ON-LINE MONITORING OF AQUEOUS BASE METAL SOLUTIONS WITH TRANSMITTANCE SPECTROPHOTOMETRY

M.J. Phiri and C. Aldrich

Research highlights

- o Conditions on a zinc plant in Namibia were simulated with Cu, Co and Zn solutions.
- o The metals could be measured online with an industrial NIR-vis spectrophotometer.
- o The transmittance spectra could be visualized with linear discriminant analysis.
- o PCR, PLSR and SVR models could estimate the metal concentrations accurately.

ON-LINE MONITORING OF AQUEOUS BASE METAL SOLUTIONS WITH TRANSMITTANCE SPECTROPHOTOMETRY

M.J. Phiri[a] **and C. Aldrich**[a,b]∗

[a]Department of Process Engineering, University of Stellenbosch, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa,

[b]Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines, Curtin University, Perth, WA, Australia

*Corresponding Author Email: chris.aldrich@curtin.edu.au

ABSTRACT

Transmittance spectrophotometry was used to monitor copper, cobalt and zinc in solution in laboratory experiments. The samples simulated plant conditions encountered on the Skorpion zinc mine in Namibia and were prepared using a simplex centroid mixture design. Principal component, partial least squares and support vector regression models were calibrated from visible and near infrared absorption spectra. All models could accurately estimate the concentrations of all the metals in solution. Although these models were affected by nickel contamination, the Cu models were less sensitive to this contamination than the Co and Zn models. Likewise, elevated temperatures led to degradation of the calibrated models, particularly the Zn models. The effects of these conditions could be visualized by a linear discriminant score plot of the spectral data.

Keywords: on-line analysis, modelling, hydrometallurgy, process instrumentation

1. Introduction

Efficient operation of hydrometallurgical plants often depends on routine analysis of valuable elements after every unit operation (Gaft et al., 2007). In this respect, the availability of on-line data is a critical element in process control and continuous process improvement, as suggested by Sowerby (2002) and Gaft et al. (2007). Since on-line measurements are typically non-intrusive, there is no interference with the process, and thus plant operations can run without interruption. In addition, on-line analysis reduces the health hazards that may be experienced during the collection of samples for laboratory analysis (Sowerby, 2002). Moreover, laboratory analysis could be subject to large sampling errors, whereas on-line methods usually provide high quality sampling data through the analysis of large continuous volumes of material (Gaft et al., 2007).

Several techniques are available for on-line analysis. These include on-stream analysers, image analysis and other optical sensors. In general, on-stream analysers are effective in monitoring and controlling plant processes, since they can determine the elemental composition of a sample with high accuracy (Remes et al., 2007). Optical fibre sensors in particular have shown considerable promise as a means to obtain elemental analyses in the areas of food processing, pharmaceuticals and mineral processing (Lewis et al., 2007). Even though the method of analysis is not the same in each case, the operating principles underlying the measurements are the same. Different wavelengths typically ranging from ultraviolet (UV) to near infrared (NIR) reflected from the sample are measured by an optical fibre sensor. These reflection patterns can usually be related to the characteristics of the sample material. Recent studies have shown that visible to near infrared (vis-NIR) spectra can be used as a supplementary information for accurate monitoring and control of large mineral plants (Haavisto and Kaartinen, 2009). The benefit is a potential increase in stability in plant operations that can lead to improved plant performance and higher financial returns (Haavisto et al., 2008). Despite its potential, online spectrophotometry is not a mature technology in the metallurgical industries yet, as some of the equipment may be sensitive to external disturbances and would therefore require a high degree of maintenance (Bergh et al., 2001).

The optical sensor used in this study was developed by Blue Cube Systems, based in Stellenbosch, South Africa and is among other used to monitor zinc purification in the Skorpion zinc mine in southwest Namibia. Although the sensor can be used as both a diffuse reflectance and a transmittance spectrophotometer, only the latter mode was considered in this study to simulate implementation on the plant as closely as possible. Zinc operations in the Skorpion complex comprise an open pit mine, mill, an electrohydrometallurgical refinery, as well as a melting and casting shop. The essential processing steps of zinc production consist of roasting of the zinc concentrates from the mine to convert the zinc sulphides into oxides. Sulphur dioxide is converted into sulphuric acid that is used to leach the zinc oxide to produce an impure zinc sulphate solution. The solution is subsequently purified by precipitation of copper with zinc dust in a first stage, to limit the copper concentration in solution to 0-2 mg/L. Other impurities are removed in a subsequent stage and the purified zinc sulphate solution is further processed by electrowinning, where zinc metal is recovered from electroplated aluminum cathodes.

At this site, the spectrometer is applied in the first stage of solution purification, where it is used to measure the concentration of copper in solution in the refinery. This requires periodic recalibration of the sensor, and the quality of the models can have a significant influence on the real time monitoring of the process. Moreover, the prediction of the copper concentration may be influenced by the presence of other metals in solution, such as cobalt and cadmium and in this investigation the effect of some of these metals on the calibration model of the instrument is also investigated.

The data processing and interpretation in this study were done by evaluating and assessing the quality and performance of principal component (PCR), partial least squares (PLSR) and support vector regression (SVR) models to simultaneously quantify copper (Cu), cobalt (Co) and zinc (Zn) in a sample solution. All models were based on the vis-NIR spectra of standard solutions, while the effects of process conditions, such as temperature and contaminants were evaluated and determined under laboratory conditions.

2. Experimental Work

2.1 Experimental Design

Statistically designed experiments provide a mathematical framework for changing all significant variables or factors simultaneously with a minimum number of samples (Anderson-Cook and Montgomery, 2008). Design Expert v8 software was used to generate a simplex centroid mixture design, as shown in Fig. 1, in which three factors, i.e. the concentrations of copper, cobalt and zinc were varied. The design points (blue circles) correspond to the composition of the calibration samples used in the derivation of the models. In this design, the measured response was assumed to depend only on the relative proportions of the components in the mixture, not the amount or quantity of the mixture (Cornell, 1981).

Permutations of the pure blends are located at the vertices, the binary blends are located along the edges and permutations of the three-component blends are located inside the triangle. The centroid point was repeated twice to yield 12 runs for the design. The values given in the figure are the percentage of the metals from their respective stock solution concentration, as described in Section 2.2. The experimental runs for each set of samples were performed randomly to accommodate any instrumental error that might have occurred during measurement.

2.2 Sample Preparation

The sulphate salts of copper, cobalt and zinc metals were used to prepare the stock solutions with concentration values of 500 mg/L, 50 mg/L and 50 000 mg/L, respectively. Each salt was first dissolved in 100 mL of distilled water and then 50 mL of 98% sulphuric acid was added to give a 500 mL solution after further dilution. High purity sulphate salts were used to avoid impurities in the metal solutions interfering with the absorption spectra of the metals. The calibration samples were prepared by diluting the stock solutions with 10% sulphuric acid to obtain the required composition of the mixture, as indicated by the design points in Fig. 1.

In order to test the models and evaluate the effect of interference on the predictive ability of the models, the proportions of the metals were chosen arbitrary so that the percentage of total metals

added up to 100. In addition, the temperature of the sampling solution was increased from 20 to 80 °C in 20 °C increments to assess the effect of temperature on the performance of the calibration models. To simulate the effect of nickel contamination, nickel was added in increments of 50 mg/L to the sampling solution in concentrations ranging from 50 mg/L to 200 mg/L. The experimental runs are summarized in Tables 1-4.

2.3 On-line Measurement Set-up

The Blue Cube instrument used for analysis consists of four basic components, i.e. an interface, data processor, sample holder and spectrophotometer. The interface is connected to a main switch and it supplies a steady 12 volts to both the data processor and spectrophotometer. The data processor contains the computer that processes the data captured from the sample. The spectrophotometer has an optical sensor with a scratch resistant sapphire window that uses the principle of diffused reflective spectroscopy over a wavelength range from approximately 350-1050 nm (Lottering and Aldrich, 2006).

A light source forms an integral part of the spectrophotometer and light is transmitted from the source to the sample holder via 6 light + 1 dark (reference) 400 μm diameter optical fibres, which are also used to transmit the resulting signal from the solution to the data processor for interpretation. The 100 W instrument is controlled at a temperature of between 50 and 55 °C. Fig. 2 shows a photograph of an industrial prototype of the Blue Cube optical fibre sensor instrument with the features mentioned above. The solution was poured thorough a funnel into the sample holder, with a volumetric capacity of 180 mL. The plastic clamp, shown in Fig. 2, was used to reduce the flow rate of the samples by constraining the inlet tube leading to the sample holder.

