J. Kaartinen and H. Hyötyniemi. 2005. Combining multi-camera-data of flotation circuit with PCA and PLS. In: Graeme J. Jameson (editor). Proceedings of the Centenary of Flotation Symposium. Brisbane, Australia. 6-9 June 2005, pages 121-125.

Combining Multi-Camera-Data of Flotation Circuit with PCA and PLS

J Kaartinen[1] and H Hyötyniemi[2]

ABSTRACT

In this paper multivariate data analysis performed on a new multi-camera system implemented to zinc flotation circuit of Pyhäsalmi mine is presented. Image-analysis-based system was developed to take advantage of the commonly known fact that the changes in the flotation process are reflected to the visual appearance of the froth surface. Most often the changes are seen first in the froth and later in other process measurements.

The image variables are already utilised in closed loop control and considerable savings are achieved with the current setting, but the results coming from the separate cells are used individually. This is clearly not the optimal way to take advantage of the many measurements coming from different locations of the process. A better approach would be for example to use the measurements coming from the early stages of the process to predict the behaviour (or improve the quality of measurements) in the later stages. The methods described in this paper are aimed to improving the situation in this sense.

INTRODUCTION

Pyhäsalmi mine is located in Finland, some 500 km north from Helsinki. The main minerals produced are copper (0.8 per cent), zinc (2.8 per cent), sulfur (37.0 per cent) and iron (33.0 per cent). Also, there are small amounts of gold and silver present in the ore (Hätönen, 1999). Flotation is divided into three sections, where copper, zinc and sulfur are processed.

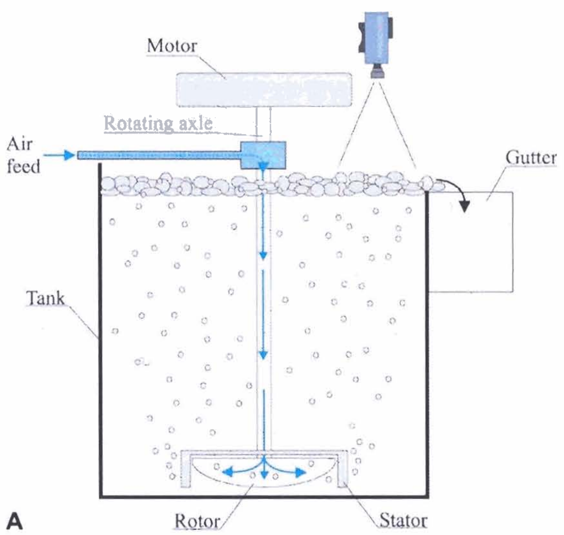
Research at Pyhäsalmi started in 1997 with an EU-funded project called 'The characterization of flotation froth structure and colour by machine vision' (ChaCo). After the ChaCo-project ended the research was continued with a Finnish project called VÄSY, where the single-camera system developed earlier was extended to multi-camera-system.

1. Helsinki University of Technology, Control Engineering Laboratory, PO Box 5500, 02015 HUT, Finland. Email: Jani.Kaartinen@hut.fi
2. Helsinki University of Technology, Control Engineering Laboratory, PO Box 5500, 02015 HUT, Finland. Email: Heikki.Hyotyniemi@hut.fi

The structure of the measurement system is shown in Figure 1. The camera is placed as near the edge of the flotation cell as possible so that the froth under the imaging area would characterise the properties of the outflowing material. The camera is located inside a protective hood, which protects not only the camera against dirt but also the imaging area against ambient light coming from the flotation hall. This is important since a big portion of the different image analysis algorithms used in the system are using the total reflectance point on top of each bubble as a basis for further calculations. Thus, the illuminating halogen lamp has to be located as near the optical axis of the camera as possible and no other light sources are allowed. Also, the intensity level of the illumination is stabilised by using uninterrupted power supply and by converting the electricity from alternating to direct current (Kaartinen, 2001).

