5)), the validity of each model for prediction has to be evaluated using the predictor (x) sample, simply because the correct response value y is unknown.

The OSC preprocessing is a key factor in selecting the correct model for prediction. Instead of using the raw x sample to calculate the model validities, the OSC deflated predictor x_o is utilized. This remarkably improves the model selection performance since the disturbing variations are first eliminated from the data.

For a new predictor sample x the OSC deflation related to model i , x_o^i , is obtained using (11), and the corresponding PLS response estimate is

$$\hat{y}^i = B^i(k)x_o^i. \quad (14)$$

To compare how well the deflated sample x_o^i belongs to the local model i , the distribution of the deflated predictor samples for each local model is assumed to be normal. Now the probability density function value in x_o^i gives the validity value for the sample, i.e. a Gaussian validity function is used also for the model estimation phase. However, since the covariance matrix $\hat{R}_{xxo}^i(k)$ is typically ill-conditioned and cannot be inverted, only the main principal component directions of the distribution are utilized. Assuming that the n_o main eigenvectors of $\hat{R}_{xxo}^i(k)$ are collected in the columns of matrix $V^i(k)$, the PCA score covariance matrix is

$$R_{tto}^i(k) = V^i(k)^T \hat{R}_{xxo}^i(k) V^i(k). \quad (15)$$

If t_o^i contains the PCA projected x_o^i , the validity value is obtained as

$$\begin{aligned} \rho_x^i(k) &= ((2\pi)^{n_o} \det(R_{tto}^i(k)))^{-1/2} \\ &\quad \times \exp[-(t_o^i - \mu_{t_o}^i(k))^T R_{tto}^i(k)^{-1} (t_o^i - \mu_{t_o}^i(k))], \end{aligned} \quad (16)$$

where $\mu_{t_o}^i(k)$ is the OSC deflated (11) predictor mean (3) projected to the PCA base:

$$\mu_{t_o}^i(k) = V^i(k)^T (\mu_x^i(k) - P^i(k)W^i(k)^T \mu_x^i(k)). \quad (17)$$

The final rMM prediction for the sample x is calculated as the weighted average of the local response estimates (14):

$$\hat{y} = \left(\sum_i \rho_x^i(k) \right)^{-1} \sum_i \rho_x^i(k) \hat{y}^i = \sum_i \tilde{\rho}_x^i(k) \hat{y}^i, \quad (18)$$

where $\tilde{\rho}_x^i(k)$ is the normalized prediction validity of model i .

2.3.3. Final rMM algorithm

To clarify the calculation of the proposed rMM algorithm, the key steps of the model updating and sample prediction are collected here.

Model adaptation for each local model i given a new sample pair $\{x(k), y(k)\}$:

1. Calculate the forgetting factor $\lambda^i(k)$ using Eqs. (4) and (5).
2. Update the covariance matrix estimates $\hat{R}_{xx}^i(k)$ and $\hat{R}_{xy}^i(k)$ using Eqs. (1) and (2).
3. Update the predictor mean $\hat{\mu}_x^i(k)$ using Eq. (3).
4. Calculate the OSC weights $W^i(k)$ as the f most important eigenvectors of $M^i(k)\hat{R}_{xx}^i(k)$.
5. Calculate the f OSC loadings $P^i(k)$ using Eq. (10).
6. Calculate the OSC deflated predictor covariance matrix estimate $\hat{R}_{xxo}^i(k)$ using Eq. (13).
7. Use the covariances $\hat{R}_{xxo}^i(k)$ and $\hat{R}_{xy}^i(k)$ in the kernel PLS algorithm [28] to obtain the local PLS model matrix $B^i(k)$.

Prediction calculation given a new predictor x :

1. For each local model i , calculate the deflated predictor x_o^i and local response estimate \hat{y}^i using Eqs. (11) and (14).
2. For each local model i , collect the main eigenvectors of $\hat{R}_{xxo}^i(k)$ in the columns of $V^i(k)$ and calculate the model validity $\rho_x^i(k)$ using Eqs. (15) and (17).
3. To obtain the final estimate \hat{y} , combine the local estimates as a weighted average using Eq. (18).