The solution inside the measuring cell was allowed to stabilise for 30 seconds before analysis was done. The start button on a remote control was pressed to initiate data capture and three scans or repetitions were done for each sample over a period of six minutes. Light was passed through the sample and the reflected light signals were fed to a data processor via optical fibres.

The desktop computer that was connected to the on-line instrument was used to capture the vis-NIR spectra for further processing. In between collection of samples, the system was washed twice with deionised water to remove possible contamination from the previous sample and finally rinsed with the sample to be analysed to improve the accuracy of the measurement. Finally, the collected spectral data were preprocessed and analysed by the calibration models.

2.4. Multivariate calibration models

Multivariate calibration models have been applied successfully in a range of industries, such as pharmaceuticals, food processing and the mineral processing industries (Dinç and Üstündag, 2003; Haaland et al., 2000; Haavisto et al., 2008). There are several different approaches to model construction, such as classical least squares calibration and prediction algorithms that are based on the explicit linear additive relation of the Beer-Lambert law, inverse least squares models based on the inverse of the Beer-Lambert law, principal component regression, as well as partial least squares regression. Of these models, the latter two are used widely, (e.g. Moneeb, 2006; Da Costa et al., 2009; McElroy et al., 2012; Otto and George, 1987; Peralta-Zamora et al., 1997; García et al., 1997; Real et al., 2007; Haavisto, and Hyötyniemi, 2009) as it can deal with large numbers of correlated variables often arising in spectrophotometric models. It is also the approach that will be considered here.

2.4.1. *Principal component regression models*

Principal component regression (PCR) is used to determine the relationship between a set of predictor variables, $\mathbf{X} \in \mathbb{R}^{nxm}$ and a response or predicted variable, $\mathbf{Y} \in \mathbb{R}^{nxl}$ consisting of n observations or samples. Instead of regression involving the predictor variables themselves, the principal components of the predictor variables are used, i.e. $\mathbf{X}$ is mean centred, scaled and approximated by the first k principal components of the predictor variables according to eq. 1, obtained by singular value decomposition which gives a score ($\mathbf{T}_k$) and loading ($\mathbf{P}_k$) matrix. The response variable can subsequently be estimated via regression on the k retained principal components.

$$\mathbf{X} = \mathbf{T}_k \mathbf{P}_k{}^T + \mathbf{E} \tag{1}$$

$$\hat{\boldsymbol{y}} = \mathbf{T}_k \mathbf{b} \tag{2}$$

Selection of the number of components to retain in the model can be done by a number of approaches, most often ensuring that the variance of the predictor variables explained by the components is high.

When dealing with data ($\mathbf{X}_{new}$) not used in the construction of the model, the data are scaled with the calibration data parameters, their principal component scores $\mathbf{T}_{new,k}$ are calculated by projection onto the principal component loading matrix ($\mathbf{P}_k$), followed by regression according to eqs 3-4 to yield estimates of the response variable ($\hat{\boldsymbol{y}}_{new}$).

$$\mathbf{T}_{\text{new,k}} = \mathbf{X}_{\text{new}}\mathbf{P}_{\text{k}}{}^{T} + \mathbf{E} \tag{3}$$

$$\widehat{\boldsymbol{y}}_{\boldsymbol{new}} = \mathbf{T}_{\text{new,k}}\mathbf{b} \tag{4}$$

2.4.2. *Partial least squares calibration models*

Like PCR, partial least squares regression (PLSR) is a multivariate calibration technique used to determine the relationship between a set of predictor variables, $\mathbf{X} \in \mathbb{R}^{nxm}$ and set of responses or predicted variables, $\mathbf{Y} \in \mathbb{R}^{nxl}$ consisting of *n* observations or samples. It is particularly useful when there are many correlated predictor variables and relatively few samples, as is often the case with the calibration of spectrophotometric models. Unlike multiple linear regression, PLS regression yields stable predictions, even when the variables in $\mathbf{X}$ are highly correlated. The algorithm is briefly summarized here, as more detailed descriptions can be found elsewhere (Garthwaite, 1994; Wold et al., 2001). Essentially, it is based on a bilinear decomposition of $\mathbf{X}$ and $\mathbf{Y}$, as follows

$$\mathbf{X} = \mathbf{TW}^{T} + \mathbf{E} \tag{5}$$

and

$$\mathbf{Y} = \mathbf{UQ}^{T} + \mathbf{F} \tag{6}$$

where $\mathbf{T}$, $\mathbf{W}$, $\mathbf{U}$ and $\mathbf{Q}$ are the score and loading matrices of $\mathbf{X}$ and $\mathbf{Y}$ respectively, and $\mathbf{E}$ and $\mathbf{F}$ are the residuals of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Decomposition takes place by projecting $\mathbf{X}$ and $\mathbf{Y}$ in directions that maximize the covariance between $\mathbf{T}$ and $\mathbf{U}$. Eqs 5-7 constitute the PLS model

$$\boldsymbol{y}_{new} = \mathbf{X}_{new}\mathbf{b}_{PLS} + \mathbf{e} \tag{7}$$

Determination of the number of components to retain in the model is usually done by means of cross-validation.

2.4.3. *Support Vector Regression*

Given training data of the form $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_1, y_1), \dots (\boldsymbol{x}_n, y_n)\}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ are the input or predictor variables and $y_i \in \mathbb{R}$, the output or predicted variable, support vector regression amounts to identification of the relationship

$$f(\boldsymbol{x}) = <w, x> + b \tag{8}$$

This is an optimization problem, with the tacit assumption in (4) that the function *f* that approximates all pairs $(\boldsymbol{x}_i, y_i)$ with ε-precision for *i* = 1, 2, … *n* actually exists, i.e. that the convex optimization problem is *feasible*. Since this may not be the case in general, some allowance for

6

errors needs to be made. This is done by using slack variables ξ_i, ξ_i^* to deal with otherwise infeasible constraints to the optimization problem, leading to a formal statement of the optimization problem according to eqs (5)-(8) (Vapnik, 1995). In these equations, n denotes the number of samples, $\boldsymbol{x}_i$ is the i'th sample mapped to a high dimensional space Φ by the kernel functions ϕ, ξ_i represents the upper training error, and ξ_i^* the lower training error, subject to the ε-insensitive loss function.

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{9}$$

subject to

$$y_i - (\boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}_i) + b) \leq \varepsilon + \xi_i \tag{10}$$

$$(\boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}_i) + b) - y_i \leq \varepsilon + \xi_i \tag{11}$$

$$\xi_i, \xi_i^* \geq 0 \tag{12}$$

In essence support vector regression maps the data set $\boldsymbol{x}_i$ from the input space χ into a high-dimensional feature space Φ by means of kernel functions. Three parameters determine the quality of the support vector regression model, namely the error cost, C, the sensitivity of the loss function, ε and the kernel function ϕ. In this study, linear (eq. 13) and Gaussian kernels (eq. 14) were used, i.e.

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j \tag{13}$$

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{\|\boldsymbol{x}_i, \boldsymbol{x}_j\|^2}{2\sigma^2}\right) \tag{14}$$

The solution of this problem is described in detail elsewhere (Drucker et al., 1997; Smola and Schölkopf, 1998) and won't be derived here.

3. Results and discussion

3.1 Data processing

The raw data from the Blue Cube optical sensor were expressed in units of transmittance counts. Fig. 3 shows the four vis-NIR spectra of pure copper (500 mg/L), cobalt (50 mg/L) and zinc (50 000 mg/L) together with a mixture containing 33.33% of each metal from its respective stock solution, over a wavelength range of 400 to 1000 nm. Each spectrum is the average of 54 individual measurements. The optical sensor is calibrated to differentiate between the different metals based on these spectra.

Before calibration models could be constructed, the spectral data were preprocessed and converted to absorbance values. The absorbance (A) of a metal of interest was approximated using the principles of the Beer-Lambert law, as given by eq. (15). The approximation holds for all the samples used in the study, owing to their dilute concentrations (Skoog et al., 2008).

$$A = -log \frac{I_{solution}}{I_{solvent}} \qquad (15)$$

From eq. (15), $I_{solvent}$ and $I_{solution}$ are the intensity values of the light that passes through the solvent and metal solutions respectively.

3.2. Principal component, partial least squares and support vector regression models

Three principal component regression models were fitted to the calibration data in Table 1, i.e. with the 307 wavelength intensities as input variables and the three metal concentrations as output variables. The test data in Table 2 were not used in the calibration of the model, but were used to validate the model. The effect of the number of principal components in the prediction of the calibration and validation data is shown in Fig. 4. Based on these results, six components were extracted from the input variables.

Likewise, PLS regression models were fitted to the calibration data in Table 1, while using the test data in Table 2 to validate the models. The effect of the number of PLS components prediction of the calibration and validation data is shown in Fig. 5. Based on these results, six components were extracted from the input variables as well, which also allowed direct comparison with the principal component models.

