

Establishing Multivariate Specification Regions for Raw Materials using SMB-PLS

Kamran Azari*, Julien Lauzon-Gauthier*, Jayson Tessier**, Carl Duchesne*

* *Aluminium Research Centre-REGAL, Department of Chemical Engineering, Laval University
1065 avenue de la Médecine, Quebec, QC, G1V 0A6, Canada (e-mail: carl.duchesne@gch.ulaval.ca)*

***Alcoa Smelting Center of Excellence, Deschambault, QC, Canada, G0A 1S0 (e-mail: jayson.tessier@alcoa.com)*

Abstract: The new Sequential Multi-block PLS algorithm is applied for establishing multivariate specification regions jointly on multiple types of raw materials. SMB-PLS distinguishes between process variations associated with raw materials from other orthogonal sources of variations such as operating policies. Combinations of raw material properties not compensated by the control schemes are clearly identified. Multivariate specifications are required for these combinations to avoid negative impact on process performance and product quality. The method was applied to an industrial aluminum smelter. Bad combinations of raw material properties were identified and validated against process knowledge.

Keywords: Multivariate specifications, Raw materials, Sequential pathway, MB-PLS, Aluminum.

1. INTRODUCTION

The properties of raw materials entering a production process influence its performance and final product quality. Thus, it is essential to develop specification regions for the incoming materials to ensure that the desired product quality or process performance can be achieved. Establishing specification regions has economic benefits for both materials suppliers and producers. Suppliers who meet the specification regions may increase their market share. Incoming materials that meet the specifications ensure the manufacturers of a smooth operation and desired product properties. In addition, manufacturers may be able to choose lower cost materials if those have the right combinations of properties.

Materials quality is a multivariate property where the right balance of properties is required to meet the specifications. Thus, multivariate specification regions are required to be set. Multivariate latent variables methods such as Principal Component Analysis (PCA), Partial Least Squares (PLS) regression (Eriksson et al., 2001) are well established data-driven multivariate methods for process analysis and monitoring in different industries (Kourti, 2005), and could be used to extract those combinations of raw material properties that have a negative impact on product quality (to be avoided).

De Smet (1993) proposed a PLS-based approach for defining multivariate specifications for raw materials properties assuming that product quality solely depend upon the incoming materials. However, as argued by Duchesne and MacGregor (2004), specification regions must account for raw materials and process variations because of the interaction between them. For instance, wider specs regions could be used if an effective process control system attenuating most raw material variations is implemented. In their approach, the relationships between the raw material properties, the process conditions and product quality were modelled using Westerhuis's Multi-block PLS (MB-PLS)

algorithm (Westerhuis et al., 1998), and the specification region was defined through a mapping between the product property space, accounting for the impact of process variations, and that of raw materials. The method was shown to be effective for defining specification regions for a single type of raw materials but did not address the case where multiple types of materials are used and synergy exists between them. In addition, MB-PLS does not clearly identify feedforward and/or feedback control actions used to compensate raw material variations, due to the way it handles the correlations between the blocks.

The Sequential MB-PLS algorithm (SMB-PLS) proposed recently (Lauzon-Gauthier and Duchesne, 2014) is a better alternative for establishing multivariate specification regions for raw materials because it explicitly takes into account the feedforward/feedback control actions taken to compensate for raw material variations. This allows establishing specifications for those combinations of properties that are not compensated for by the currently implemented process control system. The SMB-PLS method imposes a sequential pathway between the regressor blocks according to the process flowsheet (i.e. prior knowledge), and then uses orthogonalization to separate correlated information between the blocks from orthogonal variations (i.e. new information). Hence, variations in process variables that are correlated with raw material properties (e.g. control actions) are extracted in the same latent variables as opposed to MB-PLS which tends to spread correlated information between blocks into several latent variables (Lauzon-Gauthier and Duchesne, 2014). The algorithm also takes into account the synergy between the properties of different types of raw materials when more than one type is required for making the final product.

This work presents a preliminary application of SMB-PLS for establishing multivariate specification regions on multiple raw materials. An industrial case study is presented (aluminium smelting plant) in which two incoming materials

are involved. The multi-block set of data is first modelled using SMB-PLS and the relationships between the raw material blocks and the process block are interpreted. The combinations of incoming materials properties that are not compensated by the actual control system and result in lower performance are then identified. Finally, boundaries are established in the latent variable space of the SMB-PLS model to help avoid those materials at the purchasing stage. The proposed approach is consistent with the Quality by Design (QbD) initiative in the pharmaceutical industry in the sense that it could allow establishing the so-called design space for the properties of multiple types of raw materials and process conditions jointly.