3. Experimental

The data for this study were obtained from the final zinc concentrate slurry flow in the concentration plant of Inmet Mining Corporation's Pyhäsalmi mine, located in Finland. The spectra were measured using an imaging spectrograph (SpecIm ImSpector V10) having the wavelength range of 400–1000 nm and nominal spectral resolution of 5 nm. The spectrograph was connected to a monochrome CCD camera (Basler A102f), and the spectra were collected using an ordinary desktop computer. The on-line measuring was performed by leading a constant slurry flow through a jet flow cell with a sapphire window. The window was illuminated with a regular halogen light source and the spectrum of the light reflected from the slurry was measured. The XRF data were collected from the XRF analyzer of the concentrator plant (Outotec Courier®6 SL). For more information on the equipment used in data collection, please refer to [6].

In Pyhäsalmi, the solids in the zinc concentrate typically contain about 55% zinc, 33% sulfur, 8% iron and 0.5% copper, which are the elements continuously measured by the XRF analyzer. The solids content of the flow is around 30–40%. However, during a process failure the zinc content decreases, whereas the iron and copper contents may significantly increase. In this study, the main focus is on modeling the copper contents where the relative changes are the largest.

The rMM algorithm was written and the modeling performed in Matlab R2007b (The Mathworks Inc.) running on an ordinary desktop computer. All the Matlab code used in the rMM algorithm was written for this study, except the kernel PLS algorithm that was adapted from [28]. The OSC preprocessing of the data for the reference PLS calculations was done with the PLS Toolbox 4.0.2 (Eigenvector Research, Inc.).

3.1. Collected data

In total about 432,000 VNIR spectra and 4500 XRF values were collected during the measurement period of 50 days. The data col-

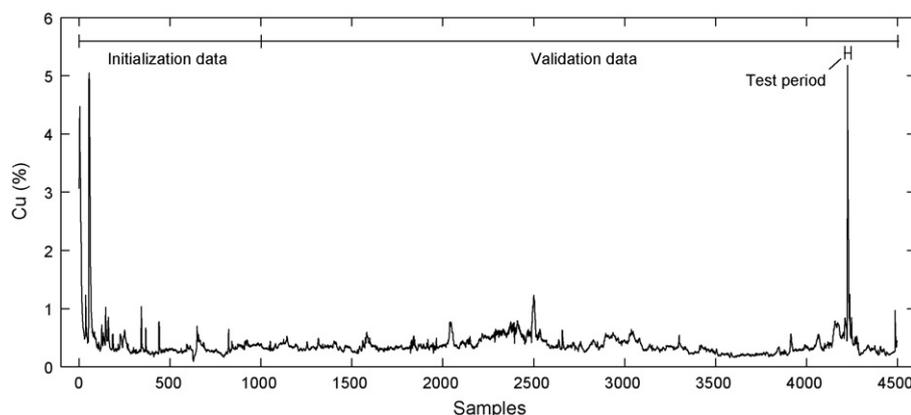


Fig. 2. Copper content variation in the final zinc concentrate during the 50 days sampling period as measured by the XRF analyzer.

lection was performed continuously using the sampling interval of 10 s for the spectrum and the normal sampling interval of the XRF analyzer (about 16 min) for the XRF values. A small number of samples were omitted due to disturbances or maintenance breaks either in the XRF analyzer or in the spectrum measurement equipment. To reduce the amount of data and speed up the computations, the measured wavelength range was averaged to 20 values reducing the dimension of the spectrum from 960 wavelength values to 20. This was detected not to decrease the modeling performance [6].

The collected matching XRF and spectrum sample pairs (4500) were divided into two parts: the first 1000 samples were reserved for initializing the models, whereas the rest of the samples (3500) were utilized in model performance validation (Fig. 2). The beginning of the initialization data contained two short but major process failures, where the copper concentration measured by the XRF analyzer reached as high values as 4.5–5%. Near the end of the validation data set a third process failure with a very sharp increase in the copper content was detected.

3.2. Model validation

All recursive modeling was performed maintaining the temporal ordering of the samples to simulate the application of the models in on-line process monitoring. For validating the different calibration methods, the root mean square error (RMSE) for all the validation samples was calculated using always the prediction given by the “old” model, i.e. the model updated with all the samples preceding the current sample. Thus, the RMSE describes well the general prediction ability that can be achieved in real applications.

However, since the main interest of the spectral analysis is to detect sudden major process changes as soon as possible, also the RMSE during the copper content increase related to the process failure in the validation data was used to compare the modeling performances. This error (RMSE2) was calculated as the root mean square prediction error for the eight validation samples where the copper content was greater than 1.5% due to the process failure. These samples are referred in the following as the test period (Fig. 2).

