Light Metals 2011 *Edited by: Stephen J. Lindsay*
TMS (The Minerals, Metals & Materials Society), 2011

Light Metals 2011

ELECTRODE TECHNOLOGY for ALUMINUM PRODUCTION

Anode Quality and Rodding Processes

SESSION CHAIR
Nigel Backhouse
Rio Tinto Alcan
Saguenay, Quebec, Canada

Light Metals 2011 *Edited by: Stephen J. Lindsay*
TMS (The Minerals, Metals & Materials Society), 2011

MULTIVARIATE MONITORING OF THE PREBAKED ANODE MANUFACTURING PROCESS AND ANODE QUALITY

Julien Lauzon-Gauthier[1], Carl Duchesne[1], Jayson Tessier[2], Katie Cantin[3], Isabelle Petit[3]
[1] Aluminium Research Centre-REGAL, Chemical Engineering Department, Université Laval, Quebec, City, QC, G1K 7P4, Canada
[2] STAS, Alcoa-STAS R&D Team, Chicoutimi, QC, G7K 1H1, Canada
[3]Alcoa inc., Aluminerie de Deschambault, Deschambault, QC, Canada, G0A 1S0

Keywords: Prebaked anode, anode quality, statistical analysis, PLS

Abstract

Prebaked anode quality is available through a weekly average of core sample laboratory measurements. Unfortunately, long delays between production and results (approximately 4-6 weeks) can lead to poor abnormal operation and faulty anode detection, and difficult process control. Extensive raw material and process data are available at Alcoa Deschambault smelter's carbon plant. Using projection to latent structures, a multivariate statistical method, it was possible to correlate raw material and process conditions to weekly lab results. The effect of different petroleum coke and coal tar pitch was analyzed and instant weekly prediction of anode properties was achieved.

Introduction

Baked anode properties are evaluated using core sample laboratory analysis. Usually, the number of samples taken during production (approximately 1%) is too small to accurately measure the variability of anode properties produced at the carbon plant. Weekly averaged data collected on a number of variables are used to monitor process performance and deviations. Furthermore, laboratory results used to monitor process performance are only available 4 to 6 weeks after the anodes have been produced. Typically, results are available after the anodes have been set in the pots. Detecting deterioration of anode properties and feedback correction of possible root-causes is slow due to these limitations. Critical process conditions are monitored at a much higher frequency at the green mill and baking furnace, but final product quality cannot be measured online.

Alcoa Deschambault has an extensive database on raw material properties, green mill and baking furnace operating conditions. Industrial databases are often not exploited to their full potential due to the enormous quantity of information that needs to be analyzed to extract useful information. The goal of this analysis was to use Deschambault's process database to model and explain variability in baked anode properties. Using multivariate statistical techniques, it is possible to correlate variations in raw material properties and process operating conditions to variations in anode properties. Using this model, it is possible to investigate which combinations of parameters have the greatest influence on anode property variation and to predict anode properties on a weekly basis. This analysis is based on historical process data, with no special measurements other than data normally collected during plant operation. It will be demonstrated that traceability of the effects of raw material on anode quality can be achieved.

This paper explains in detail how the analysis was performed. First, the dataset used for the study is described in detail. A brief description of the method used is presented, followed by an investigation of the causes of anode quality variations. Finally, the anode quality prediction will be discussed.

Process and data

The data used in this article were collected from the Alcoa Deschambault smelter (ADQ), located in Deschambault near Quebec City. The ADQ smelter operates 264 AP-30 reduction cells with 40 anodes each. The smelter's carbon plant produces more than 150 000 anodes annually in order to fulfill the potroom needs. The green anode plant has a capacity of 30 tons per hour with continuous mixers and vibrocompactors. There are two baking furnaces with 34 sections each. Weekly data was collected from: raw material lab analysis and supplier's certificate of analysis (COA), green mill process and baking furnace data historian, and core sample laboratory analysis. The time spanned by the dataset is from December 29, 2008 to July 26, 2010 for a total of 82 weeks. During this period, Deschambault used 6 different coke suppliers and 2 pitch suppliers. The anode recipe usually combines 2-3 types of cokes and a single type of coal tar pitch. The different raw material blends used throughout the analysis period are presented in Table I.

Table I: Raw material blends processed during the analysis period.

Raw material blend	Coke 1	Coke 2	Coke 3	Pitch
1	A	B	D	1
2	A	C	D	1
3	A	C	E	1
4	A	C	E	2
5	D	E	-	2
6	D	F	-	2

*Coke suppliers are identified by letters A-F and pitch suppliers with numbers 1-2.

Deschambault has some blending capability that was discussed in the paper by Gendron et al. [1]. All the data presented in this paper have been mean-centered and scaled to unit variance (i.e. auto-scaled). The variables included in the analysis are presented in Table II.

Multivariate analysis

The dataset used in this analysis contains a large number of variables. Industrial databases are noisy and typically contain a certain percentage of missing data. The variables are also generally highly collinear. All of these situations can cause problems when using classical multilinear least-squares regression analysis. To handle these data issues, a multivariate latent variable called Projection to Latent Structure (PLS) was used.