Support vector regression models were trained on the same data as the linear models, i.e. 14 calibration samples (Table 1) and validated on 9 test samples (Table 2). All the wavelength absorptivities were used as predictor variables, i.e. 307 variables in total, while the concentrations of the metals were used as response variables. The parameters C, ε and the width of the Gaussian basis function (σ) were determined by trial and error and are summarized in Table 5. N_{SV} indicates the number of support vectors extracted from the calibration data during construction of the models.

The results for the three models on the calibration in Table 1 and validation data in Table 2 are summarized in Table 6. These models are compared more formally in the following section.

3.3. Comparison of the models

The performance of the PCR, PLSR and SVR models are compared in Table 6 in terms of the squared correlation coefficients and the root means square errors (RMSE) of the experimental and predicted response variables. The hypotheses that the models are different were assessed by making use of Fisher's Z transform. This transform is given by Z which is approximately normally distributed as given by:

$$Z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right) \tag{16}$$

When comparing any two models, the null hypothesis was tested that the difference between the correlation coefficients of the two models is zero, i.e. H_0: $Z_1 - Z_2 = 0$. The alternative hypothesis was simply that the models are different, i.e. the difference between the correlation coefficients is not zero, or H_1: $Z_1 - Z_2 \neq 0$. In this case the standard error of the test statistic $Z_1 - Z_2$ is

$$s = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \tag{17}$$

In eq. 17, n_1 and n_2 denote the number of samples on which models 1 and 2 were based (the overall fits on data from Tables 1 and 2 were used, i.e. $n_1 = n_2 = 23$ in all cases). This allows the calculation of confidence intervals around $Z_1 - Z_2$ of the form $(Z_1 - Z_2) \pm \alpha s$, with the value of Z at a given confidence level, e.g. $\alpha = 1.96$ for 95% certainty.

Although all the models yielded statistically similar results, the computational demands of the SVR models based on grid searching for optimal parameters, C, ε and σ, were considerably higher. As a consequence, use of the PCR and PLSR models over SVR models would be preferable from a practical perspective and further analyses were conducted with the PLSR models only.

3.4. Influence of process disturbances on model calibration.

3.4.1 Temperature

Fig. 6 shows the absorption spectra of the mixture of 50 mg/L Cu, 15 mg/L Co and 30 000 mg/L Zn at different temperatures. In this study, the intensities of the absorption spectra increased with an increase in temperature, as seen in the wavelength range of 700-1000 nm.

From Fig. 6, as the temperature increases from 20 to 80 °C, the spectra associated with Cu result in three distinct absorption peaks. These peaks have their origin in the behaviour of the water

molecules in the solutions. At higher temperatures, water changes its state from liquid to vapour and water vapour shows several maximum absorbance peaks in the near infrared (NIR) region. Water vapour exhibit more intense peaks than the metals of interest in this study, since its absorbance values are affected by both electronic (absorption) and vibrational transitions, whereas the Cu, Co and Zn metals exhibit only electronic (absorption) transitions.

As shown in Fig. 7, temperature had a significant effect on all calibrated models, even at 40 °C. However, the Co model was least affected, with residual values approximately an order of magnitude smaller than those of the Zn model.

3.4.2 Nickel contamination

Fig. 8 illustrates the absorption spectra of a solution of Cu, Co and Zn metals in the presence of nickel. The solution was composed of 150 mg/L Cu, 15 mg/L Co and 20 000 mg/L Zn metals on a mass basis. An increase in the Ni concentration leads to a shift in the peak of the spectrum to a shorter wavelength. At a Ni concentration of 200 mg/L, the peak has shifted from approximately 800 nm to 750 nm.

Moreover, Ni metal is characterised by two absorption peaks, i.e. one sharp and one broad peak, located around 410 nm and 720 nm, respectively. The positions of the two peaks of Ni are close to the maximum absorption peaks of Cu and Zn metals, and therefore a poor quantitative analysis of these metals can be expected in the presence of Ni. The presence of Ni as a contaminant in the solutions degrades the models, as it is not explicitly accounted for in the calibration of the models, as indicated in Fig. 9. Based on the magnitudes of the model residuals, the Cu model appears to be less sensitive to nickel contamination than the Co and Zn models.

3.4.3 Visualization of the spectral data

The spectra from the calibration and test data sets combined (black), the spectra obtained at elevated temperatures (red) and the spectra obtained in the presence of nickel in the solutions are shown in Fig. 10. Differences in the spectra are not readily apparent from this figure.

The effect of the presence of Ni and elevated temperatures on the spectral absorbance of the metal solutions can be visualized better by projecting the data to a lower-dimensional map by any of a number of multivariate methods. In this case, Fisher linear discriminant analysis was used to maximally separate the clusters in the spectral data associated with the different experimental runs. That is, given a set of grouped observations

$$\begin{bmatrix} (x_1, y_1) \\ (x_2, y_2) \\ \vdots \\ (x_n, y_n) \end{bmatrix}, where \ (x_i, y_i) \in \mathbb{R}^p \otimes Q \tag{18}$$

where x_i is a vector of l observed measurements and Q is the set of q categories or groups to which x_i is assigned, then discriminant analysis seeks a transformation w that maps x_i to a lower-dimensional latent variable z_i, so that the groups of data are maximally separated. That is

$$z_i = w^T x_i + b \tag{19}$$

where b is a bias term. This is mathematically equivalent to finding the vector w that maximizes the coefficient $\eta(w)$ as a function of the between-group and within-group scatter matrices, C_B and C_W.

$$\eta(w) = \frac{w^T C_B w}{w^T C_W w} \tag{20}$$

This is accomplished via the solution of an eigenvalue decomposition problem, the details of which can be found in standard references on multivariate statistics, e.g. Fukunaga (1990).

Not all the spectra were used to construct the model, since the spectra were highly correlated and also to avoid practical problems with the model handling a relatively large input space. Therefore the absorbances at the $l = 10$ wavelengths 929-947 nm were used as the set of input variables and the $q = 51$ experimental run indices as shown in Tables 1-4 were used as the categorical response variables. The projection of a total of $n = 2754$ samples (54 per experimental run) is shown in Fig. 11 on which the data are indicated by different colours and numbers (the experimental runs or groups).

Runs 4, 10 and 13, at the centre runs of the mixture design, give an indication of the repeatability of the measurements. The scores of runs 10 and 13 overlap almost completely, while run 4 can be seen as a cluster in close proximity to the other two clusters. Likewise, the clusters associated with runs 1, 14 and 15 are also closely proximate, as these runs were also identical (pure water).

The nickel contaminants had a moderate effect on the model. In Fig. 8, for example, the effect that the presence of nickel has on the spectra of a solution with the Cu-Co-Zn composition represented by run 18 (located at approximately $z_1 = 18$ and $z_2 = 0$) can be seen. The addition of 50, 100, 150 and 200 mg/L Ni to this solution is shown by runs 24-27 respectively. Similarly, the effect of nickel of 50, 100, 150 and 200 mg/L Ni on the spectra of a solution with the Cu-Co-Zn composition represented by run 19 (located at approximately $z_1 = 7$ and $z_2 = -2$) are indicated by runs 28-31 respectively. The effect is again moderate. The same goes for the Cu-Co-Zn compositions

represented by runs 20 and 23 around which the sets of runs 32-35 and 36-39 are clustered respectively.

The influence of the nickel in solution on the reliability of the optical sensor can be quantified by assessing the reliability of the SVR models (calibrated on the data in Tables 1 and 2) when nickel is present in solution (Table 3). The results are summarized in Table 8, showing the R^2-values of the models in the presence and absence of nickel in solutions. Here the concentrations of the nickel were not taken into account, as too few data were available to analyze the effect of different nickel concentrations. As can be seen from these data and the p-values in the table, the presence of Ni up to 200 mg/L had a significant effect on estimation of all the metal concentrations.

The effect of temperature can be assessed by visual inspection of run 16 in Fig. 8 in conjunction with runs 40, 44 and 48 associated with temperatures of 40, 50 and 80 °C. The clusters associated with elevated temperatures move progressively further away from the cluster associated with run 16 at a temperature of 20 °C. The same effects can be seen by considering runs 17, 41, 45 and 49, all of which are associated with identical solution compositions, but at temperatures increasing from 20 to 80 °C. The same also applies to runs 21, 42, 46 and 50, as well as runs 22, 43, 47 and 51.

On the whole, the effect of temperature on the spectra appears to be significantly more pronounced than the effect of the presence of nickel in the solutions, and quantification is therefore not shown here, as all *p*-values could be expected to be very close to 100%.