Instead of establishing specification regions, one may be in a situation where it desired to select in an optimal way the raw materials, from those available on the market, their ratios as well as the process conditions under which they should be processed to obtain the desired quality. In that case, the reader is referred to the work of Muteki and MacGregor (2008).

2. METHODOLOGY

SMB-PLS uses the MB-PLS hierarchical structure, but imposes the sequential pathway of the process flowsheet to order the regressor blocks (\mathbf{X}_b), and sequentially extract information from each of them. The model structure is given by equations 1-12 and is shown schematically in Fig. 1. Note that although any number of \mathbf{X}_b blocks ($b=1, 2, \dots, B$) can be used in SMB-PLS, the calculations are presented only for the first block in the sequence \mathbf{X}_1 (i.e. $b=1$ in eqs 1-12) and for two regressor blocks ($B=2$) for sake of simplicity in explaining the algorithm. Prior to using SMB-PLS, all the data blocks are mean-centered and scaled to unit variance and block scaling is applied to each \mathbf{X}_b blocks. Then the blocks are ordered according to the process flowsheet with the first block \mathbf{X}_1 containing incoming raw material properties, and process data in the second block \mathbf{X}_2 (could be divided into several blocks according to the sequence of process units but not done here for simplicity). After initializing \mathbf{u} with a column of \mathbf{Y} , the iterative algorithm starts by regressing \mathbf{u} onto \mathbf{X}_1 to obtained the block weights $\mathbf{w}_{X_1}^T$ and block scores \mathbf{t}_{X_1} (eqs. 1,2). The subsequent block \mathbf{X}_2 is then orthogonalized with respect to \mathbf{t}_{X_1} (eq. 3) to yield two new blocks: variations correlated with and orthogonal to \mathbf{X}_1 (\mathbf{X}_2^{corr} and \mathbf{X}_2^{orth} respectively). The loadings and scores of the 2nd block $\mathbf{w}_{X_2}^{corr,T}$ and $\mathbf{t}_{X_2}^{corr,T}$ are obtained by regressing \mathbf{u} onto \mathbf{X}_2^{corr} (eq. 4,5). All block scores are then combined in the super level score matrix \mathbf{T} (eq. 6) and a PLS model is built between \mathbf{Y} and \mathbf{T} to obtain the super weights \mathbf{w}_T^T and super scores \mathbf{t}_T (eqs. 7-10), just as in MB-PLS. Up to this point, \mathbf{X}_1 and all information in the subsequent blocks that is correlated with \mathbf{X}_1 is used to explain the variations in \mathbf{Y} . Equations 1-10 are repeated until convergence on \mathbf{u} . Upon convergence, the \mathbf{X}_b blocks are deflated using the super-scores as in MB-PLS (eqs. 11-12). Then \mathbf{X}_b and \mathbf{Y} are replaced respectively by \mathbf{E}_b and \mathbf{F} and the iterations are repeated to compute the next component. When additional components have depleted all information in \mathbf{Y} related with \mathbf{X}_1 , the procedure is repeated for the next block in the sequence by setting $b=2$ in eq. 1-12 and replacing the \mathbf{X}_b 's ($b=2, \dots, B$) and \mathbf{Y} by their residuals left after modelling

with \mathbf{X}_1 . The next components now based on \mathbf{X}_2^{orth} will only capture new orthogonal information not explained by \mathbf{X}_1 . Note that for the last block in the sequence ($b=B$), \mathbf{X}_2 in this example, a normal PLS is built between \mathbf{E}_B and \mathbf{F} .