Table II: Variables included in the analysis divided into a regressor block **X** and a response block **Y**

Var ID	X-variable	Var ID	X-variable	Var ID	X-variable	Var ID	Y-variable
1	Coke real density	17	Pitch Dist	33	Mixer 1power	47	% Air dust
2	Coke Na	18	VC bellows P	34	Mixer 1 delta power	48	% Air Rx
3	Coke Ca	19	% Pitch	35	Mixer 2 power	49	% CO2 dust
4	Coke S	20	% Coarse	36	Mixer 2 load	50	% CO2 Rx
5	Coke V	21	% Inter	37	HX paste T	51	Flexu Strength
6	Coke app dens	22	% Fines	38	Pitch Temp	52	Fracture Energy
7	Rec butts Ca	23	% Butts recyc	39	Pitch/paste Temp delta	53	Thermal conduc
8	Rec butt Na	24	% Green recyc	40	% Green scrap	54	Coeff Expansion
9	Rec butt S	25	Coarse Rt4	41	Fire weight loss	55	Elec resistance
10	Rec butt V	26	Inter Rt50+Rt100	42	Baked weight	56	Lc
11	Rec butt Na/Ca	27	Fines Pt200	43	Fire cycle time	57	Young Mod
12	Pitch SP	28	Butts rec Rt3/8	44	Fire start Temp	58	App dens baked
13	Pitch TI	29	Butts rec Rt3/8+Rt4	45	Fire final Temp	59	Real dens baked
14	Pitch QI	30	Aggregate Rt3/8	46	% Baked scrap		
15	Pitch Coking val	31	Aggregate Pt200				
16	Pitch S	32	Green app dens				

This technique, and other multivariate methods, have been presented in recent TMS papers [2-5] for the analysis of primary aluminum smelter data and in other journals for different applications [6]. The PLS regression method uses all the available **X** and **Y** variables and projects them onto a lower dimensional subspace called the latent variable space. These new latent variables (LV) are a linear combination of the original variables and are computed so as to maximize the covariance between the **X** and **Y** variables. The latent variables can be viewed as a small number of lurking variables driving the process, and hence the **X** and **Y** data, in certain correlated directions. The structure of the PLS model is shown below.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{F} \quad (2)$$

$$\mathbf{T} = \mathbf{XW^*} \quad (3)$$

$$\mathbf{W^*} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \quad (4)$$

$$\mathrm{VIP}_{j,A} = \sqrt{J\sum_{a=1}^{A} w_{aj}^2 \left(SSY_a / SSY_{tot}\right)} \quad (5)$$

In the equations above, the **P** and **Q** matrices contain the loading vectors that best represent the **X** and **Y** spaces respectively, whereas **W*** contains the loading vectors that define the relationship between the **X** and the **Y** spaces. Projection residuals of each space are stored in **E** and **F**. The variable importance in projection (VIP) is an indication of the importance of a variable in predicting the **Y** space. In the equation (5), w_{aj} is the loading weight of the j^{th} variable in the a^{th} PLS latent variable, SSY_a is the sum of squares of **Y** explained by the a^{th} LV of the PLS model and SSY_{tot} is the sum of squares of **Y** explained by the model.

Results

PLS regression model

A PLS regression analysis was performed on the dataset which contains 82 observations, 46 **X**-variables and 13 **Y**-variables. The model contains five latent variables (LVs). They were selected by a cross-validation procedure to maximize predictive ability and to minimize overfitting. This technique leaves a randomized group of observations out of the dataset. A PLS model with *a* latent variables is built on the remaining observations, and then used to compute the prediction error sum of squares (PRESS) of the group of data left out of the model. This procedure is repeated until each observation (i.e. row of **X** and **Y**) have been left out once. The overall prediction error sum of squares for a PLS model with *a* latent variables PRESS(*a*) is then computed. The selected number of LVs (*a*) is that one minimizing prediction errors (PRESS). A summary of the model is presented in Table III. It shows the cumulative variance of the **X** and the **Y** (R^2X and R^2Y) data blocks explained by the first 5 PCs, as well as the cumulative variance of **Y** predicted by the model (Q^2) through the cross-validation procedure.

Table III: PLS regression model overview

PC	R^2X(cum)	R^2Y(cum)	Q^2(cum)
1	0,225	0,263	0,237
2	0,316	0,395	0,346
3	0,394	0,451	0,372
4	0,449	0,498	0,381
5	0,501	0,527	0,379

A cumulative Q^2 value of 0,379 appears low, but this value is an overall value computed from all thirteen **Y** variables. Individually, most variables have good predictions. This is shown in Table IV. The same comment applies to the explained R^2X and R^2Y, some variables are well explained and some are not which lowers the overall variance explained. Table IV lists the variance explained for each individual **Y** variables.

The fit or variance explained (R^2Y) for this model ranges from 0,300 to 0,712. Some would consider such a fit as low, but these are good results considering the industrial nature and the level of noise in the data as well as the uncertainties related to the measurement of raw material properties and the residence time within each piece of equipment. The same comments apply to the variance predicted (Q^2) which ranges from 0,126 to 0,610. For some variables, (i.e. Coeff Expansion and Young Mod) the prediction ability of the model is low, but it can be considered good for most variables.