THE MULTI-CAMERA SYSTEM

As mentioned before, based on the good results obtained by using the single camera for several years the system was extended to cover the whole zinc circuit. The current system consists of six cameras that are mounted on top of six different flotation cells of the circuit as shown in Figure 2. With every camera, various variables from the froth image(s) can be calculated as explained in the next section. Since there are large amounts of data available of these individual cells three different methods were tested in order to get the overall picture of the flotation process. This was done by combining the separate image analysis results and applying different sensor fusion techniques. The tested methods were principal component analysis (PCA), principal component regression (PCR) and partial least squares (PLS) (Sharma, 1996). All of these are multivariate statistical methods and they are very suitable in the cases when trying to extract the relevant variables out of vast amounts of multivariate data.



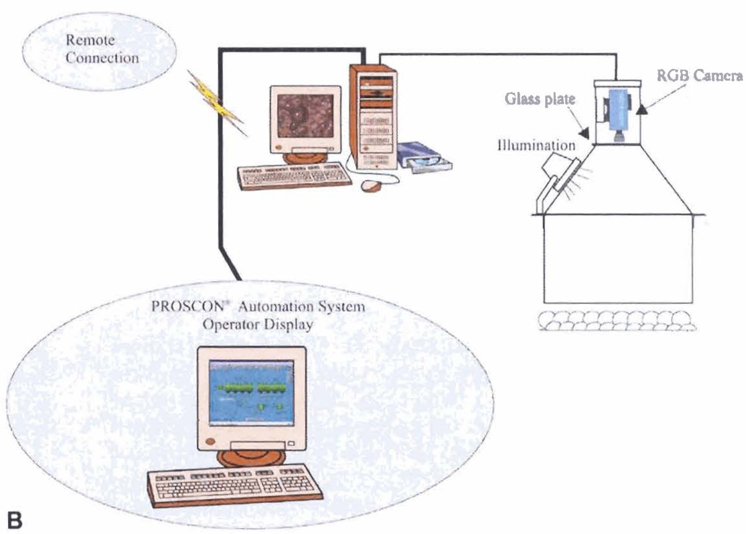


FIG 1 - (A) Location of the camera on the flotation cell, and (B) basic setup and connections for a single camera.

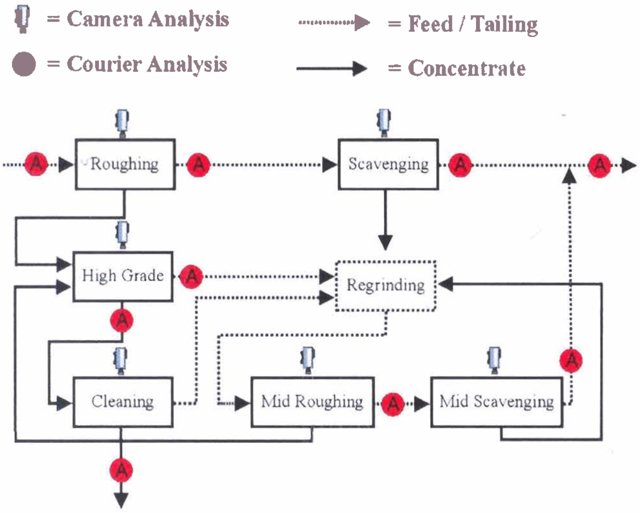


FIG 2 - Location of cameras and x-ray fluorescence (Courier) analysis points in the zinc circuit.

IMAGE VARIABLES

From each camera an image pair is taken so that time difference between the two images is about 20 ms. From these images all the necessary variables can be calculated. There must be two consecutive images in order to get information from the dynamical features of the froth (eg froth speed, bubble collapse rate, etc). Static features (eg mean bubble size, number of bubbles, etc) are calculated from the first image only. Currently it takes approximately 0.5 seconds to process one image pair resulting in a sample interval of three seconds for a single cell. An example of a froth image from the rougher circuit is shown in Figure 3.