$$\mathbf{w}_b = \mathbf{X}_b^T \mathbf{u} / \mathbf{u}^T \mathbf{u} \quad \text{and} \quad \|\mathbf{w}_b\| = 1 \quad (1)$$

$$\mathbf{t}_b = \mathbf{X}_b \mathbf{w}_b \quad (2)$$

$$\mathbf{X}_{b+k}^{Corr} = \mathbf{t}_{X_b} \left(\mathbf{t}_{X_b}^T \mathbf{t}_{X_b} \right)^{-1} \mathbf{t}_{X_b}^T \mathbf{X}_{b+k} \quad k = 1, 2, \dots, B-b \quad (3)$$

$$\mathbf{w}_{b+k} = \mathbf{X}_{b+k}^{Corr,T} \mathbf{u} / \mathbf{u}^T \mathbf{u} \quad \text{and} \quad \|\mathbf{w}_{b+k}\| = 1 \quad (4)$$

$$\mathbf{t}_{b+k} = \mathbf{X}_{b+k}^{Corr} \mathbf{w}_{b+k} \quad (5)$$

$$\mathbf{T} = \left[\mathbf{t}_b, \mathbf{t}_{b+k}^{Corr}, \dots, \mathbf{t}_B^{Corr} \right] \quad (6)$$

$$\mathbf{w}_T = (\mathbf{T}^T \mathbf{u} / \mathbf{u}^T \mathbf{u}) \quad \text{and} \quad \|\mathbf{w}_T\| = 1 \quad (7)$$

$$\mathbf{t}_T = \mathbf{T} \mathbf{w}_T / \mathbf{w}_T^T \mathbf{w}_T \quad (8)$$

$$\mathbf{q} = \mathbf{Y}^T \mathbf{t}_T / \mathbf{t}_T^T \mathbf{t}_T \quad (9)$$

$$\mathbf{u} = \mathbf{Y} \mathbf{q} / \mathbf{q}^T \mathbf{q} \quad (10)$$

$$\mathbf{P}_b = \mathbf{X}_b^T \mathbf{t}_T / (\mathbf{t}_T^T \mathbf{t}_T) \quad (11)$$

$$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_T \mathbf{P}_b^T \quad \text{and} \quad \mathbf{F} = \mathbf{Y} - \mathbf{t}_T \mathbf{q}^T \quad (12)$$

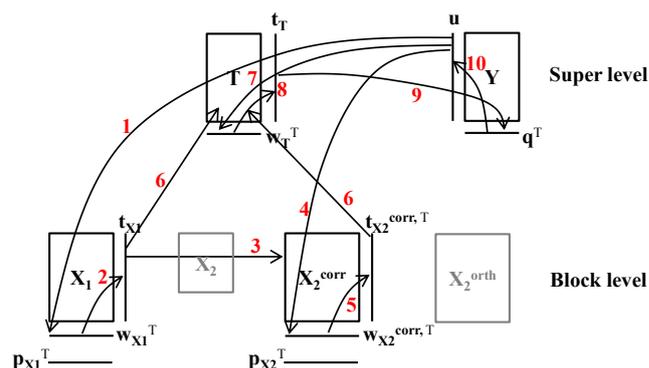


Fig. 1. Schematic of SMB-PLS algorithm for two \mathbf{X}_b blocks. Numbers correspond to each step of the iterative algorithm.

A different number of components can be selected for each \mathbf{X}_b block as opposed to MB-PLS for which the number is the same for all blocks. The number of components is selected similarly as for any other latent variable method using, for instance, cross-validation and root mean square error of prediction (RMSEP) and Q^2 statistics (variance of \mathbf{Y} explained on a prediction set). These were used in this study.

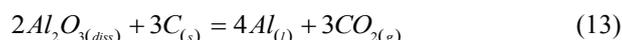
Consider that raw materials properties and process variables are grouped in \mathbf{X}_1 and \mathbf{X}_2 , respectively. SMB-PLS captures the impact of variations in raw materials properties on the process and on \mathbf{Y} in the first modelling step. This allows identifying feedback/feedforward control to compensate for raw materials properties. Establishing specifications in the latent variables space of $[\mathbf{Z} \mathbf{X}_2^{corr}]$ aims at penalizing those combinations of raw materials properties that are not compensated for by the current control schemes. In the second step, \mathbf{X}_2^{orth} captures process variations that are

independent from raw materials and also affect **Y**, e.g. process disturbances, operating policies, etc.

Note that the pathway structure impose in the current version of SMB-PLS is appropriate for a large number of “linear” flowsheets, such as those found in several industries (polymer processing, pharmaceutical, metallurgical, etc.) but is not suitable for highly integrated chemical plants with complex recycle structures. Modifications to the algorithm will be explored in the future. Also, if raw materials do not come in lots but rather continuously in pipelines, one just needs to synchronize raw material properties with process conditions prior to adding lags and modelling.