4. Results and discussion

4.1. Model estimation

The VNIR spectrum data were initially preprocessed to mean centered absorbance values ($\log(1/R)$, where R is the measured reflectance). For testing, also the normal OSC preprocessing was

calculated using the mean centered absorbance values of the initialization data. These preprocessed data sets are referred in the following as ‘ $\log(1/R)$ ’ and ‘ $\text{OSC}(\log(1/R))$ ’, respectively. The recursive multimodel OSC and PLS modeling method (rMM) presented in Section 2.3 was applied to the ‘ $\log(1/R)$ ’ data. As a reference, a regular PLS model as well as a regular recursive PLS model (rPLS) were applied to the both data sets, and the results were compared to the rMM modeling results. For the PLS model, all the 1000 initialization samples were utilized for estimation.

In the rMM approach, two local models centered in the response space in $\mu_y^1 = 0.5\%$ and $\mu_y^2 = 5\%$ were utilized, and the width parameter τ in the response validity functions (4) was $\tau = 2.5$. The locations and width of the local model regions were selected so that the first model could handle the normal operation conditions with low copper values and the second model the unusually high copper levels related to process failures. The nominal recursive forgetting factor in rMM was $\lambda_0 = 0.95$ and the forgetting factor in rPLS was $\lambda = 0.95$. The number of factors in both static and rMM OSC deflation was $f = 1$. More OSC factors were also tested, but the modeling results were not improved. The number of PCA latent variables that were used to invert the deflated predictor covariance matrix (15) was selected so that the condition number of $\hat{R}_{t_{ro}}^i(k)$ was not too high. With eight latent variables the maximum condition number was kept around 10^6 , which ensured that no numerical problems occurred when inverting the matrix.

The number of latent variables in each model was optimized according to the RMSE and RMSE2 values. Fig. 3 shows the modeling performances with respect to the latent variable space dimension. In the rMM method the specified number of latent variables was used in both local PLS models. Clearly the two ordinary PLS models give poor modeling results for the whole validation data regardless of the number of latent variables. On the other hand, increasing the latent dimension improves their modeling performance during the test period. The rPLS models give generally better predictions, but they are not able to match the modeling performance of the rMM method even with a high number of latent variables.

According to the results presented in Fig. 3, six latent variables were selected for the rMM approach, ten for the both rPLS models and thirteen for the both PLS models for the further analysis. The differences in the number of optimal latent variables is easily explained by the fact that the local models in rMM describe simpler local dependencies whereas the rPLS models are required to handle the whole data range. The regular PLS models on the other hand are trained with a data set collected during a long time period, so that they attempt to model also the temporal changes in the relation of the VNIR spectra and XRF measurements, which further increases the required latent dimension. It also seems clear that the

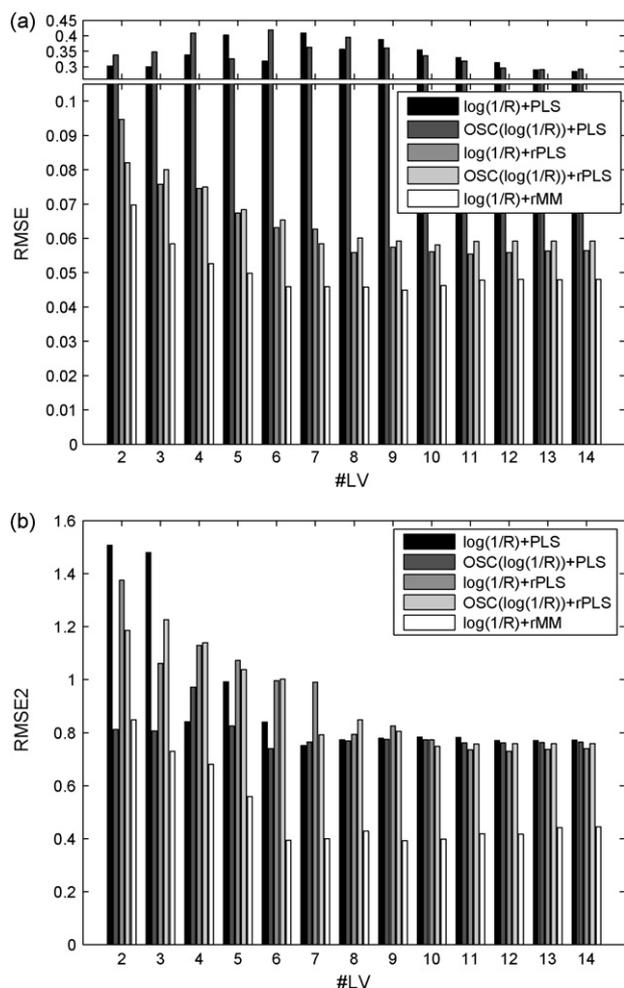


Fig. 3. Root mean square errors for the whole data (a) and the test period containing only a process failure (b) with respect to the number of latent variables (#LV).

