**Proceedings of the 17th World Congress**
**The International Federation of Automatic Control**
**Seoul, Korea, July 6-11, 2008**

# Prediction of Concentrate Grade in Industrial Gravity Separation Plant – Comparison of rPLS and Neural Network

**Remes A.*, Vaara N. **, Saloheimo K. ***, Koivo H. ***

**Helsinki University of Technology,*
*department of Automation and systems Technology,*
*Control Engineering Laboratory,*
*P.O.Box 5500, FIN-02015 TKK, Finland (e-mail: heikki.koivo@tkk.fi)*
*** Outokumpu Tornio Works, Kemi Mine P.O.Box 172, FIN-94101 Kemi, Finland*
**** Outotec Minerals, P.O.Box 84, FIN-02201 Espoo, Finland*

**Abstract:** Control of the concentrate quality is usually one of the main targets in the operation of mineral concentrator processes. Availability of the estimates of end product properties in advance – based on upstream process measurements – offers an opportunity to develop higher level control strategies for the unit processes. Here, the recursive PLS and adaptive neural network models are compared in the prediction of the concentrate grade at a gravity separation plant. The methods are applied in the Outokumpu Tornio Works Kemi Mine plant data. The chromite concentrate grade can be predicted relatively accurately based on the slurry properties measured in the grinding circuit. Accordingly, the predicted chromite grade decreases about 0.2 %-units when the slurry D50 passing size is increased by 10 %. This enables further development of the grinding control, especially the control of the slurry particle size, to meet the concentrate specifications. 

**Keywords:** gravity separation, grinding, recursive PLS, neural network, process modelling

## 1. INTRODUCTION

When operating complex industrial processes, on-line modelling is a useful tool. Updated models are needed for early reaction to disturbances that affect the process efficiency and end product quality through the process chain.

In gravity mineral separation plants there are usually only few control variables. The slurry particle size distribution substantially affects the separation performance. Changes in the ore mineralogy introduce the majority of the process disturbances. Predictive plant models can offer tools for adjusting grinding conditions in advance, to keep the process in predefined targets, especially for final concentrate quality.

A number of modelling studies for mineral processing plants have been reported previously. The applicability of *partial least squares* (PLS) modelling method for mineral processing data was demonstrated using Brunswick mine grinding and flotation data (Hodouin *et al*., 1993). Afterwards, Dayal and MacGregor (1997a) showed the *recursive* PLS (rPLS) method to be much better when compared to recursive least squares algorithm, using Brunswick's sulphide flotation data as well. The same case process was also used for *adaptive neural network* modelling with good results, despite that the variable selection was considered intractable (Forouzi and Meech, 1999). More recently the concentrate grades of a flotation plant have been modelled by applying a dynamic ARMAX (autoregressive moving average with exogenous inputs) model (Casali *et al*., 2002). Gonzalez *et al*. (2003) compared several model types and structures for the copper grade of the Codelco Andina flotation plant. They concluded that (linear) PLS and (nonlinear) neural networks are nearly equally good in their prediction ability. The PLS algorithm was considered to be good for variable selection – also when neural networks are applied.

The use of PLS algorithm for identifying dynamic models was already suggested by Hodouin *et al*. (1993). However, the rPLS and adaptive neural network methods with dynamic model structures have not been widely studied in mineral processing plants. Casali *et al*., (2002) suggested that the dynamic models, when used in prediction of the concentrate grade, could be used as a part of control strategy.

In this study the advantages of continuously adapted predictive models are discussed in contrast to non-adaptive models. The linear recursive PLS, both with constant and variable forgetting factors, are compared with non-linear neural network structures with adaptive training. The results are evaluated using a concentrator plant case, and the applicability of the model types is assessed.

## 2. APPLIED ADAPTIVE PREDICTION METHODS

The partial least square (PLS) regression is widely used methods for linear model parameter estimation. A good description of the method can be found in Wold *et al*. (2001).

Since the studied industrial process is time variant, adaptive predictive models, both linear and non-linear, were applied

here. The selected techniques were a recursive partial least squares method and an adaptive neural network.