3. CASE STUDY: ALUMINUM SMELTING

The Hall-Heroult process (Grjotheim, 2010) is used for commercial production of primary metal aluminum. Liquid aluminum is produced by electrolytic reduction of aluminum oxide (alumina) dissolved in cryolite. The reduction cells (or pots) consist of a steel shell lined with refractory and thermal insulation materials, cathode block and collector bars. Several consumable carbon anodes are dipped into the electrolyte. The alumina is consumed electrochemically by which aluminum is reduced on the cathode surface and oxygen is released on the anode. A pool of liquid aluminum is formed on top of the cathode block underneath the cryolite (bath) layer. Oxygen reacts with carbon and CO/CO₂ is produced. The overall alumina reduction reaction is as follows:



Alumina is charged periodically into the cells to control the alumina concentration close to the lowest bath electrical resistivity. Molten aluminum is tapped over 24 h periods and consumed anodes are replaced every 20-30 days.

Current efficiency (CE), energy consumption (EC) and profit were used as performance metrics in the **Y** block. Current efficiency (%) is the ratio of metal produced over 24 hours to theoretical metal production obtained from Faraday’s law. Energy consumption (kWh/kg) is the energy required to produce a unit mass of aluminum. Profit (\$) is the economical profit per pot per day and is a function of CE, operating current and voltage (more details in Tessier et al., 2012).

Data was obtained from Alcoa’s Deschambault smelter (Quebec, Canada), operating prebaked anode technology (AP-30). It was collected from 26 pots with similar operating policy between September 2013 and May 2014. It is known in the field that the properties of both raw materials, alumina (Wang, 2009) and baked carbon anodes (Jentoftsen et al., 2009), strongly affect reduction cell performance. Control actions periodically implemented at the plant to counteract negative effects cannot attenuate them completely.

Alumina was supplied by two different suppliers (binary variables), and characterized by suppliers for particle size distribution, chemical impurities, specific surface area, LOI, bulk density and angle of repose (24 properties in total). Baked anodes were characterized at the smelter for physical and mechanical properties, chemical impurities, and air and CO₂ reactivities. These anode properties were measured from anode core samples on a weekly basis (16 properties total).

The properties of both raw materials were collected in a single block **Z**. Pot operating data are recorded for each pot and used for process control and monitoring. It includes bath level and temperature, excess AlF₃, electric current, feeding strategy and intervals, bath and metal composition, anode set cycle, and voltage variations were used for this study. These 49 variables were stored in the pot block **X**.

Weekly averages of variables were used in the data blocks because some variables, e.g. anode properties, are not measured on daily basis and the variables have different sampling or measurement frequencies. In addition, using weekly averages helps reducing the uncertainties that exist in the amount of metal produced within a 24 hour period which, in turn, corrupts CE values. More details on CE measurement uncertainties can be found in Tessier et al. (2012).

Variations in raw materials and process operation do not show instantaneous influence on the pot performance. Hence, a lag structure was added to the data prior to modelling in order to account for process dynamics. A dead-time of 4 weeks was used between the anode block and **Y** because the anodes are set in the pots a month after they are produced. Then, the anode block was augmented with 4 lags of 1 week to account for the fact that the 40 anodes populating a pot at any given time have a different age. The anodes are replaced every about 20-30 days (about 2/day/pot) because they are consumed by the reaction. At a given time, the newest anode was produced 4 weeks earlier (dead-time) while the oldest was manufactured 4 weeks plus 20-30 days in the past. Adding 4 lags was found adequate to cover the range of properties of the population of anodes in a pot in a given week. The alumina and pot blocks were also augmented with 4 lags of one week to account for dynamics. Adding more lags did not improve the model performance. Figure 2 presents the final data arrangement.

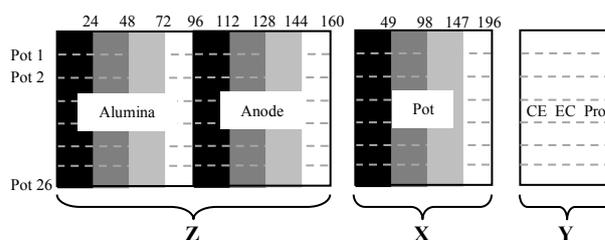


Fig. 2. Data sets with process dynamics. Colors indicate the **Z** and **X** lag structure to account for process dynamics.