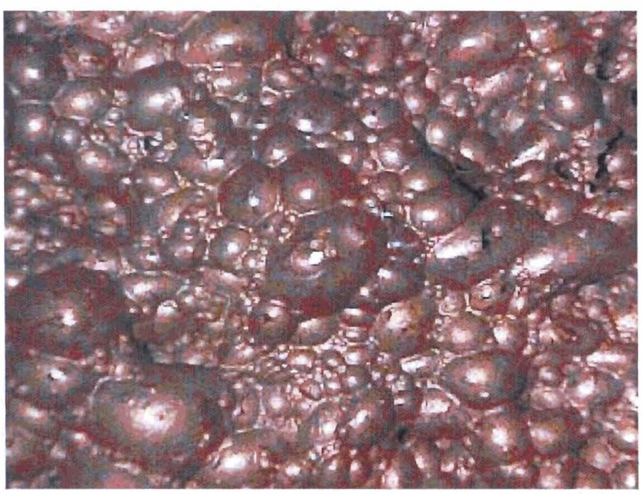


FIG 3 - An example of a froth image.

It is possible to get approximately 60 - 70 different variables from each image but there is considerable redundancy in these measurements. A good example is different colour plane representations of the image; the grabbed image can be presented both in RGB (red, green, blue) and HSV (hue, saturation, value) colour planes and if for all these six variables the first four moments are calculated, 24 different measurements are obtained to describe the colour statistics of a given image. However, it is clear that these measurements are highly redundant, and hence it is important to pick up only the most important image variables for the end user (Kaartinen and Koivo, 2002). Another possibility – which is used in this paper – is the use of intelligent data compression techniques; see also Hyötyniemi (1999) and Hyötyniemi and Ylinen (2000).

Currently the most interesting variables measured from the flotation froth of the zinc circuit are *speed of the froth*, *mean bubble size*, *bubble collapse rate*, *redness of the froth* (since redness correlates with the amount of zinc in the froth) and *transparency* of the froth (ie the load variable, see Miettunen *et al* (2001).

DATA ANALYSIS

In the data analysis there were two independent data sets collected during 26 August 2004 - 1 September 2004 (data1) and 6 September 2004 - 10 September 2004 (data2). The actual data sets consist of 96 variables collected in six minute intervals (averaged from one minute data). Eventually 18 image variables from three cells were selected for further analysis. The cells being studied were roughing, mid roughing and cleaning. Roughing and cleaning are an obvious choice since their impact on the whole circuit is considerable. The results from the mid roughing circuit were also consistent with the other data, but high grade results – although expected to be useful – could not be utilised since the high grade flotation tanks in Pyhäsalmi are such that the flow under the camera is not continuous. This more or less ruins that data. The image data from scavenging was available and it was used in testing but was eventually dropped since it did not improve the accuracy of the estimation of the final product properties. The selected image variables were the same for each cell and are presented in Table 1.

TABLE 1

The selected image variables.

Variable	Description
Correlation	Peak value of cross correlation matrix of an image pair
Redness	Mean-value of red-component of an RGB-image
Load	Load-variable that estimates the mineral content in the surface of the bubble (see Miettunen *et al*, 2001)
Speed	Speed of the froth (obtained from peak value position in the cross correlation matrix)
Bubble size (BS)	Mean bubble size in the image (obtained from segmentation results)
Intensity	Mean intensity of the image

The delays between different cells were estimated and removed from both data sets in a same way. The delays were (in six minute time steps): From rouging to mid roughing four, from roughing to cleaning four and from roughing to measured zinc content of the final product (ZnR Zn per cent) six time steps. The appropriate delays were estimated partly from the data and partly by using the process knowledge of the plant engineers at Pyhäsalmi. The delays are somewhat tricky since the nature of the flotation process is quite complex and there are many internal feedback flows in the circuit. This leads to changing time delays. Because the tested methods are static and time invariant, they are very sensitive to the fact that the time delays are removed properly. Therefore the changing time delays introduce unpredicted error to the results of the following analysis.

THE PCA AND PCR APPROACH

This section describes the principal component analysis (PCA) and principal component regression (PCR) analysis. Both of these methods are multivariate statistical methods and they are very suitable in the cases when trying to extract the relevant variables out of vast amounts of multivariate data. There were two goals; the first goal was to identify correlations (principal components) in the data from different cells and compare them. The second goal was to evaluate the practical relevance of the image analysis: how well can one predict the zinc concentration of the final product by using only the image variables. The analysis was carried out by using MATLAB® and its PLS toolbox (Wise and Gallagher, 1998), where all the needed calculation routines are readily available.