### *2.1 Recursive Partial Least Squares Regression*

Recursive partial least squares regression was first introduced by Helland *et al*. (1992). In this study, the kernel-base recursive PLS algorithm presented by Dayal and MacGregor (1997b) was applied. The adaptation is based on the update of the PLS covariance matrices when a new observation ($\boldsymbol{x}_t$ and $\boldsymbol{y}_t$) is available. The old data are exponentially discounted with the forgetting factor $\lambda_t$ by updating the (unscaled) covariance matrices $(\boldsymbol{X}^T\boldsymbol{X})_t$ and $(\boldsymbol{X}^T\boldsymbol{Y})_t$ as follows (Dayal and MacGregor, 1997b):

$$\left(X^T X\right)_t = \lambda_t \left(X^T X\right)_{t-1} + x_t^T x_t \qquad (1)$$

$$\left(X^T Y\right)_t = \lambda_t \left(X^T Y\right)_{t-1} + x_t^T y_t \,. \qquad (2)$$

Additionally, the forgetting factor can be adjusted continuously, to only discount the old data when the process is persistently excited, thus containing some new information. The variable forgetting factor can be calculated, as shown by Fortescue *et al.* (1981), with

$$\lambda_t = 1 - \frac{\left[1 - x_t \left(X^T X\right)_t x_t^T\right] e_t^2}{\sigma_o^2 N_o}\,, \qquad (3)$$
$$\lambda_t = \lambda_{\min} \quad if \quad \lambda_t < \lambda_{\min}$$

where $\sigma_o^2$ is the expected measurement noise variance of the output variable, $N_o$ is the nominal asymptotic memory length (determining the adaptation speed) and $e_t$ is the error between the PLS estimate and the measurement.

### *2.5 Adaptive MLP Neural Networks*

Neural networks are common structures in modelling of the non-linear complex processes. *Multilayer perceptron* (MLP) neural networks consist of an input layer, one or more hidden layers and an output layer. The hidden and output layer contains computation *nodes* with *activation functions* that are often sigmoidal type, thus introducing the non-linearity. The network input layer can also contain past output estimates as an input, forming a *recurrent dynamic network* structure. The network is trained with data in supervised manner with back-propagation algorithm, fixing the *weights* of the neurons. In adaptive learning, the training passes are performed continuously when a new data set becomes available. For more information on the neural networks the reader is referred e.g. to Lin and Lee (1996) and Haykin (1999).

## 3. DESCRIPTION OF THE OUTOKUMPU TORNIO WORKS KEMI PLANT

### *3.1 Kemi Chromite Concentrator*

The Outokumpu Tornio Works Kemi Mine is located in Northern Finland. The concentrator processes 1.2 Mt of chromite ore annually. The products are upgraded lumpy ore and metallurgical grade concentrate. The concentrate is produced using gravity and high-gradient magnetic separation methods, preceded by a rod mill - ball mill grinding stage. The gravity separation circuit includes Reichert cones and spiral separators. The performance of the separation process is strongly dependent on the feed slurry properties, especially on the particle size distribution. The flow sheet of the Kemi concentrator plant is shown in Fig. 1.

The Kemi Mine is integrated to the ferrochrome smelter of Outokumpu Tornio Works, at 20 km distance from the mine. At the Outokumpu Tornio Works high chromite grade of the concentrate is advantageous for ferrochrome production. Therefore the main operating goal at the Kemi Mine is to maximize the chromite content of the concentrate, used subsequently in the ferrochrome smelters, while keeping the concentrate production rate in a predefined value. Hence accurate prediction of the product grade, based on the grinding circuit slurry properties, gives a good basis for process and production management.

### *3.1 Applied Process Data*

The selected predictive models were compared using two data sequences. The time series consist of 74 and 79 samples of ten-minute average data, respectively. The output variable is the concentrate grade $HR_{Cr2O3}$ (expressed in $\%Cr_2O_3$), measured by an on-belt XRF analyzer after the drum filter. The selected input variables were the feed slurry chromite on-line assay ($\%Cr_2O_3$) ($TMT_{Cr2O3}$) and the on-line analysis of the 50 % passing size of the particles (μm) (*D50*), measured from the grinding circuit. In addition, to describe the ore in terms of grindability, the Bond operating work index (kWh/t) (*WIo*) was calculated by applying (Napier-Munn *et al.*, 2005)

$$WIo = \frac{W}{10\left(\frac{1}{\sqrt{P_{80}}} - \frac{1}{\sqrt{F_{80}}}\right)}, \qquad (4)$$

where W (kWh/t) is the work input of the grinding mills and $P_{80}$ and $F_{80}$ (μm) are the 80 % passing sizes of the grinding circuit product slurry and the ore feed.