OSC preprocessing is not significantly reducing the required number of latent variables in the PLS and rPLS models, thus indicating that the static OSC preprocessing is also suffering from the temporal changes in the data distributions.

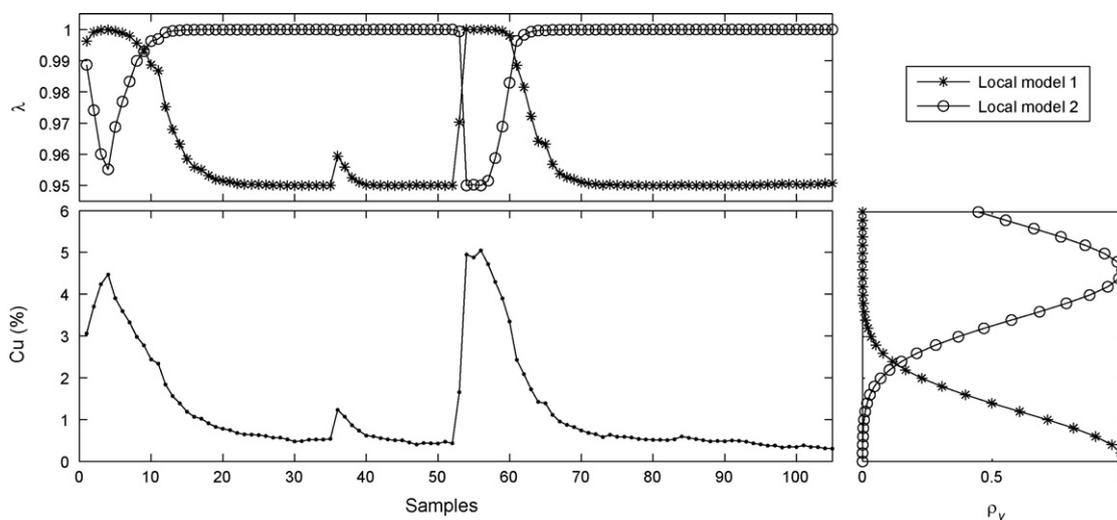


Fig. 4. The beginning of the recursive rMM estimation. On the bottom left are shown the first 110 XRF samples with the two process failures, above that is the behaviour of the forgetting factors (5) during the adaptation and on the right are the response validity functions (4) for the two local models.

Table 1

Comparison of the modeling performances.

Preprocessing	Model	#LV	Q^2	RMSE	RMSE2
log(1/R)	PLS	13	0.27	0.290	0.77
OSC(log(1/R))	PLS	13	0.22	0.291	0.76
log(1/R)	rPLS	10	0.92	0.056	0.77
OSC(log(1/R))	rPLS	10	0.92	0.058	0.75
log(1/R)	rMM	6	0.95	0.046	0.40

To analyze the adaptation of the local models in the rMM model, the first 110 data points containing the two process failures are shown in the bottom part of Fig. 4 with the corresponding forgetting factors $\lambda^i(k)$ (5) above them. Additionally, the validity functions (4) used to determine the forgetting factor values are drawn on the right. The forgetting factor values indicate how much each model is adapted with the corresponding data point: Clearly the local model 2 is adapted ($\lambda^2(k)$ is less than one) only when the copper values are large, and the local model 1 is adapted ($\lambda^1(k)$ is less than one) when the copper values are in the normal region. According to Fig. 4, the about 20 process failure samples with $\lambda^2(k)$ less than one are used for the adaptation of the second local model. Since the copper values stay small after that (Fig. 2), practically no adaptation of the second model takes place until the third process failure (test period) near the end of the whole data set is reached. This means that the information of the large copper values is stored in the local model 2. On the other hand, the local model 1 is adapted continuously after the process failures and thus it gives accurate predictions in the normal operating region.