4. RESULTS AND DISCUSSION

Two components (t_1 and t_2) were found sufficient to capture the impact of raw material properties **Z** (and correlated process data X^{corr}) on **Y** in the 1st modelling step. Two additional components (t_3 and t_4) were also needed to model the effect of orthogonal variations in the pot data X^{orth} on the remaining variations in **Y**. Table 1 presents the cumulative variance of the **Z** and **X** used to explain **Y** variations. The first two components captured 43.3% of information in **Z** and 8.6% of information in **X** that was correlated with **Z** (X^{corr}), to explain 11.9% in **Y**. Components 3 and 4 explained an additional 20% of the variations in **X** (i.e. X^{orth} block) and 21% of the variations in **Y**. Fig. 3 shows observed vs.

predicted profit for calibration and validation models. Although explaining 33% of Y seems low, the Y -residuals showed no significant autocorrelation after 4 components (white noise), hence there was no more structured information to extract from Y . This is explained by the fact that production data was used (no designed experiments) and because control actions attenuate an important proportion of the variance in raw material properties and other process disturbances. Would the controllers have been turned off, a much larger amount of variance could have been explained. Another reason for low prediction ability of the model was the uncertainties in CE are important. Yet, the fact that 33% of the variance of Y is explained by the model means that room for improvement exist, and establishing specs regions for raw materials and better operating regions for the process could help removing further variance. Also note that the model performance on validation data (1/3 of the data set) is has as good as on the training set (no over-fitting).

Table 1. Cumulative percent variance explained by the SMB-PLS model for Z , X and Y data matrices.

Block	Comp.	Z	X	Y
Z and X^{corr}	1	38.31	7.79	4.94
Z and X^{corr}	2	43.27	8.56	11.89
X^{orth}	3	43.27	21.36	24.64
X^{orth}	4	43.27	28	33.23

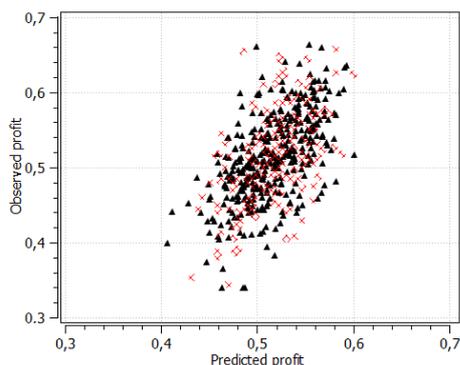


Fig. 3. Profit predicted by SMB-PLS vs. observed profit (Solid triangles: calibration data, Crosses: validation data).

Fig. 4 compares the Y -variance explained by MB-PLS and the new SMB-PLS algorithm. The number of components were selected using the same criteria for both methods. Although the results look similar there are some difference worth discussing. MB-PLS after 4 components (obtained from cross validation) had a higher R^2 for calibration model but a lower for validation compared to SMB-PLS with 2 components for each block. This may lead to suspect a slight over-fitting of the data with MB-PLS, likely caused by imposing the same number of components to each block.

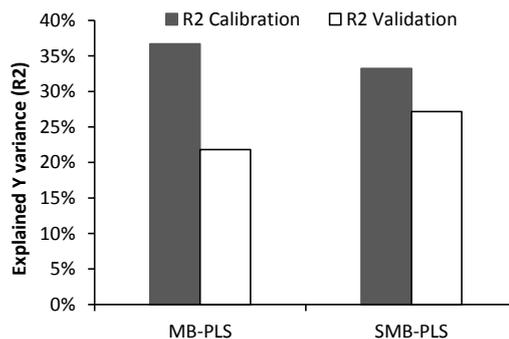


Fig. 4. Variance of pot performance metrics (Y) explained by MB-PLS and SMB-PLS on calibration and validation data.

Fig. 5 shows the relative contribution of the Z and X blocks to each component of the MB-PLS and SMB-PLS models, defined as the square of the super weights of a block in each component. The figure clearly shows that MB-PLS distributes the information captured from X almost evenly in all components due to between block correlations. Thus model interpretation becomes more complex. In SMB-PLS, however, adjustments to manipulated variables in response to raw material variations and changes in state variables caused by raw materials are captured in the first two components as will be shown next. The last two components capture only information from process variables that are orthogonal to raw materials (e.g. process disturbances, operating policies, etc.).