Furthermore the 40 minutes (equal to 4 samples) process delay between the input and output data was compensated by shifting the data in time. The data was normalized to zero average and unit variance, and median filtering was applied for noise reduction. During the data collection, the grinding circuit control variables – the ore feed rate and the rod mill rotation speed – were varied stepwise to enhance the excitation of the data. The combined data sequences are shown in Fig. 2., and the mean values and the standard

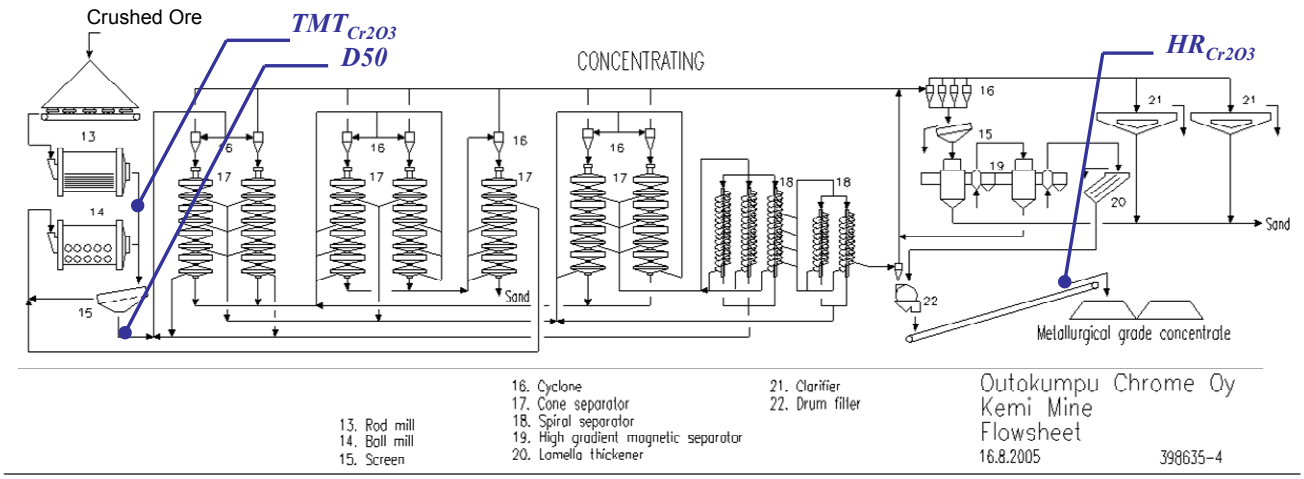


Fig. 1. Flowsheet of the metallurgical grade concentrate plant at the Outokumpu Tornio Works Kemi concentrator.

deviation are in Table 1. Autocorrelation of the concentrate grade HRCr2O3 (%) after one and two sample lags is 0.99 and 0.97; indicating relatively slow process dynamics.

**Table 1. Statistics of applied data**

| | Average | Standard deviation |
|---|---|---|
| $HR_{Cr2O3}$ (%) | 45.20 | 0.17 |
| $TMT_{Cr2O3}$ (%) | 26.32 | 0.90 |
| *WIo* (kWh/t) | 10.34 | 0.89 |
| *D50* (µm) | 66.89 | 4.71 |