Principal components

By using the image variables described above, the first and second principal components were calculated separately for each cell. The idea was to compare the different cells and see if they would behave similarly.

The analysis showed that, mathematically speaking, the three cells work systematically. Figure 4 illustrates this by showing the loadings of the first two principal components.

As can be seen, the correlations between the different variables are always in the same direction and surprisingly consistent in the first principal component. However, when similar analysis was performed on the second data set (data2), the results were not so evident, but clearly the same kind of behaviour can be seen (see Figure 5).

The amount of variance captured from the whole data is presented in the legends of the figures. The first two principal components were selected since on the average they cover 84 per cent of the total variance in the data.

Estimators

Since the analysis in the previous chapter clearly showed a consistent behaviour between the selected flotation cells, the next step was to try to predict the zinc percentage of the final product. This was done by using only the selected image variables and nothing else. The Courier analysis was used as a reference to validate the estimate.

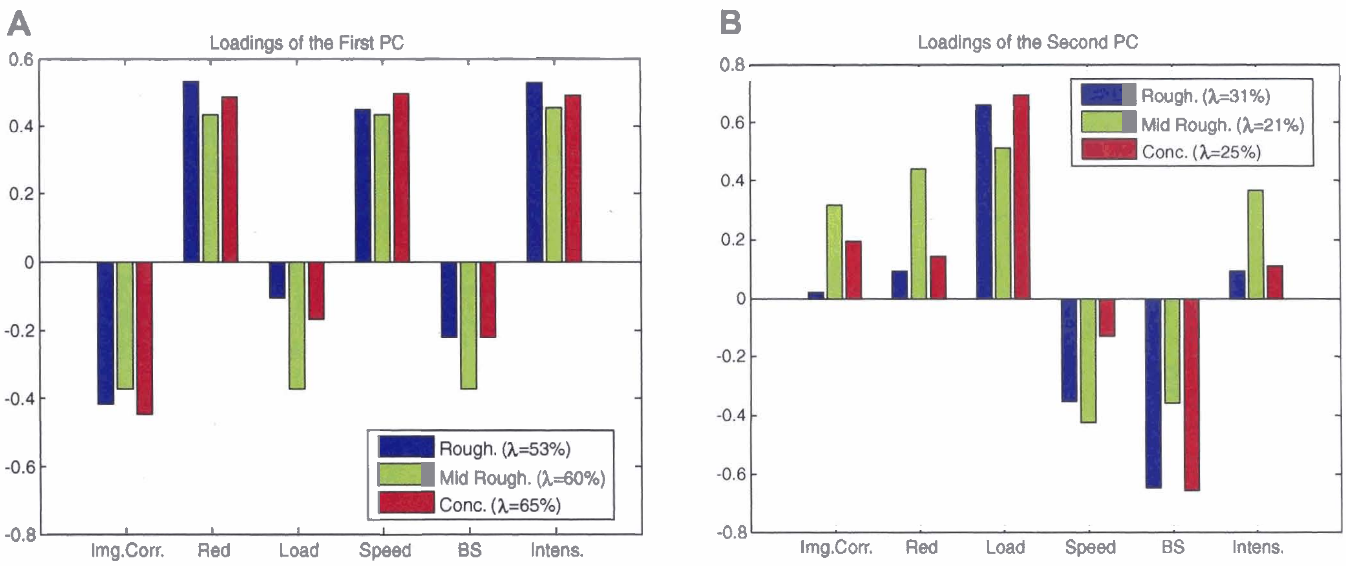


FIG 4 - (A) Loadings of the first principal component (data1), and (B) loadings of the second principal component (data1).