4.2. Modeling performance

The RMSE, RMSE2 and Q^2 values calculated for each model are presented in Table 1. In general, the regular PLS models are not able to model the validation data and give considerably higher error values than the rPLS and rMM models. This clearly shows the advantage of recursive modeling when the models are to be used for longer periods on real process data. Mainly due to its better performance on the test period samples, the rMM model is capable of modeling the whole validation data better than the rPLS models. Because the static OSC preprocessing before utilizing PLS or rPLS is practically not affecting the overall modeling results in the following only the models with the 'log(1/R)' data are analyzed.

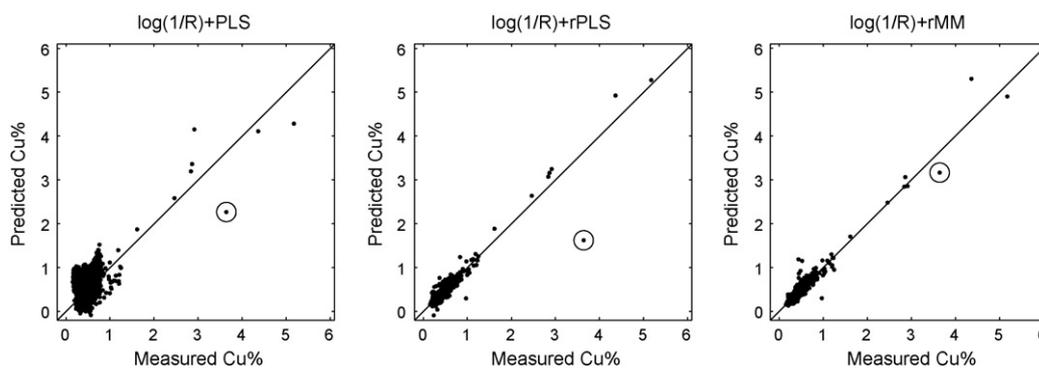


Fig. 5. Scatter plots of the model predictions for the mean centered absorbance validation data. The first test period sample corresponding to the increase in the copper content during the test period is circulated.

When comparing the test period performance related to the rapid increase in the copper content (RMSE2) there is a great difference between the models. Fig. 5 shows the scatter plots for the PLS, rPLS and rMM models with the 'log(1/R)' preprocessing and highlights the most critical first test period sample. rMM is able to decently predict the first test sample, whereas the rPLS fails completely. This is simply because the recursive updating has caused it to forget the existence of the rare high copper content values. The estimation data set of the PLS model contained also the two process failures, and since no adaptation has been performed, it gives a little better prediction for the first process failure sample. On the other hand, in the rMM model the second local model in rMM was adapted only to the large copper values, and can therefore predict this first test sample.

After the rPLS and rMM models are adapted with the first test period sample, the rest of the large copper values are predicted by them with approximately equal accuracy. This shows that the rPLS model can quickly adapt to the change in the process; however, it is not capable of predicting the change beforehand, which is the key point when the VNIR spectrum measurements are to be utilized in the flotation plant monitoring and control.

4.3. Predicting the intermediate points

The real benefit of the VNIR spectrum measurements in the presented application of assaying the flotation slurries is the small sampling interval. When compared with the traditional XRF-based grade measurements which in the test case have a sampling interval of about 16 min, the spectral sampling with a 10 s interval can be considered practically continuous. With an up-to-date calibration model, the unknown spectra can be used to on-line predict the elemental contents of the slurry between the XRF measurements.

Fig. 6 shows the copper content predictions calculated from the spectra with rMM and rPLS during the beginning of the third process failure (test period). The predictions are calculated on-line so that when a new XRF measurement with the corresponding spectrum is obtained, the models are adapted. After that, the adapted models are used to predict the received unknown spectra (every 10 s) until the next XRF measurement is received.

The rMM prediction detects the rapid increase in the copper level about 5 min before the first large XRF measurement arrives. Since the prediction is calculated on-line and involves only light computations, the copper value estimate is available for the automatic process control system and human operator within seconds after the corresponding spectrum is measured. In this case the XRF measurement is received quite soon after the real process failure begins, whereas in a worst case situation a 15 min delay would be possible. However, even as it is the rMM prediction confirms here

the copper level increase since it continuously analyzes the slurry flow, thus ruling out the possibility of a single outlier in the XRF measurements. Additionally, the prediction clearly indicates that after the first large XRF measurement the copper level continues to increase, and that some actions definitely are required either by the process operator or the control system.