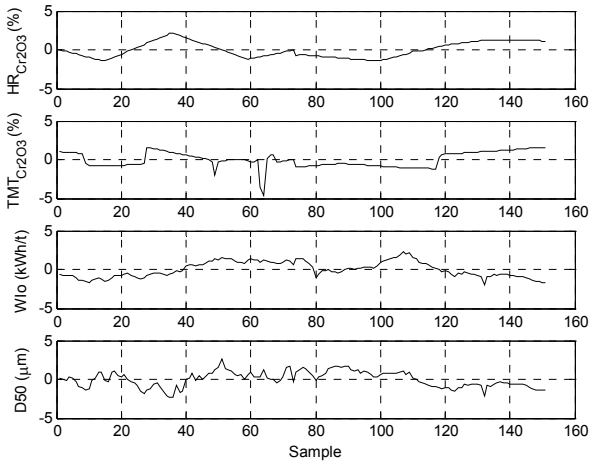


Fig. 2. Applied scaled process data from the Kemi concentrator; the concentrate grade ($HR_{Cr2O3}$ (%)) is the output variable and the input variables, measured in the grinding stage, are: the chromite assay ($TMT_{Cr2O3}$ (%)), grinding work index (*WIo* (kWh/t)) and the 50% passing size of the particles (*D50* (µm)).

## 4. RESULTS

The aim of the modelling was to predict the concentrate grade after the 40 minutes process delay, based on the prevailing grinding circuit measurements. Firstly, non-recursive PLS models were studied with the two data sequences. According to the cross-validations, two latent variables is the best selection, resulting in root mean square errors of cross-validation (RMSECV) to be 0.17 (data 1) and 0.30 (data 2). Likewise, the $R^2$ values were 0.45 and 0.80 respectively.

### *4.1 Recursive Partial Least Squares Models*

The performance of the raw PLS model presented above implies that the process conditions are varying significantly, causing prediction error and motivating use of model adaptation algorithms. The two data sequences were merged to form a sequence including totally 153 samples. As the two data sequences are from two very different situations, a change in the process conditions takes place certainly after the sample number 74. The modified kernel algorithm with recursive updates of the covariance matrices (1 and 2) was applied to the data. In addition, effect of the adaptation of the forgetting factor $\lambda_t$ (3) was studied. For adaptation of $\lambda_t$ the effective memory length $N$ was set to 10 (standing for 1.67 hours time slot), and the expected measurement noise of the output variable $\sigma_o^2$ was set to 0.04. Minimum value limit for the forgetting factor was set to 0.85. First 15 data samples were used for calculation of the initial values of covariance matrices.

The recursive PLS yields notably better measures of fit when compared to the non-recursive version. Additionally, the non-unity forgetting factor enhances the prediction performance. In terms of $R^2$, maximum absolute error and standard error of prediction (SEP) the adaptive forgetting factor yields the best results. The measures of fit for different forgetting factors of the rPLS models are presented in Table 2.

**Table 2. Performance of recursive PLS with different forgetting factors $\lambda_t$**

| | $\lambda_t = 1$ | $\lambda_t = 0.95$ | Adaptive $\lambda_t$ |
|---|---|---|---|
| $R^2$ | 0.70 | 0.77 | 0.85 |
| Max. abs. error | 1.98 | 1.19 | 0.97 |
| Standard error of prediction | 0.59 | 0.52 | 0.44 |

The static rPLS models, shown in Table 2, result in relatively large variations of the regression coefficients, as shown in Fig. 3. This apparently indicates the lack of dynamics in the model. It turned out that in this case the most suitable dynamic model (in terms of fit statistics) is a relatively simple output error (OE) type model including, in addition to the input variables, a time delayed output estimate as a fourth input. A scheme of the model structure is shown in Fig. 4.

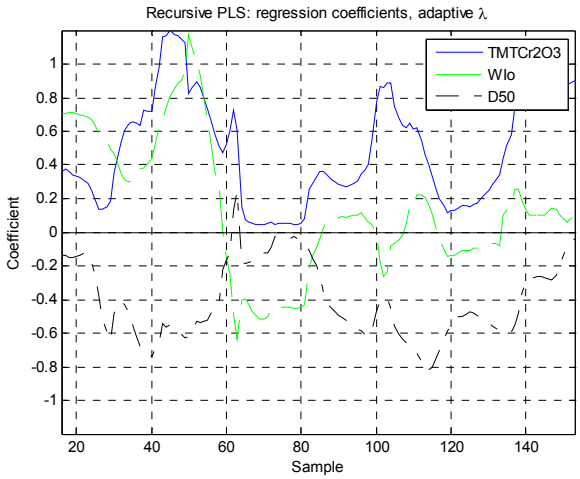


Fig. 3. Evolution of the regression coefficients of static rPLS model with adaptive forgetting factor.