When used online in real process environments, the validity of the sensor's calibration model would have to be monitored, for example by mapping new scores of the spectral data onto a map similar to the one shown in Fig. 11. This does not have to be done by plant staff, who may not have the necessary statistical expertise, but can be (and is often) done by the vendor or data analyst from a site globally remote from the plant. If these data were to deviate from the calibration data, recalibration would be required, for example based on the use of grab samples on the plant.

4. Conclusions

In this study, an industrial spectrophotometer was used for the simultaneous monitoring of copper, zinc and cobalt in aqueous solutions that simulated plant conditions encountered on the Skorpion zinc mine in Namibia. Principal component, partial least squares and support vector regression models were calibrated from visible and near infrared (vis-NIR) absorption spectra, and the following can be concluded from the study:

- The principal component regression, partial least squares and support vector regression models could all estimate the metal concentrations in solution accurately.

- Although all models were affected by nickel contamination and increased temperatures, the Co and Cu models were considerably less sensitive to increased temperature levels, while the Cu model was also less sensitive to nickel contamination than the Co and Zn models.
- The effect of these elevated temperatures and nickel contamination on the calibrated models could be visualized effectively by projection of the high-dimensional spectra onto a linear discriminant score plot. These effects would generally require recalibration of the models, which could be guided by such a discriminant score plot.

Acknowledgements

References

Anderson-Cook, C.M. and Montgomery, D.C. 2008. Response surface design evaluation and comparison. *Journal of Statistical Planning and Inference*, 139, 1-8.

Bergh, L.G., Jämsä-Jounela, S.-L. and Hodouin, D. 2001. State of the art in copper hydrometallurgical processes control. *Control Engineering Practice*, 9, 1007-1012.

Cornell, J.A. 1981. *Experiments with Mixtures*. John Wiley & Sons, New York.

Da Costa, G.M., Barrón, V., Ferreira, C.M. and Torrent. J. 2009. The use of diffuse reflectance spectroscopy for the characterization of iron ores. *Minerals Engineering*, 22{14}, 1245-1250.

Dinç, E. and Üstündag, Ö. 2003. Spectophotometric quantitative resolution of hydrochlorothiazide and spironolactone in tablets by chemometric analysis methods. *Il Farmaco*, 58(11), 1151-1161.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A. and Vapnik, V. 1997. Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155-161.

Fukunaga, K. 1990. *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, CA, USA.

Gaft, M., Sapir-Sofer, I., Modiano, H. and Stana, R. 2007. Laser induced breakdown spectroscopy for bulk minerals online analyses. *Spectrochimia Acta: Part B*, 62, 1496-1498.

García, M.A., Martí, V., Meinhardt, E., Cortina, J.L. and Granados, M. 1997. Metal cyanide control in hydrometallurgical processing of gold ores by multivariate calibration procedures. *Analytica Chimica Acta*, 353(1), 123-131.

Garthwaite, P.H. 1994. An interpretation of partial least squares. *Journal of the American Statistical Association*, 89(425), 122-127.

Haaland, D.M., Chambers, W.B., Keenan, M.R. and Melgaard, D.K. 2000. Multi-window classical least-squares multivariate calibration methods for quantitative ICP-AES Analyses. *Applied Spectroscopy*, 54(9), 1291-1302.

Haavisto, O. and Hyötyniemi, H. 2009. Recursive multimodel partial least squares estimation of mineral flotation slurry contents using optical reflectance spectra. *Analytica Chimica Acta*, 642(1–2), 102-109.

Haavisto, O. and Kaartinen, J. 2009. Multichannel reflectance spectral assaying of zinc and copper flotation slurries. *International Journal of Mineral Processing*, 93(2), 187-193.

Haavisto, O., Kaartinen, J. and Hyötyniemi, H. 2008. Optical spectrum based measurement of flotation slurry contents. *International Journal of Mineral Processing*, 88(3-4), 80-88.

Lewis, E., Sheridan, C., O'Farrell, M., King, D. and Flanagan, C. 2007. Principal component analysis and artificial neural network based approach to analysing optical fibre sensors signals. *Sensors and Actuators*, 136, 28-38.

Lottering, J.M. and Aldrich , C. 2006. Online measurement of factors influencing the electrostatic separation of mineral sands, *Journal of the Southern African Institute of Mining and Metallurgy*, 106(4), 283-290.

McElroy, L., Bao, J., Jayasundara, C.T., Yang, R.Y., and Yu, A.B. 2012. A soft-sensor approach to impact intensity prediction in stirred mills guided by DEM models. *Powder Technology*, 219, 151-157.

Moneeb, M.S. 2006. Polarographic chemometric determination of zinc and nickel in aqueous samples. *Talanta*, 70(5), 1035-1043.

Otto, M and George, T. 1987. Application of multicomponent spectrophotometry in analysis of copper electroplating bath solutions. *Analytica Chimica Acta*, 200, 379-385.

Peralta-Zamora, P., Cornejo-Ponce, L., Nagata, N. and Poppi, R.J. 1997. Chemometric alternatives for resolution of classical analytical problems: Spectrophotometric determination of lanthanide mixtures. *Talanta*, 44(10), 1815-1822.

Real, B.D., Ortiz, M.C. and Sarabia, L.A. 2007. Analysis of interferents by means a D-optimal screening design and calibration using partial least squares regression in the spectrophotometric determination of Cr(VI). *Talanta*, 71(4), Pages 1599-1609.

Remes, A., Saloheimo, K. and Jämsa-Jounela, 2007. Effect of speed and accuracy of on-line elemental analysis on flotation control performance. *Minerals Engineering*, 20, 1055-1066.

Smola, A.J. and Schölkopf, B. 1998. A tutorial on support vector regression. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK.

Skoog, D.A., Holler, F.J. and Crouch, S.R. 2008. *Fundamentals of Analytical Chemistry*. Thomson Brooks/Cole, Canada, Chapter 24, 26 pp.

Sowerby, B. 2002. Online measurement and control in sustainable mineral processing and energy production. *ATSE Focus*, 120, 1.

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

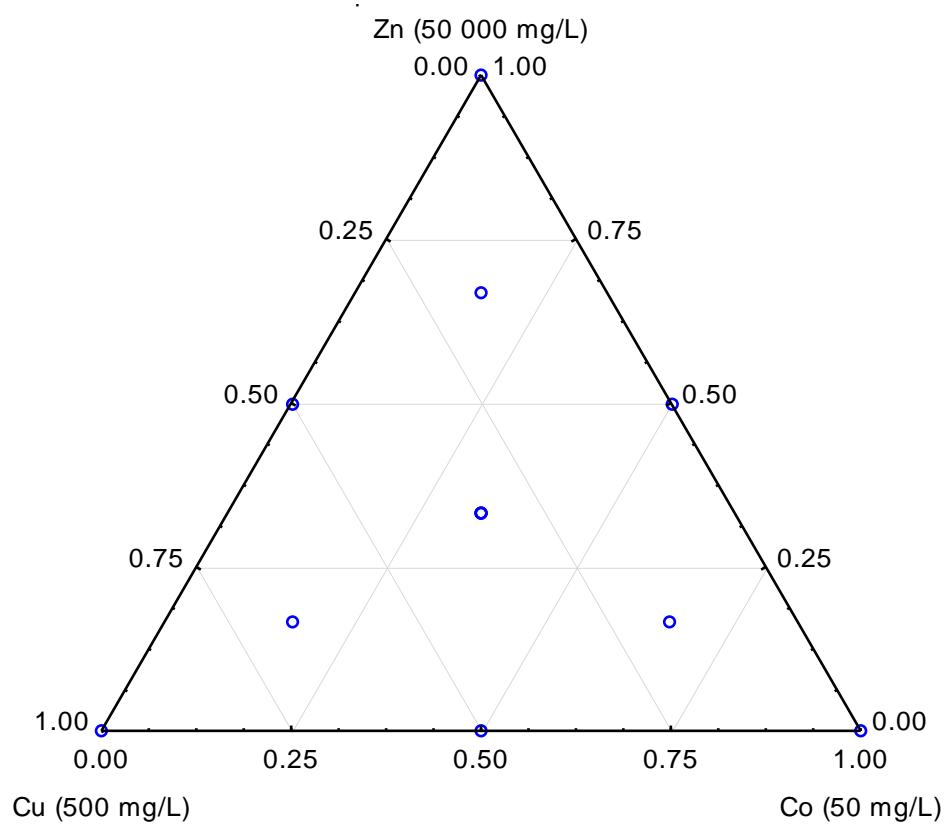Wold, S., Sjöström, M. and Eriksson, L. 2001. PLS: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130.