In the upper part of Fig. 6 the normalized prediction validity $\tilde{\rho}_x^1(k)$ of the local model 1 in rMM is shown as a function of time. It describes the importance of the first local model in the final prediction given by the rMM (see (18)). Since only two models are used, it holds that $\tilde{\rho}_x^2(k) = 1 - \tilde{\rho}_x^1(k)$. Initially the low copper values are mainly predicted by the first local model. However, when the increase in the copper content takes place around 11:56, the weights very rapidly change so that the rMM prediction is given by the second local model. This indicates that now the predicted VNIR spectra locate completely in the subset of the second model, and it should be used for the prediction.

Before the process failure there are also a couple of spectra that are not solely predicted by the first local model, even though the copper level stays quite low. However, since no corresponding XRF measurements are available, it is not possible to evaluate the correctness of these predictions.

When comparing the rMM and rPLS predictions for the unknown spectra (Fig. 6), the advantage of the local model structure

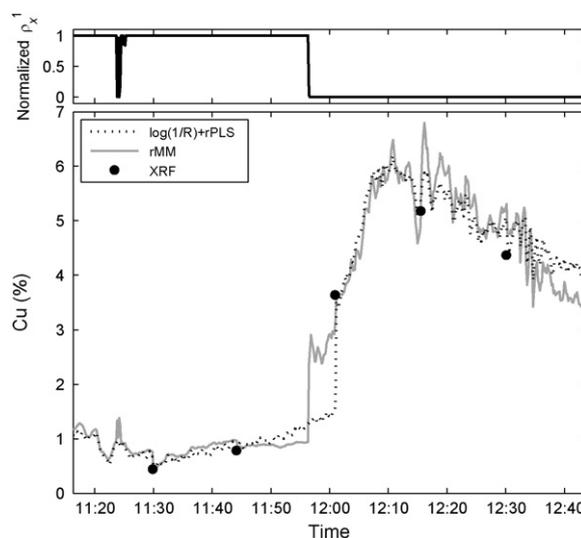


Fig. 6. The rMM and rPLS on-line predictions of all the spectra during the beginning of the process failure. The upper part shows the behaviour of the normalized predictor validity of local model 1 used to weigh the local model predictions in rMM.

is evident. Since the rPLS model does not remember the last high copper values seen in the beginning of the initialization data (about 47 days ago), it is unable to detect the new process failure until a confirming XRF measurement is received and the model is adapted with it. Even though the rPLS model then predicts the following large copper values correctly, it does not improve the detection of new process failures from the VNIR data with respect to the XRF analyzer. On the other hand, the rMM model using the same data for adaptation and giving equally good predictions for the normal copper range predicts the process failure as soon as it takes place.

5. Conclusions

In this work, the prediction of mineral flotation slurry contents based on VNIR reflectance spectrum measurements has been studied as a supplemental assaying method for XRF analyzers. It was shown by comparing the prediction performances of regular PLS and rPLS approaches during a long measurement period that a regular PLS model is not able to maintain a sufficient accuracy. It was further shown that due to the recursive adaptation and exponential forgetting, an rPLS model is not capable of correctly predicting the sudden large content changes caused by rare process failures. However, for the flotation process control and monitoring, these changes should be detected as soon as possible to minimize the effects of the failures.

To manage also the process failure situations, a new recursive multimodel approach (rMM) has been introduced. It is demonstrated that the performance of the proposed rMM approach is at least equal to that of the rPLS model in normal operation conditions. Additionally, due to the selectively updated local PLS models, rMM is capable of predicting the process failure situations where the slurry content values increase rapidly and unexpectedly. This enables the faster detection of the possible process failures that may require immediate corrective control actions. The presented rMM algorithm can also be generalized to other applications of the similar type.

Acknowledgements

The help of the personnel of Pyhäsalmi Mine Oy and Outotec Oy is gratefully acknowledged.