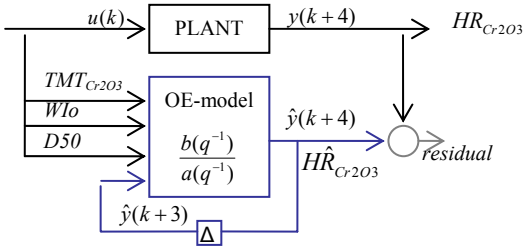


Fig. 4. A scheme of the output-error model structure; parameters *a* and *b* are the coefficients for the inputs and the precious output estimate respectively.

Next the recursive PLS was used in identification of the OE model parameters. Also in this case two latent variables were sufficient. The rPLS identified OE model yields a higher $R^2$ value and smaller maximum absolute error and standard error of prediction, when compared to non-dynamic rPLS (with adaptive $\lambda_t$). Also the prediction performance is still better when compared with the original data one lag autocorrelation. The performance of the dynamic rPLS models is shown in Table 3.

**Table 3. Performance of recursive PLS with adaptive $\lambda_t$ in identification of dynamic OE model**

| | 3 latent variables | 2 latent variables |
|---|---|---|
| $R^2$ | 0.99 | 0.98 |
| Max. abs. error | 0.38 | 0.36 |
| Standard error of prediction | 0.10 | 0.14 |

The time series of the measured concentrate grades together with predicted OE-rPLS grade estimates and the residuals are shown in Fig. 5 and the evolution of the model regression coefficients is shown in Fig. 6. The regression coefficients change abnormally between samples 60-80; this is probably due to a failure in the chromite assay slurry sampler, causing the sudden change in the measurement. This can be seen also in Fig. 2. Changes of the model adaptation rate in the recursive parameter update are shown in the forgetting factor plot of Fig. 7. According to the autocorrelation, the model residual is virtually white noise. In average, the model coefficients for the rPLS updated output error model are:

$$\hat{HR}_{Cr2O3}(k+4) = 0.185 \cdot TMT_{Cr2O3}(k) + 0.082 \cdot WIo(k) - 0.208 \cdot D50\ (k) + 0.698 \cdot \hat{HR}_{Cr2O3}(k+3) \quad (5)$$

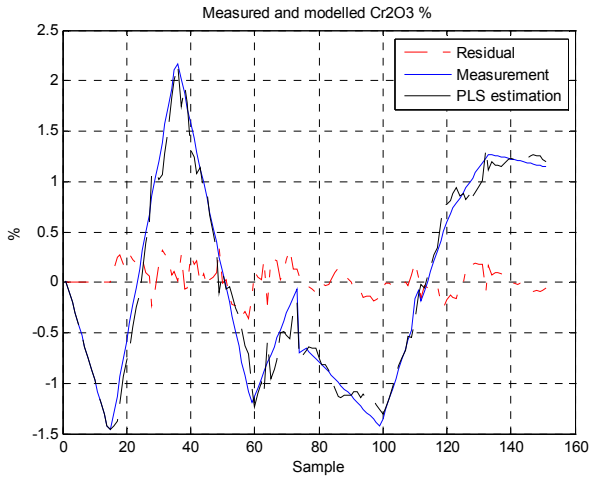


Fig. 5. Measured and predicted concentrate grades and the residuals when the dynamic OE-rPLS model with adaptive forgetting factor is applied.

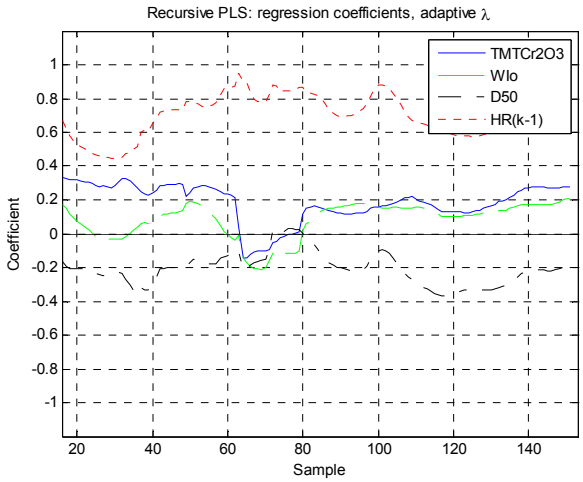


Fig. 6. Evolution of the regression coefficients of dynamic OE-rPLS model with adaptive forgetting factor.