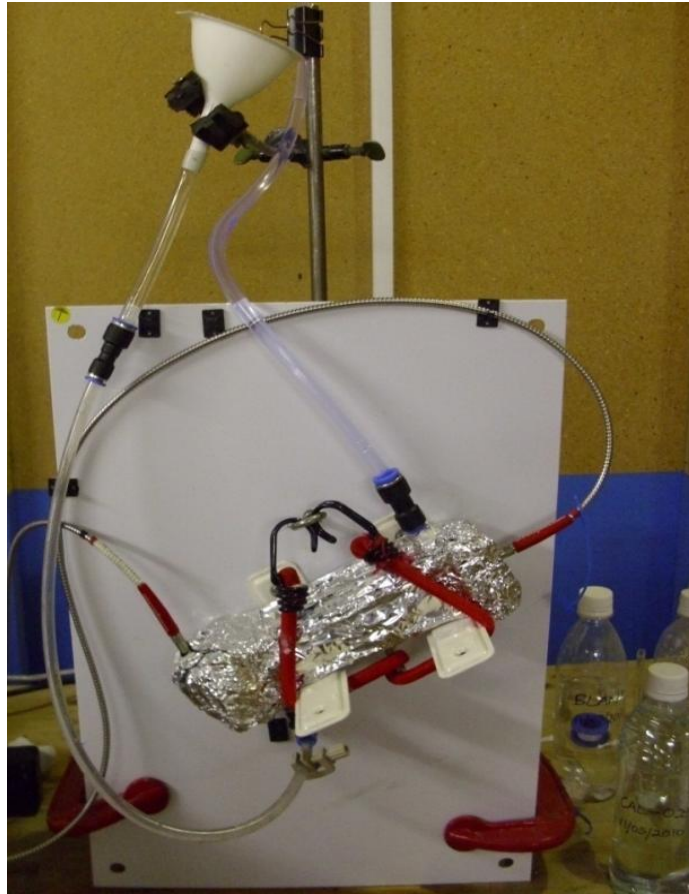Figure 1: Three factor simplex-centroid mixture design with data points.

Figure 2: Laboratory set-up of Blue Cube instrument with sample holder.
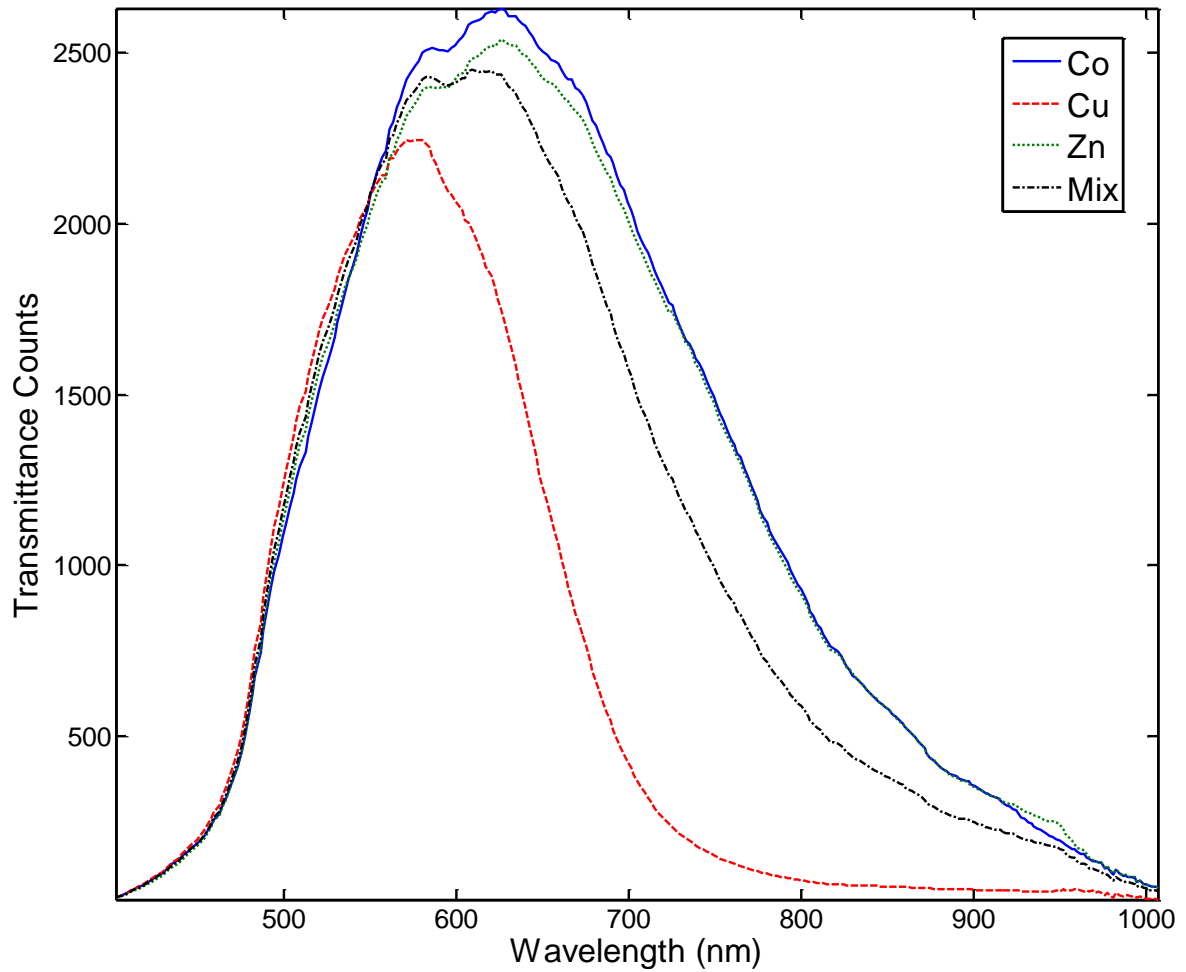
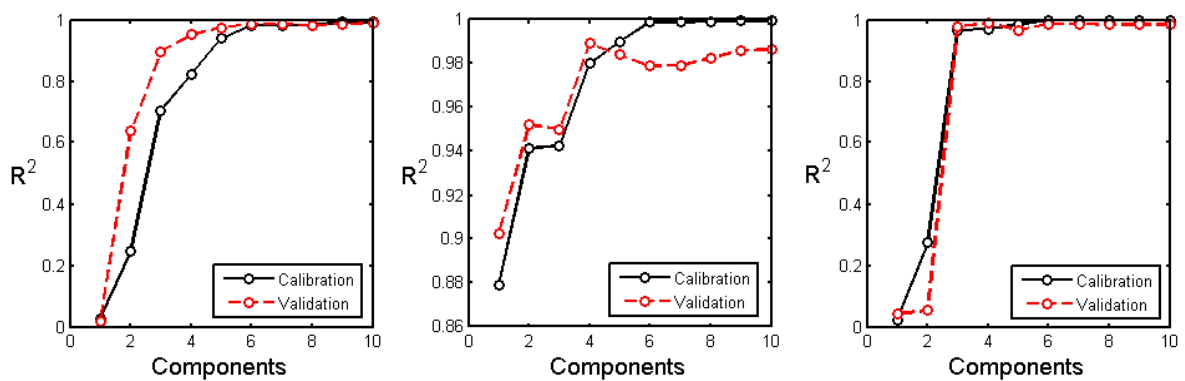Figure 3: Raw spectra of pure Cu, Co, Zn and their 1:1:1 mixture.