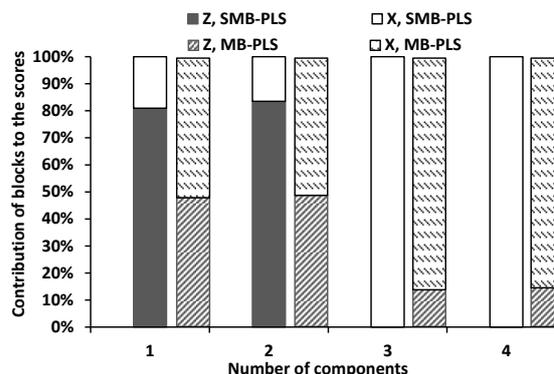


Fig. 5. Relative contributions of Z and X blocks to each component for MB-PLS and SMB-PLS algorithms.

Figure 6 presents the t_1 - t_2 latent variable space of raw material properties and correlated process variations (Z and X^{corr}) and that of the orthogonal process block X^{orth} (t_3 - t_4). The color map indicates the predicted profit ($Y_{pred} = \mathbf{t} \mathbf{q}^T$) in scaled units. Fig. 6.a clearly shows that variations driven by raw materials affected profit in spite of control actions since the observations cluster from lower left to upper right of the plot according to increasing profit. Therefore, certain combinations of raw material properties seem associated with lower performance. A similar observation can be made for orthogonal process variations (Fig. 6.b). The high profit region lies in the positive score values.

The straight line in Fig. 6.a defines the joint multivariate specification region for both incoming raw materials (anodes and alumina) corresponding to a given scaled profit. Lots of raw materials projecting below the line would need to be

avoided to help maintain a certain level of performance as these variations cannot be compensated by the actual control scheme. A profit value in the lower range of about -0.4 was selected for illustration purposes. Setting the limit at higher value would increase profit at the expense of rejecting a wider range of materials and possibly higher purchasing costs due to demand for higher quality materials. Hence, deciding where to set the limit results from a compromise.

Fig. 7 provides an overlay of Z , X^{corr} and Y loadings that will be used to interpret the relationships between the variables with respect to variations in alumina and anode properties (captured by t_1 and t_2). The variables in the same quadrant are positively correlated and those in opposite quadrants are negatively correlated. Among the 160 variables, only the most important ones (i.e. $VIP > 1.2$) are shown in the plot. The numbers before the variable names show the week index. For example, 1_A is the average of variable A for the current week for which pot performance is calculated, and 4_A is the average of variable A measured three weeks ago.

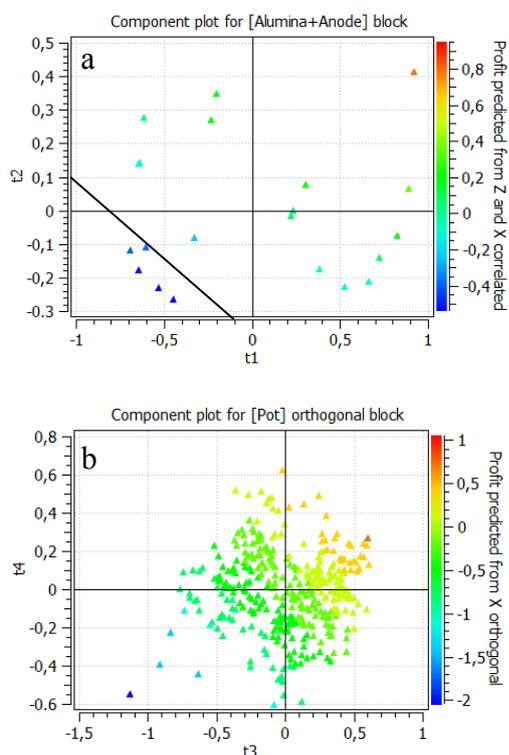


Fig. 6. Predicted profit projected on (a) Z block scores and (b) X^{orth} block scores.