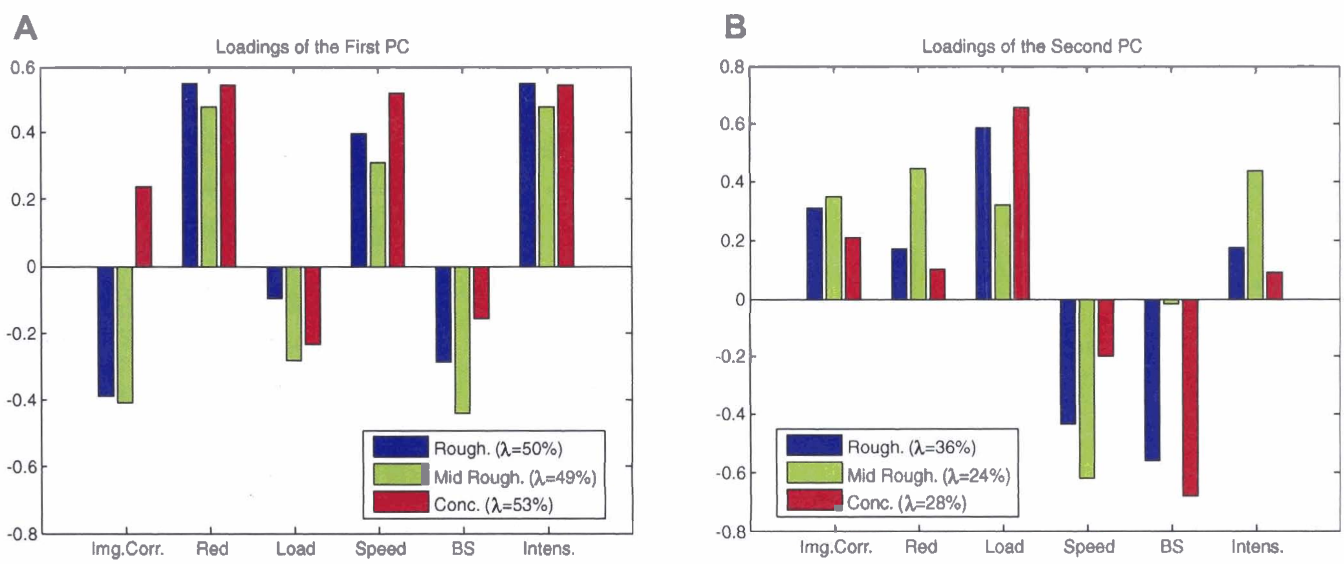


FIG 5 - (A) Loadings of the first principal component (data2), and (B) loadings of the second principal component (data2).

The PCA and PCR estimates were obtained by sliding a 48-hour history window from which the PCA model was updated at each step. After the model was calculated, the last values of the score vectors of both principal components were stored. Then the analysis was advanced by one step. The analysis cycle is presented below:

1. take the last 480 points of history data,
2. calculate the first two principal components for that data,
3. calculate the score vectors of those principal components,
4. store the final values of the score vectors, and
5. wait for the next data point, slide the history window and go to one.

Surprisingly, the score values of the first principal component obtained this way (see Figure 6a) followed the zinc content of the final product with remarkable accuracy ($R^2 = 0.86$). This is very interesting since only image variables were used to get this result. This was a good proof of the power of image analysis in the context of mineral flotation. Also, since the time delay between the image variables from the cleaning cell and the Courier analysis of the final concentrate is on the average 12 minutes, this means that the PCA approach is able to predict the zinc content in advance.

The principal component regression (PCR) was also used by combining the two score vectors. Data1 was used as a teaching data and data2 as a validation data. As can be seen in Figures 6b and 7b, the results were improved only slightly for the teaching data and degraded – because of the changing time delays and different operating point – for the validation data. This means that it is essentially the first principal component alone that correlates with the output concentration.

The same analysis was performed on the second data set and the results were similar, although not as good as with the first data set. Figure 7 illustrates the resulting curves as well as the squared correlation coefficients.