Figure 4: Root means square errors (RMSE) of Zn (left), Cu (middle) and Co (right) PCR models for the calibration and validation data as a function of the number of components in the models.
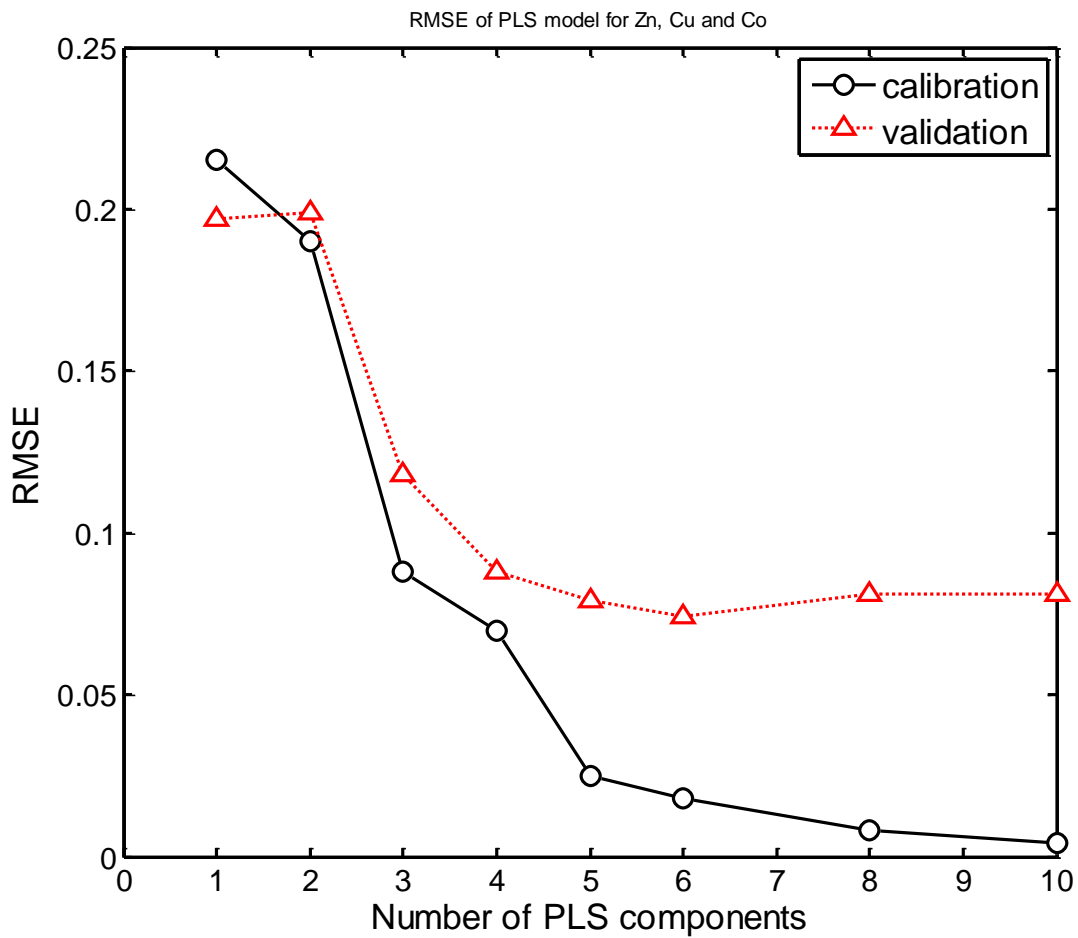
Figure 5: Root means square errors (RMSE) of PLSR models for the calibration and validation data as a function of the number of components in the models.
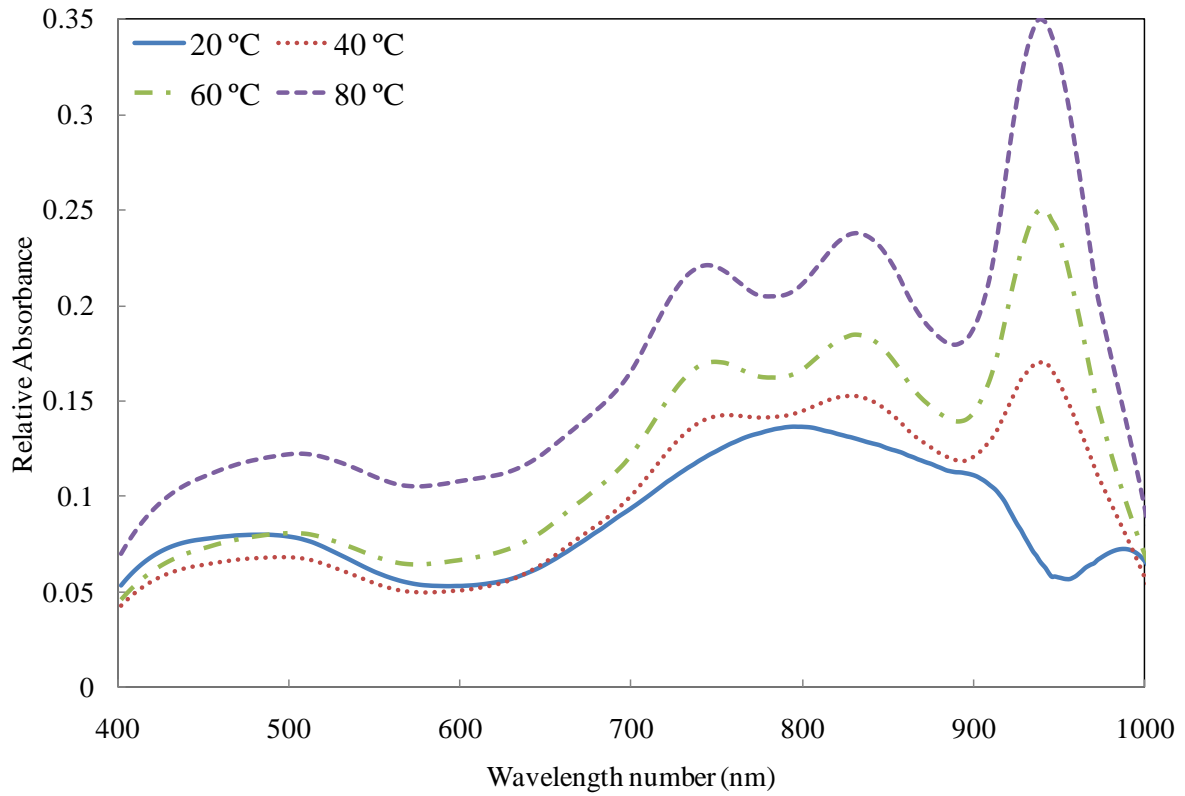
Figure 6: Influence of temperature on absorption spectra of a mixture of
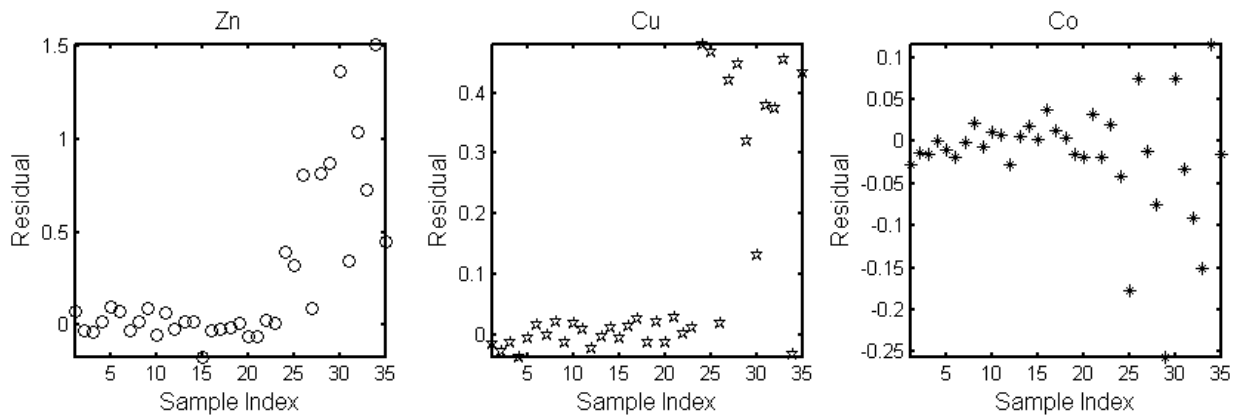50 mg/L Cu, 15 mg/L Co and 30 000 mg/L Zn.