Anode air permeability is clearly one of the main factors that cannot be compensated for and contributed to lower pot performance (cluster opposite to CE and profit in Fig. 7), a well known effect (Grjotheim, 2010). V_2O_5 impurity in alumina is also known as a factor that reduces the current efficiency (Grjotheim, 2010). High angle of repose for alumina powder enhances the crust formation on top of the pot and reduces the alumina dissolution (Grjotheim, 2010). This variable was mostly captured by t_2 and showed a negative impact on CE. Alumina particles smaller than 20 micron revealed an enhancing influence on CE. Although finer particles may improve the dissolution rate of alumina in

the bath, this observation is in contrast with general agreement in the literature where <20 micron alumina should be limited (Grjotheim, 2010). It should be noted that this work is data driven and obtained results are highly affected by datasets used to train the algorithms.

The X^{corr} weights were also overlaid in Fig. 8 to interpret the synergy between the pot operating variables and materials properties. This figure leads to identify the two main control schemes implemented to compensate for raw material properties: the pot resistance and bath chemistry controls. The first component captured the pot control decisions made to compensate the effect of materials variations on the pot resistance. Anode electrical resistivity, reactivity dust (CRD) and alumina particles smaller than 20 micron are among the variables that increase the cell resistance. Fig. 8 shows that these variables are in opposite direction with average pot resistance with respect to w_1 . This means that when materials properties cause an increase in the resistance of the cell, actions are taken to reduce the pot resistance, for example anode-to-cathode distance is reduced. The second component captures bath chemistry control actions. Low AlF_3 concentration in the bath increases the bath liquidus temperature and bath temperature. Therefore, more fluoride is added to reduce the bath temperature. When Na_2O in alumina is increased more AlF_3 is added to the bath to form cryolite.

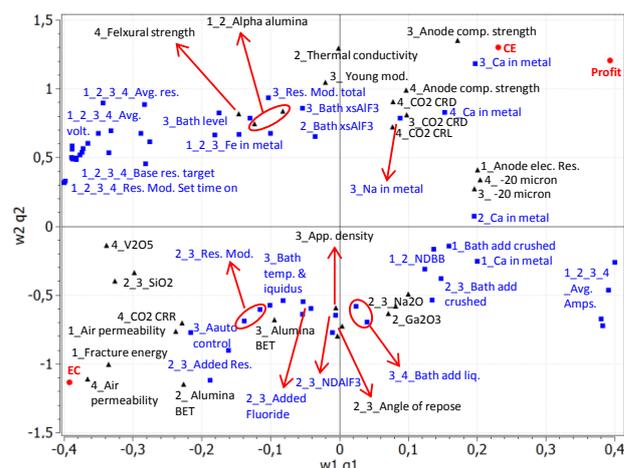


Fig. 7. Overlay of Z (black) and X^{corr} (blue) weights and Y loadings (red).

Fig. 8 shows the loading biplot of the X^{orth} weights and Y loadings again for the most important variables ($VIP > 1.2$). This plot allows interpreting those process variations that affected Y but are independent from raw materials. The pot variables appearing in this Figure, i.e. pot age, bath level, AlF_3 target and feeding strategy, are independent of raw materials properties (i.e. operating policies). When rich alumina feeding strategy is used and the number of alumina feeds is increased (lower feeding interval) the alumina concentration in the bath is increased that enhances CE. Increasing the bath excess AlF_3 and CaF_2 also helps reducing the bath liquidus temperature, accelerates alumina dissolution and subsequently lead to a higher CE. RF is an indication of the amount of tapped metal according to metal tapping table and showed a strong and positive relationship with CE.

Sodium content of metal revealed a positive correlation with CE as is widely accepted in the literature (Grjotheim, 2010, Tabereaux, 1996). The third component shows the negative impact of pot age on CE. High bath liquidus temperature and bath measured temperature resulted in higher EC.

A similar analysis of MB-PLS loadings could not be added here due to space limitation. The influence of orthogonal process variations on pot performance could be identified. However, SMB-PLS provided more insights and results were in better agreement with process knowledge for the effects of materials variations and correlated process variables (control schemes mainly). Some of the MB-PLS results were difficult to explain without considering the correlations between materials properties and process data.