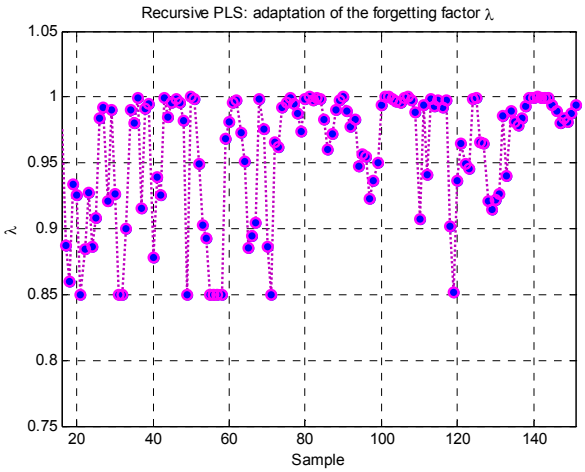


Fig. 7. Changes of the variable forgetting factor in OE-rPLS identification.

Table 4 summarizes the effects of changes in the model input variables on the predicted PLS model output (5), in terms of original unscaled $Cr_2O_3$ (%) grades. The listed numbers indicate the change of the model output resulting from a 10% increase of each input from its mean value, respectively. It can be seen that the feed chromite content ($TMT_{Cr2O3}$) and the particle size ($D50$) cause the largest responses on the estimate of the concentrate grade, but to opposite directions.

**Table 4. Effect of the input changes on the unscaled PLS estimates of the concentrate grade**

| Input variable | 10 % of the variable's mean | $\Delta\hat{H}R_{Cr2O3}$ (%) when the input change is 10 % of the mean |
|---|---|---|
| $TMT_{Cr2O3}$ | 2.6 | 0.26 |
| $WIo$ | 1.0 | 0.08 |
| $D50$ | 6.7 | -0.19 |

*4.1 Non-linear neural network models*

Finally, the same output error model type with same variable selection as in (5) was implemented using a non-linear multilayer neural networks. The network contained also a feedback connection enclosing the network output to the input layer. The network structure was one hidden layer including tangent sigmoid transfer function neurons, and one output neuron with linear transfer function. The network was trained using the backpropagation or Levenberg-Marguardt learning algorithm; the error goal was set to $1x10^{-8}$. The initial network learning was performed using with first sixteen data samples.

To determine the most appropriate network and adaptation configuration, the number of hidden neurons and number of adaptive learning passes for each new data vector were varied. Number of data passes enhances the network's ability to predict the process output. However, it is generally favourable to keep the network structure as simple as possible and at the same time to avoid overlearning of (noisy) measurement data. The networks were trained in adaptive manner using the data shown in Fig. 2. The performance of some tested variants is given in Table 5. In terms of $R^2$ Networks 2 and 4 (4 uses the Levenberg-Marguardt algorithm) are equally good. However, the Network 4 has a simpler structure having only 5 hidden neurons. Also, the same structure yields a good fit also with the backpropagation algorithm (Network 3). Thus the structure applied in the Networks 3 and 4 is the best fulfilling the requirements of data fit and simplicity of the structure.

**Table 5. Performance of adaptive neural networks with different configurations**

| | Network 1 | Network 2 | Network 3 | Network 4 -LM |
|---|---|---|---|---|
| Hidden neurons | 10 | 10 | 5 | 5 |
| Number of passes | 1 | 10 | 5 | 5 |
| $R^2$ | 0.80 | 0.99 | 0.97 | 0.99 |
| Max. abs. error | 1.08 | 0.20 | 0.41 | 0.52 |
| Standard error of prediction | 0.45 | 0.07 | 0.14 | 0.10 |