Figure 7: Effect of elevated temperatures on the PLSR models (Zn, Cu, Co from left to right).
Sample index 1:14 (calibration data), 15:23 (validation data), 24:27 (40 C°), 28:31 (60 C°),
and 32:35 (80 C°).
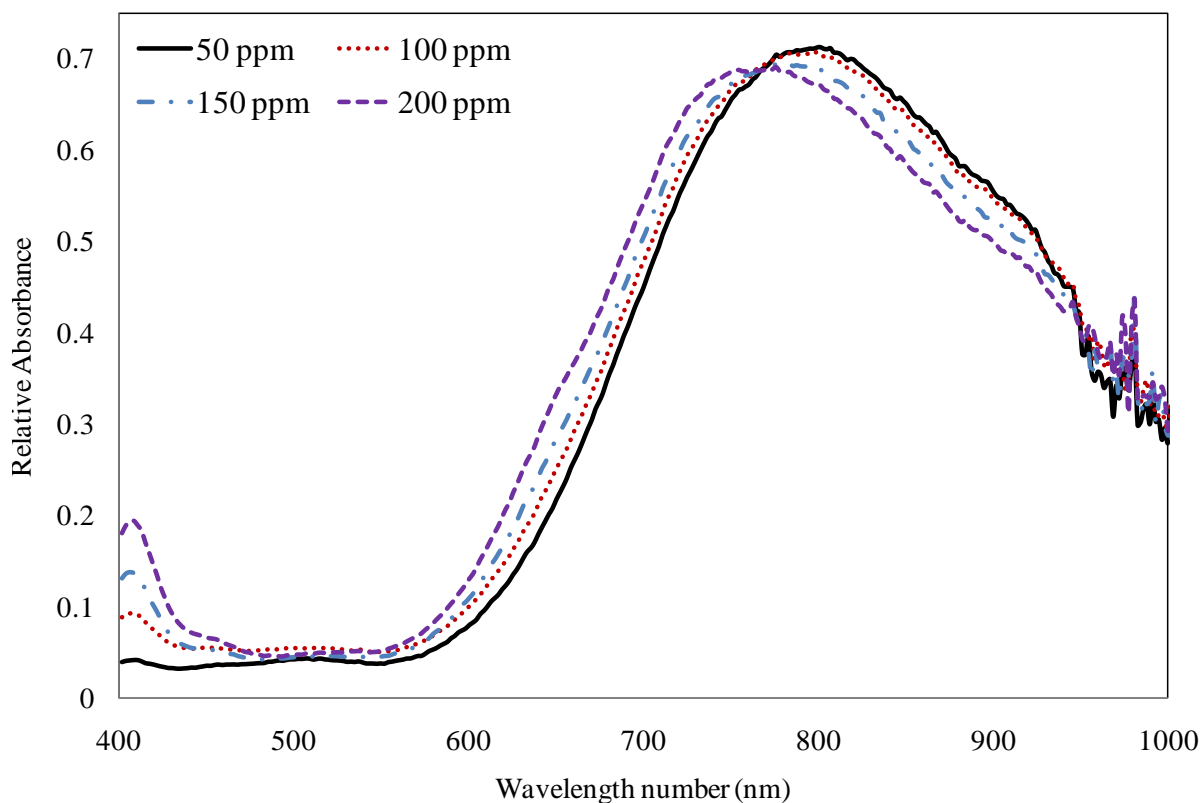
Figure 8: Absorption spectra of a mixture of 150 mg/L Cu, 15 mg/L Co and 20 000 mg/L in the presence of nickel in concentrations as indicated in the legend.
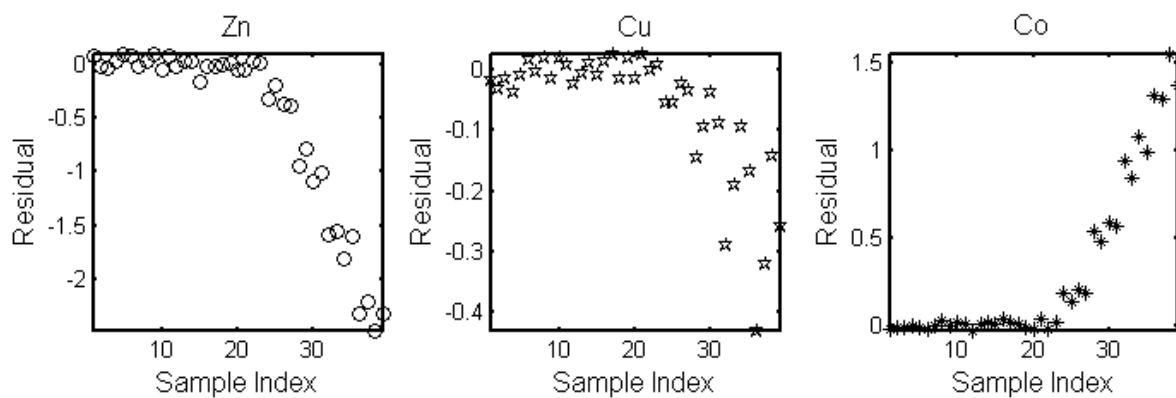

Figure 9: Effect of Ni contamination on the PLSR models (Zn, Cu, Co from left to right). Sample index 1:14 (calibration data), 15:23 (validation data), 24:27 (50 ppm Ni), 28:31 (100 ppm Ni), 32:35 (150 ppm Ni) and 36:39 (200 ppm Ni).

Figure 10: Spectra of the calibration data (top), elevated temperature conditions (middle) and simulated nickel contamination (bottom).

Figure 11: Linear discriminant analysis score plot of the spectral data associated with the calibration data (indices 1-14, black), test data (indices 15-23, magenta), nickel contaminants (indices 24-39, blue) and elevated temperature conditions (indices 40-51, red). The data corresponding to the experimental runs are summarized in Tables 1-4.

Table 1. Solution compositions and temperatures for calibration data.

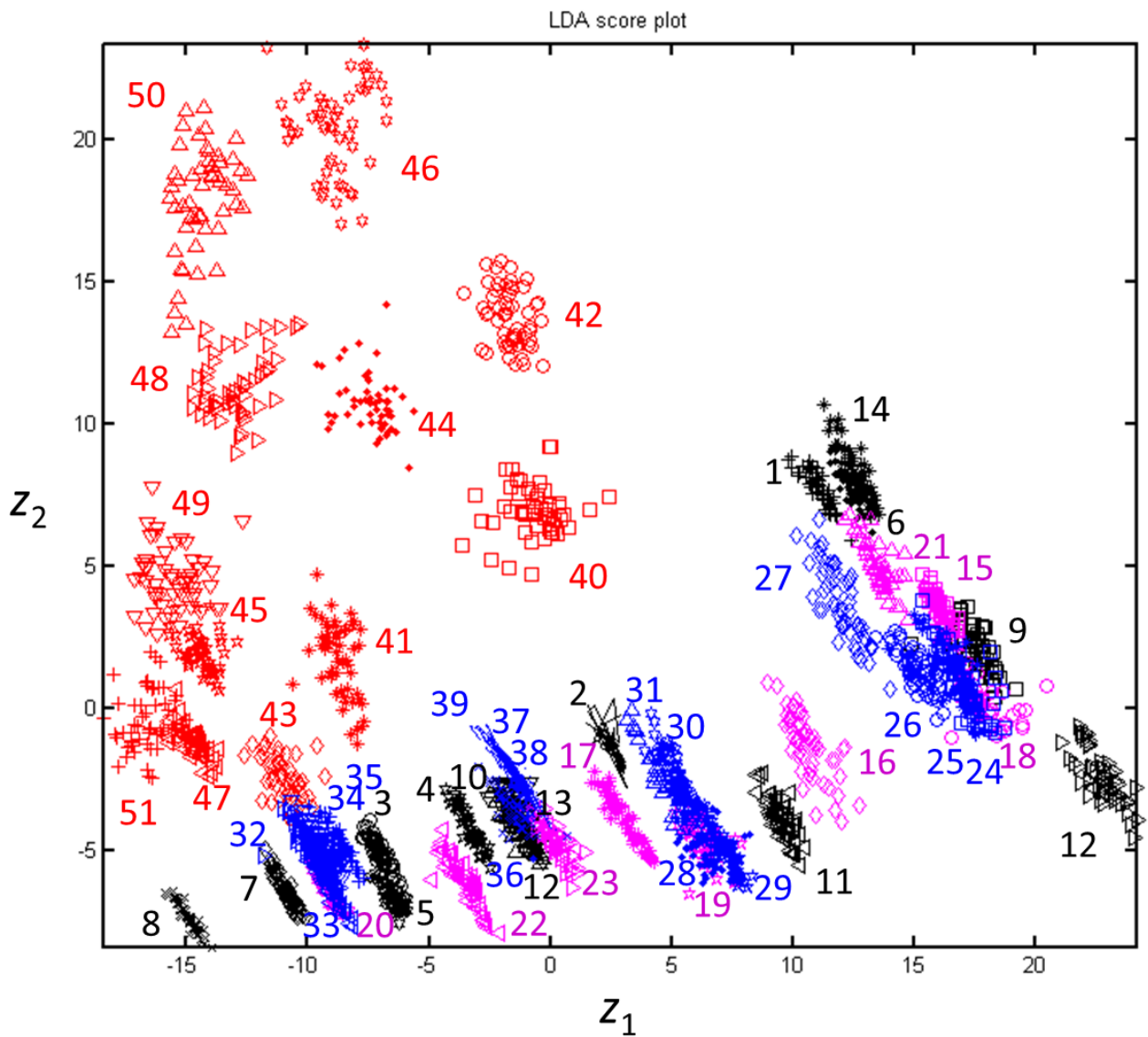| Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Cu (mg/L)	0	83.3	250	167	250	0	333.3	500	0	167	83.3	0	167	0
Co (mg/L)	0	33.3	25	16.7	0	50	8.33	0	25	16.7	8.33	0	16.67	0
Zn (mg/L)	0	8.33	0	16.7	25	0	8.33	0	25	16.7	33.3	50	16.7	0
Ni (mg/L)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T (°C)	20	20	20	20	20	20	20	20	20	20	20	20	20	20

Table 2. Solution compositions and temperatures for test data.