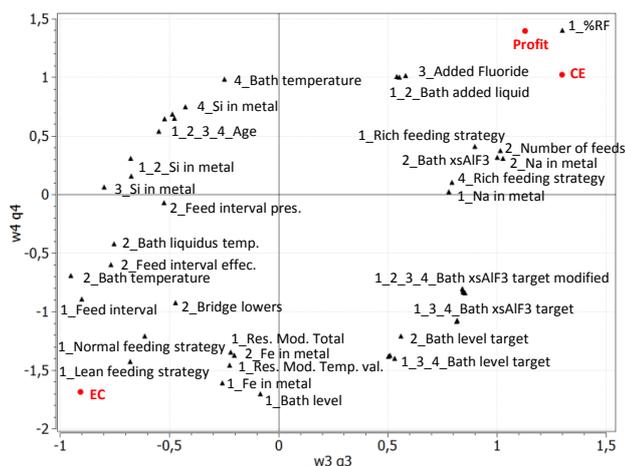


Fig. 8. Weights for X^{orth} (black) and Y loadings (red).

5. CONCLUSIONS

Efforts are being made in several industries to establish specification regions for raw materials. For processes using more than one type of raw materials and several process units where process parameters are correlated with materials properties or with previous sections parameters, it is difficult to determine the specification regions. This work presents an early attempt at establishing specification regions for multiple types of raw materials, taking into account the variations compensated for by control actions.

The Sequential multi-block PLS algorithm (SMB-PLS) was applied to differentiate the between block correlation from the orthogonal (new) information available in each blocks of a multi-block data structure by imposing the sequential pathway of a process flowsheet. Each step of this approach focuses on the effect of a given regressor block on subsequent ones and Y and thus captures only new information that is not modelled by previous blocks. In this way, specification regions could be established jointly on all materials in a way to identify those combinations of properties that are not attenuated by the current control schemes. Although SMB-PLS is appropriate for many sequential production routes where process dynamics can be defined, modifications may be needed for specific processes.

The approach is illustrated using a dataset from an aluminum smelting process and it was observed that variations in raw

materials, i.e. alumina and anode, induced some effects on the process operation. In particular, it was found that variations in anode air permeability and some mechanical properties affected process performance in spite of the control schemes currently implemented on that process. These combination of properties (high permeability, low compressive strength, etc.) need to be avoided to prevent negative impact on process performance.

REFERENCES

- De Smet, J. A. (1993). *Development of multivariate specification limits using partial least squares regression*. McMaster University.
- Duchesne, C. and MacGregor, J. F. (2004). Establishing multivariate specification regions for incoming materials. *Journal of Quality Technology*, 36 (1), 78-94.
- Eriksson, I., Johansson, E., N. Kettaneh-Wold, N. and Wold, S. (2001). *Multi- and Megavariate Data Analysis. Principles and Applications*. Umetrics Academy, Umeå.
- Grjotheim, K. (2010). *Introduction to aluminium electrolysis: Understanding the Hall-Héroult process*, Alu Media, Düsseldorf.
- Jentoftsen, T. E., Linga, H., Holden, I., Aga, B. E., Christensen, V. G. and Hoff, F. (2009). Correlation between anode properties and cell performance. *In: Bearne, G., ed. Light Metals, 2009 Warrendale. Minerals, Metals & Materials Soc*, 301-304.
- Kourti, T. (2005). Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing*, 19 (4), 213-246.
- Lauzon-Gauthier, J. and Duchesne, C. (2014). A new multi-block PLS algorithm including a sequential pathway. *In: EuroPact 2014, May 6-9 2014 Barcelona, Spain*. 130.
- Muteki, K. and MacGregor, J. F. (2008). Optimal purchasing of raw materials: A data-driven approach. *Aiche Journal*, 54 (6), 1554-1559.
- Tabereaux, A. T. (1996). The role of sodium in aluminum electrolysis: A possible indicator of cell performance. *In: Hale, W., ed. Light Metals, 1996 Warrendale. Minerals, Metals & Materials Soc*, 319-326.
- Tessier, J., Duchesne, C., Tarcy, G. P., Gauthier, C. and Dufour, G. (2012). Multivariate Analysis and Monitoring of the Performance of Aluminum Reduction Cells. *Industrial & Engineering Chemistry Research*, 51 (3), 1311-1323.
- Wang, X. W. (2009). Alumina dissolution in aluminum smelting electrolyte. *In: Bearne, G., ed. Light Metals, 2009 Warrendale. Minerals, Metals & Materials Soc*, 383-388.
- Westerhuis, J. A., Kourti, T. and MacGregor, J. F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, 12 (5), 301-321.