## 5. DISCUSSION

The comparison of the prediction performances of static non-recursive PLS with the recursive PLS algorithm, shown in Table 2, clearly points out the benefits of adaptive updating of the model parameter. Certainly, the adaptation is advantageous especially in mineral process modelling, since the process involves numerous unmeasured disturbances and the process operating conditions are highly time variant. In addition, the variable forgetting factor in the rPLS algorithm typically enhances the prediction performance, even though the difference was relatively small (see Table 2). However, by using the variable forgetting factor in the model, the effective memory length of the model adaptation is certainly more suitable for prevailing operating conditions. The prediction performance was further improved by introducing dynamics to the model. This can be seen also from the regression coefficients; in the dynamic case (Fig. 6) the input variable coefficients are more stable when compared to the static model (Fig. 3). The model accuracy can be still improved slightly by introducing non-linearity to the adaptive model. Nevertheless, the neural network model parameters cannot be interpreted so easily. For this reason the regression coefficient models are more practical, especially when the model purpose is, in addition to the prediction, to find out the effect of each input variable to the process output. The linear models can be also more robust in contrast to nonlinear models when unmeasured disturbances are present. As a future work, the model should be tested longer periods with process data from different normal operating conditions.

## 6. CONCLUSIONS

In this paper, the feasibility of the adaptive models in prediction of the concentrate grade in gravity separation plant at the Outokumpu Kemi concentrator was studied. Adaptive models are advantageous for the case process where a lot of unmeasured disturbances exist. The identification of dynamic output error (OE) model with recursive PLS – instead of static model – improves the prediction: then the maximum absolute deviation from the measured $Cr_2O_3$ (%) assay decreases from 0.17 to 0.06 percentage units (unscaled). Instead, the application of a non-linear neural network model did not cause any drastic improvements to the prediction performance. The selected OE-rPLS model type indicates the slurry particle size to be an important factor in estimation of the concentrate grade. For instance, by increasing the 50 % passing size of the grinding circuit outlet slurry (*D50*) by ten percent from the mean value decreases 0.19 % units of the resulting estimate of the chromite concentrate grade.

## REFERENCES

Casali, A., G. Gonzalez, H. Agusto and G. Vallebuona (2002). Dynamic simulator of a rougher flotation circuit for a copper sulphide ore, *Minerals Engineering*, **15**, pp. 253-262.

Dayal, B.S. and J.F. MacGregor (1997a). Recursive exponentially weighted PLS and its applications to adaptive control and prediction, *Journal of Process Control*, **7**, pp 169-179.

Dayal, S.B. and J.F. MacGregor (1997b). Improved PLS algorithm, *Journal of Chemometrics*, **11**, pp. 73-85.

Forouzi, S. and J.A. Meech (1999). An adaptive artificial neural network to model Cu/Pb/Zn flotation circuit, In: *IEEE Industry Applications Society, Advanced Process Control Applications for Industry Workshop 29-30 April 1999, Vancouver, Canada*, pp. 75-82.

Fortescue, T.R., L.S. Kerhenbaum and B.E. Ydstie (1981). Implementation of self-tuning regulators with variable forgetting factors, *Automatica*, **17**, pp. 831-835.

Gonzalez, G.D., M. Orchard, J.L. Cerda, A. Casali and G. Vallebuona (2003). Local models for soft-sensors in rougher flotation bank, *Minerals Engineering*, **16**, pp. 441-453.

Haykin, S. (1999). *Neural networks a comprehensive foundation* (2nd Ed.), Prentice Hall, New Jersey.

Helland, K., H. Berntsen, O. Borgen and H. Martens (1992). Recursive algorithm for partial least-square regression, *Chemometrics and intelligent laboratory systems*, **14**, pp. 129-137.

Hodouin, D., J.F. MacGregor, M. Hou and M. Franklin (1993). Multivariate statistical analysis of mineral processing plant data, *CIM Bulletin*, **86**, pp. 23-34.

Lin, C.-T. and C.S. Lee (1996), *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*, Prentice Hall, New Jersey.

Napier-Munn, T.J., S. Morrell, R.D. Morrison and T. Kojovic (2005). *Mineral comminution circuits, their operation and optimization*, JKMRC, Queensland.

Wolds, S., M. Sjöström and L. Eriksson (2001). PLS-regression: a basic tool of chemometrics, *Chemometrics and intelligent laboratory systems*, **58**, pp. 109-130.