Run	15	16	17	18	19	20	21	22	23
Cu (mg/L)	0	50	125	0	100	300	0	200	150
Co (mg/L)	0	15	13	15	0	20	40	0	15
Zn (g/L)	0	30	25	35	40	0	10	30	20
Ni (mg/L)	0	0	0	0	0	0	0	0	0
T (°C)	20	20	20	20	20	20	20	20	20

Table 3. Solution compositions and temperatures for nickel effect data.

Run	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
Cu (mg/L)	0	0	0	0	100	100	100	100	300	300	300	300	150	150	150	150
Co (mg/L)	15	15	15	15	0	0	0	0	20	20	20	20	50	50	50	50
Zn (g/L)	35	35	35	35	40	40	40	40	0	0	0	0	20	20	20	20
Ni (mg/L)	50	100	150	200	50	100	150	200	50	100	150	200	50	100	150	200
T (°C)	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

Table 4. Solution compositions and temperatures for temperature effect data.

Run	40	41	42	43	44	45	46	47	48	49	50	51
Cu (mg/L)	50	125	0	200	50	125	0	200	50	125	0	200
Co (mg/L)	15	12.5	40	0	15	12.5	40	0	15	12.5	40	0
Zn (g/L)	30	25	10	30	30	25	10	30	30	25	10	30
Ni (mg/L)	0	0	0	0	0	0	0	0	0	0	0	0
T (°C)	40	40	40	40	60	60	60	60	80	80	80	80

Table 5: Parameters of support vector regression models.

	Cu	Co	Zn
Kernel	RBF	RBF	Linear
C	10	17	2
ε	0.1	0.1	0.1
σ	0.003	0.003	-
N_{SV}	5	11	11

Table 6: Summary statistics for the PCR, PLS and SVR models for copper, cobalt and zinc.

	Correlation Coefficients						Root Mean Square Errors					
Metal	R^2_{calib}			R^2_{test}			$RMSE_{calib}$			$RMSE_{test}$		
	PCR	PLS	SVR	PCR	PLS	SVR	PCR	PLS	SVR	PCR	PLS	SVR
Cu	0.999	0.999	0.983	0.979	0.980	0.979	0.00643	0.00843	0.0406	0.0400	0.0496	0.0472
Co	0.999	0.999	0.990	0.987	0.988	0.980	0.00400	0.00741	0.0470	0.0507	0.0390	0.0470
Zn	0.982	0.984	0.970	0.987	0.986	0.990	0.0263	0.0370	0.055	0.051	0.135	0.136

List of Figure Captions

Figure 1: Three factor simplex-centroid mixture design with data points.

Figure 2: Laboratory set-up of Blue Cube instrument with sample holder.

Figure 3: Raw spectra of pure Cu, Co, Zn and their 1:1:1 mixture.

Figure 4: Root means square errors (RMSE) of Zn (left), Cu (middle) and Co (right) PCR models for the calibration and validation data as a function of the number of components in the models.

Figure 5: Root means square errors (RMSE) of PLSR models for the calibration and validation data as a function of the number of components in the models.

Figure 6: Influence of temperature on absorption spectra of a mixture of 50 mg/L Cu, 15 mg/L Co and 30 000 mg/L Zn.

Figure 7: Effect of elevated temperatures on the PLSR models (Zn, Cu, Co from left to right). Sample index 1:14 (calibration data), 15:23 (validation data), 24:27 (40 C°), 28:31 (60 C°), and 32:35 (80 C°).

Figure 8: Absorption spectra of a mixture of 150 mg/L Cu, 15 mg/L Co and 20 000 mg/L in the presence of nickel in concentrations as indicated in the legend.

Figure 9: Effect of Ni contamination on the PLSR models (Zn, Cu, Co from left to right). Sample index 1:14 (calibration data), 15:23 (validation data), 24:27 (50 ppm Ni), 28:31 (100 ppm Ni), 32:35 (150 ppm Ni) and 36:39 (200 ppm Ni).

Figure 10: Spectra of the calibration data (top), elevated temperature conditions (middle) and simulated nickel contamination (bottom).

Figure 11: Linear discriminant analysis score plot of the spectral data associated with the calibration data (indices 1-14, black), test data (indices 15-23, magenta), nickel contaminants (indices 24-39, blue) and elevated temperature conditions (indices 40-51, red). The data corresponding to the experimental runs are summarized in Tables 1-4.

List of Table Captions

Table 1. Solution compositions and temperatures for calibration data.

Table 2. Solution compositions and temperatures for test data.

Table 3. Solution compositions and temperatures for nickel effect data.

Table 4. Solution compositions and temperatures for temperature effect data.

Table 5: Parameters of support vector regression models.

Table 6: : Summary statistics for the PCR, PLS and SVR models for copper, cobalt and zinc.

Graphical Abstract
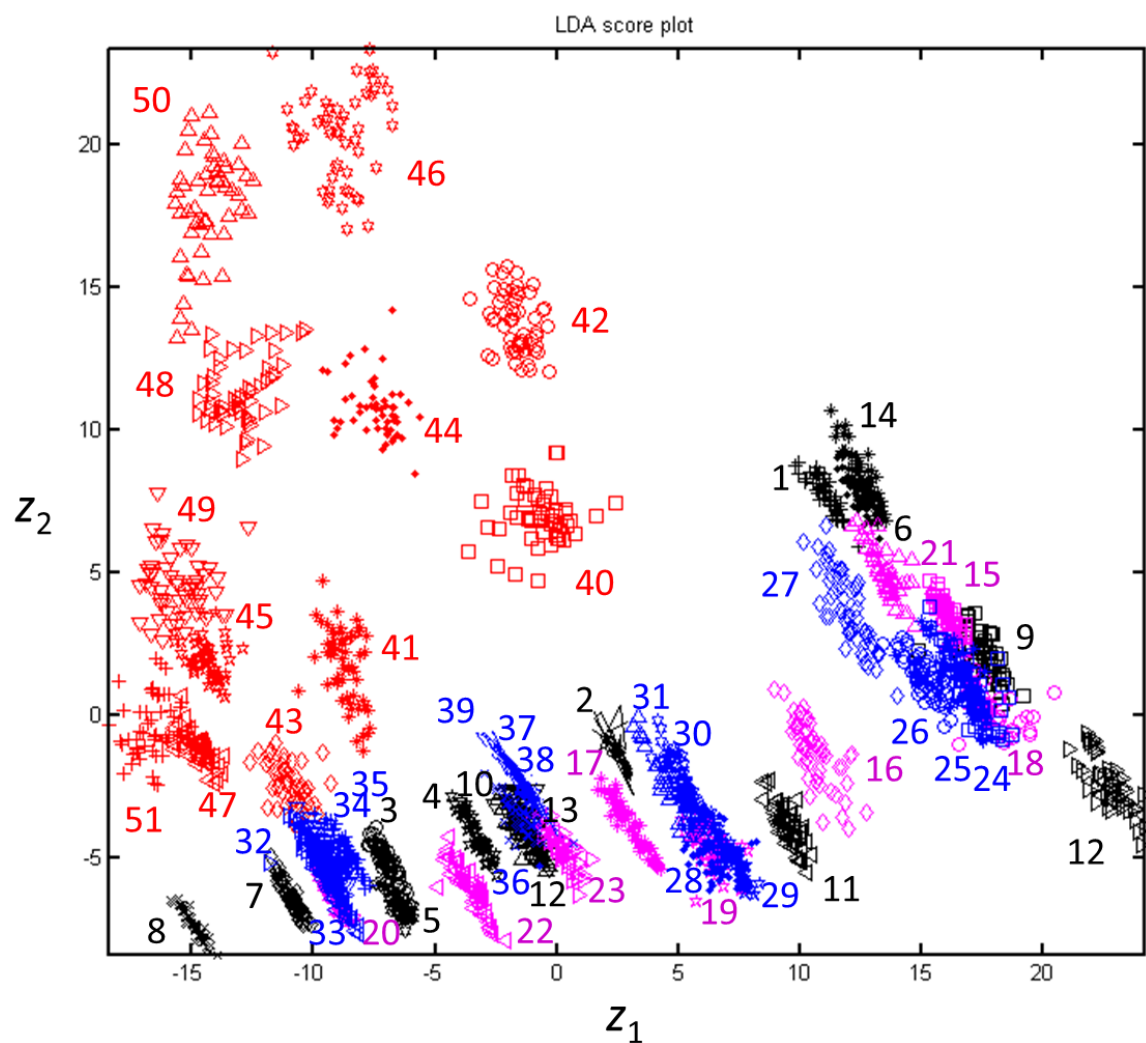
Score plot generating by linear discriminant analysis, showing the projection of 307 spectral variables associated with different conditions in an aqueous solution of Cu, Co, Zn and Ni at different temperatures.

.