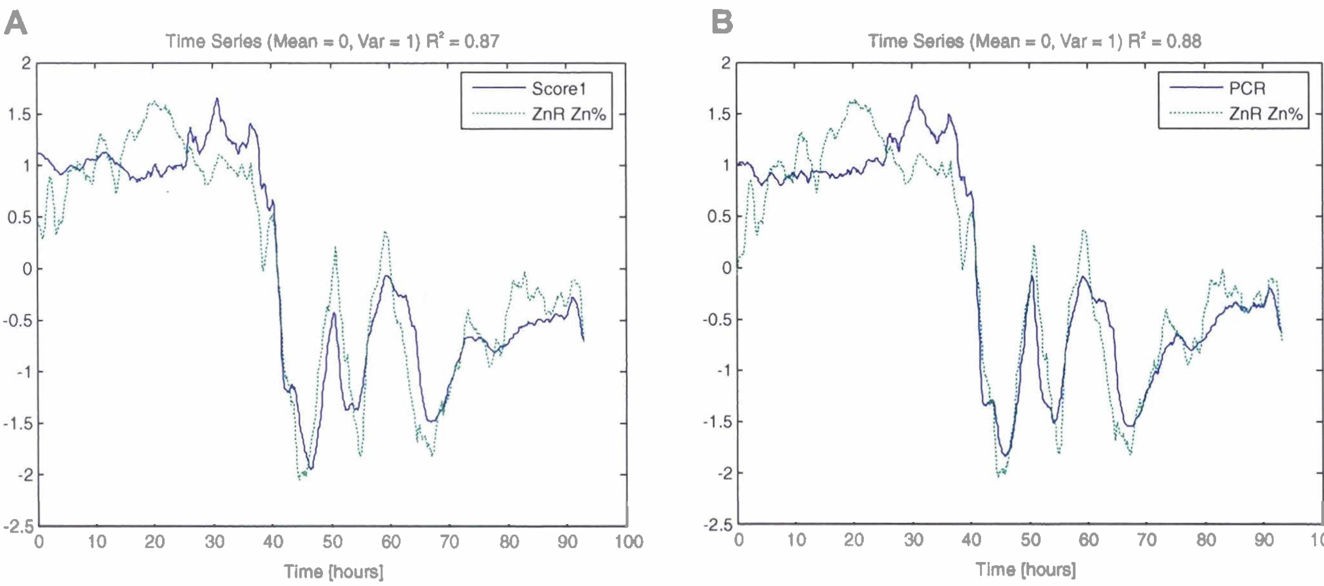


FIG 6 - (**A**) Score values of the first principal component, PCA approach (data1), and (**B**) PCR approach, teaching data (data1).