References

- [1] B.A. Wills, T. Napier-Munn, *Wills' Mineral Processing Technology*, 7th edition, Butterworth-Heinemann, 2006.
- [2] J. Gebhardt, W. Tolley, J. Ahn, Color measurements of minerals and mineralized froths, *Miner. Metall. Process.* 10 (2) (1993) 96–99.
- [3] J. Oestreich, W. Tolley, D. Rice, The development of a color sensor system to measure mineral compositions, *Miner. Eng.* 8 (1/2) (1995) 31–39.
- [4] G. Bartolacci, P. Pelletier, J. Tessier, C. Duchesne, P.-A. Bossé, J. Fournier Jr., Application of numerical image analysis to process diagnosis and physical parameter measurement in mineral processes—part I: Flotation control based on froth textural characteristics, *Miner. Eng.* 19 (6–8) (2006) 734–747.
- [5] J. Liu, J. MacGregor, C. Duchesne, G. Bartolacci, Flotation froth monitoring using multiresolutional multivariate image analysis, *Miner. Eng.* 18 (1) (2005) 65–76.
- [6] O. Haavisto, J. Kaartinen, H. Hyötyniemi, Optical spectrum based measurement of flotation slurry contents, *Int. J. Miner. Process.* 88 (3–4) (2008) 80–88.
- [7] S. Qin, Recursive PLS algorithms for adaptive data modeling, *Comput. Chem. Eng.* 22 (4/5) (1998) 503–514.
- [8] L. Ljung, T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MIT, 1983.
- [9] K.J. Åström, B. Wittenmark, *Adaptive Control*, 2nd edition, Prentice Hall, 1994.
- [10] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, *Chemom. Intell. Lab. Syst.* 44 (1998) 175–185.
- [11] B. Dayal, J. MacGregor, Recursive exponentially weighted PLS and its applications to adaptive control and prediction, *J. Process Control* 7 (3) (1997) 169–179.
- [12] K. Helland, H. Berntsen, O. Borgen, H. Martens, Recursive algorithm for partial least squares regression, *Chemom. Intell. Lab. Syst.* 14 (1991) 129–137.
- [13] S. Wold, Exponentially weighted moving principal components analysis and projections to latent structures, *Chemom. Intell. Lab. Syst.* 23 (1994) 149–161.
- [14] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.* 74 (369) (1979) 829–836.
- [15] T. Naes, T. Isaksson, Locally weighted regression in diffuse near-infrared transmittance spectroscopy, *Appl. Spectrosc.* (1) 46 (1992) 34–43.
- [16] Z. Wang, T. Isaksson, B.R. Kowalski, New approach for distance measurement in locally weighted regression, *Anal. Chem.* 66 (2) (1994) 249–260.
- [17] A.M.C. Davies, H.V. Britcher, J.G. Franklin, S.M. Ring, A. Grant, W.G. McClure, *Mikrochim. Acta* 94 (1–6) (1988) 61–64.
- [18] J.S. Shenk, M.O. Westerhaus, Investigation of a LOCAL calibration procedure for near infrared instruments, *J. Near Infrared Spectrosc.* (4) 5 (1997) 223–232.
- [19] T. Fearn, A.M.C. Davies, Locally-biased regression, *J. Near Infrared Spectrosc.* 11 (6) (2003) 467–478.
- [20] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Non-linear regression methods in NIRS quantitative analysis, *Talanta* 72 (1) (2007) 28–42.
- [21] L. Aarhus, Nonlinear empirical modeling using local PLS models, 1994, <http://heim.ifi.uio.no/ftp/publications/cand-scient-theses/LAarhus/the%25sis.pdf>.
- [22] S. Schaal, C.G. Atkeson, Constructive incremental learning from only local information, *Neural Comput.* 10 (8) (1998) 2047–2084.
- [23] C.G. Atkeson, A.W. Moore, S. Schaal, Locally weighted regression for control, *Artif. Intell. Rev.* 11 (1–5) (1997) 75–113.
- [24] S. Schaal, C.G. Atkeson, S. Vijayakumar, Real-time robot learning with locally weighted statistical learning, in: *Proceedings of the International Conference on Robotics and Automation*, IEEE Press, 2000, pp. 288–293.
- [25] S. Vijayakumar, A. D'Souza, S. Schaal, Incremental online learning in high dimensions, *Neural Comput.* 17 (12) (2005) 2602–2634.
- [26] O. Haavisto, H. Hyötyniemi, Recursive partial least squares estimation of mineral flotation slurry contents using optical reflectance spectra, in: *Proceedings of the 11th Conference on Chemometrics in Analytical Chemistry*, vol. 2, 2008, pp. 81–85.
- [27] T. Fearn, On orthogonal signal correction, *Chemom. Intell. Lab. Syst.* 50 (2000) 47–52.
- [28] B. Dayal, J. MacGregor, Improved PLS algorithms, *J. Chemom.* 11 (1) (1997) 73–85.