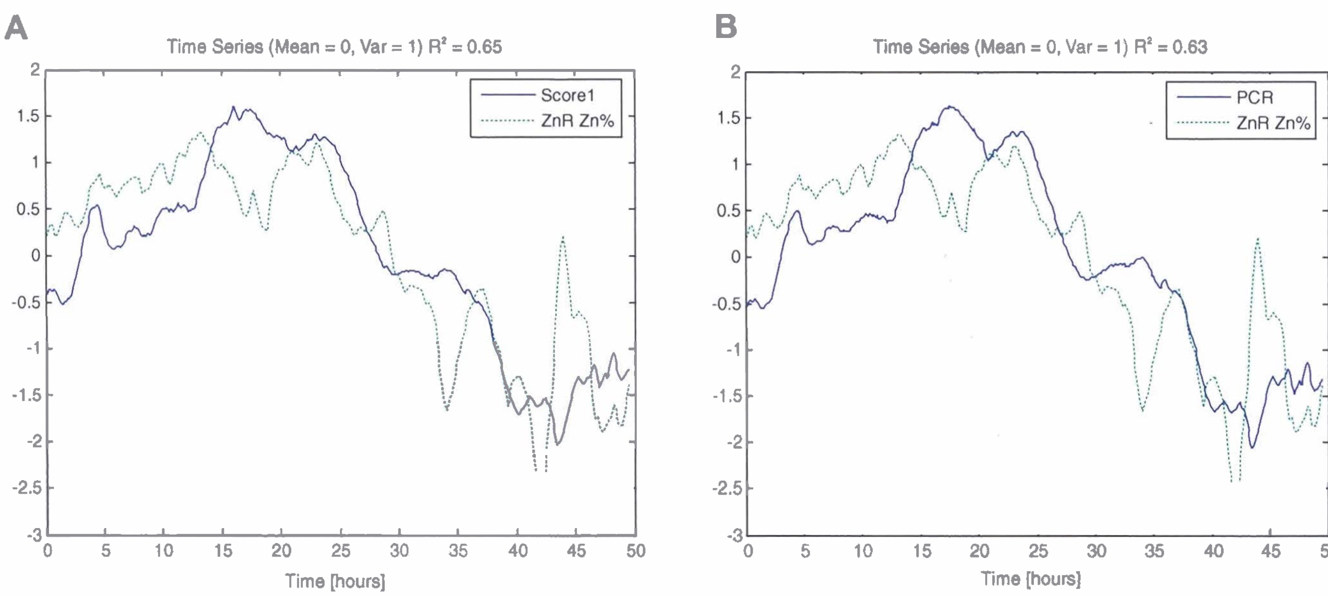


FIG 7 - (**A**) Score values of the first principal component, PCA approach (data2), and (**B**) PCR approach, validation data (data2).

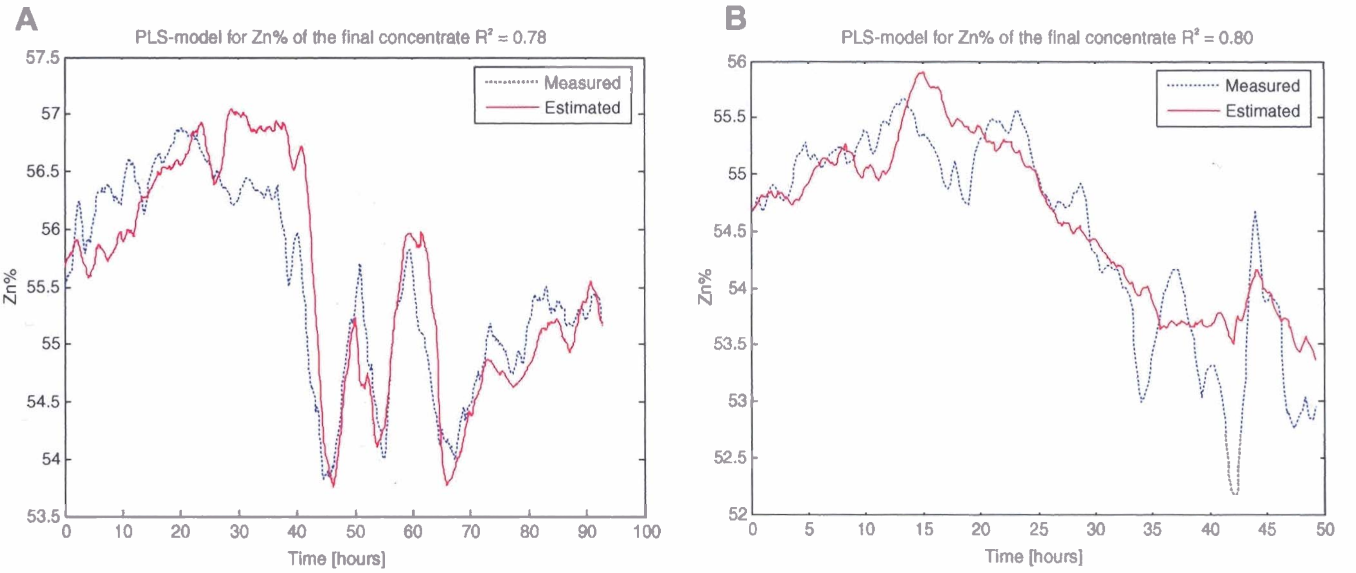


FIG 8 - (A) PLS model for data1, and (B) PLS model for data2.

THE PLS APPROACH

Finally, a PLS model was tested in a similar fashion as the PCA and PCR models above. The same 48-hour sliding history window was used as a teaching data and since the *y*-side of teaching data (ie the zinc content of the final product) is actually coming two time steps later than the *x*-data (the image variables), the estimate can be obtained 12 minutes in advance.

The results of the PLS analysis are shown in Figure 8. One interesting thing is the degradation in the correlation with the first data set, since one would assume improvement – at least when compared to PCA approach – because of the additional *y*-side information. This is due to the changes in the time delay between the image variables and the Courier analysis of the final product.

For the second data set the results are improved as shown in Figure 8b. However, even if the improvement is clear in terms of correlation coefficients it is still only minor improvement when compared to the PCA approach (Figure 7a), which accomplished more or less the same thing completely without the aid of the *y*-side data.

FUTURE RESEARCH

Since such promising results were obtained in this study, the goal for future research is to develop these results to be applied in on-line control of the flotation process. One interesting possibility is to complement the Courier analysis using image analysis: The x-ray analyses are now obtained only once for every 20 minute period and having accurate estimates between these analyses would be important. It seems that if the models based on image analysis are calibrated to match the past x-ray analyses, the information gaps in the measurements can be filled. This issue will be approached applying the static methods (PCR and PLS) as well as dynamic multivariate methods, such as *subspace identification*.

CONCLUSIONS

The image analysis equipment currently installed at Pyhäsalmi mine as well as the most important image variables were presented in this paper. However, the main focus was on the multivariate data analysis that was performed on two independent data sets that were collected in the fall 2004. Although the methods described here are time invariant and the process delays clearly are not, still the results obtained were speaking strongly in favour of image analysis in the control of flotation process. Although the benefits of image analysis have been reported before (eg Guarini *et al*, 1995; Moolman *et al*, 1995; Cipriano *et al*, 1998; Miettunen *et al*, 2001), the data analysis performed in this paper clearly shows that the visual information obtained from the flotation froth, by itself, is able to characterise the state of the flotation process. This gives the research group confidence and motivation to continue with image analysis and to derive new implementations of closed loop control.

REFERENCES

Cipriano, A, Guarini, M, Vidal, R, Soto, A, Sepulveda, C, Mery, D and Briseno, H, 1998. A real time visual sensor for supervision of flotation cells, *Minerals Engineering*, 11(6):489-499.

Guarini, M, Cipriano, A, Soto, A and Cuesalaga, A, 1995. Using image processing techniques to evaluate the quality of mineral flotation process, in *Proceedings 6th International Conference on Signal Processing, Applications and Technology*, Boston, 24 - 25 October.

Hätönen, J, 1999. Image analysis in mineral flotation, Helsinki University of Technology, Control Engineering Laboratory, Report 116, Finland.

Hyötyniemi, H, 1999. GGHA Toolbox for Matlab, Helsinki University of Technology, Control Engineering Laboratory, Report 117, Finland.

Hyötyniemi, H and Ylinen, R, 2000. Modeling of visual flotation froth data, *Control Engineering Practice*, 8(3):313-318.

Kaartinen, J, 2001. Data acquisition and analysis system for mineral flotation, Helsinki University of Technology, Control Engineering Laboratory, Report 126, Finland.

Kaartinen, J and Koivo, H, 2002. Machine Vision Based Measurement and Control of Zinc Flotation Circuit, *Int Journal of Studies in Informatics and Control*, 11(1):97-105.

Miettunen, J, Kaartinen, J and Hätönen, J, 2001. Image analysis based control of zinc flotation, in *Proceedings Fourth Regional Symposium on Computer Applications in the Mineral Industries (APCOM 2001)*, Tampere, Finland, 3 - 5 September, pp 267-275.

Moolman, D W, Aldrich, C, van Deventer, J S J and Bradshaw, D B, 1995. The interpretation of flotation froth surfaces by using digital image analysis and neural networks, *Chem Eng Sci*, 50(22):3501-3513.

Sharma Subhash, 1996. *Applied Multivariate Techniques* (John Wiley and Sons, Inc: New York).

Wise, B M and Gallagher, N B, 1998: PLS_Toolbox 2.0 for use with MATLAB™, Eigenvector Research, Inc.