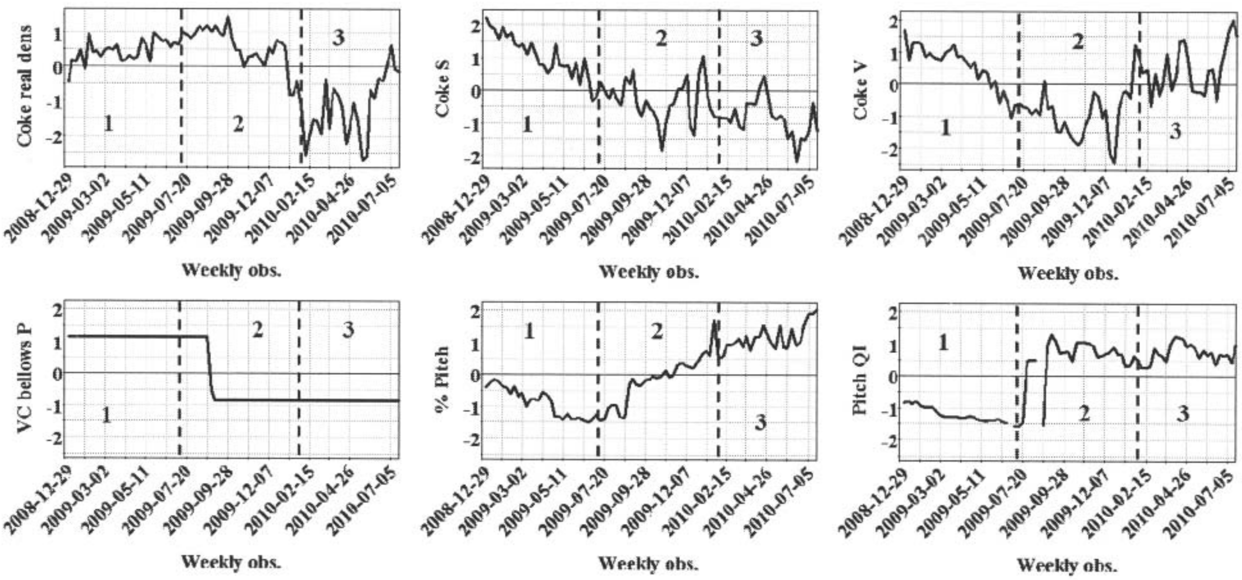

Figure 1: Some raw material properties and process conditions variations.

Table IV: Variance explained for each **Y** variables

Var ID	Variable	R^2Y(cum)	Q^2(cum)
47	% Air dust	0,505	0,341
48	% Air Rx	0,583	0,450
49	% CO2 dust	0,495	0,318
50	% CO2 Rx	0,559	0,381
51	Flexion Strength	0,712	0,610
52	Fracture Energy	0,501	0,253
53	Thermal Conduc	0,509	0,354
54	Coeff Expansion	0,300	0,126
55	Elec Resistance	0,675	0,520
56	Lc	0,613	0,514
57	Young Mod	0,362	0,186
58	App Baked Dens	0,664	0,549
59	Real Baked Dens	0,450	0,341

From the PLS regression model using 5 LVs a list of the 15 most important variables in the prediction (VIP) is presented in Table V, As a rule of thumb, as discussed in papers by Chong and Jun and Ericksson et *al* [7, 8], variables having a VIP over 1 are considered important. Most of the variables listed in Table V are raw material properties. This was expected since Deschambault dealt with six different coke-pitch blends (i.e. different coke/pitch supplier combinations) over the time spanned by this analysis. However, more operating conditions were expected to show a greater influence on the model. For example coke or dry blend size distribution for coarse, intermediate and fines fractions and their ratio in the dry blend mix all had a VIP of less than 1. This can be explained by their lack of variability in the dataset. Except for the pitch ratio in the paste, almost all other operating conditions in the green mill were kept constant. A design of experiments on the operating conditions would have enabled the capture of more information from the process variables. Figure 1 shows the variation in some coke properties, pitch QI and two process operating conditions (vibrocompactor bellows pressure and pitch ratio) with the three different groups of similar process conditions identified.

Table V: VIP

Rank	Var ID	Variable	VIP
1	42	Baked weight	1,539
2	9	Rec butt S	1,518
3	18	VC bellows P	1,458
4	10	Rec butt V	1,421
5	19	% Pitch	1,413
6	4	Coke S	1,369
7	14	Pitch QI	1,345
8	10	Coke V	1,297
9	45	Fire final T	1,252
10	13	Pitch SP	1,239
11	43	Fire cycle t	1,216
12	32	Green app density	1,203
13	15	Pitch Coking val	1,178
14	1	Coke real dens	1,173
15	24	% Green rec	1,070

Process variability investigation

One of the goals of this analysis was to investigate process variability. The plot of the first two latent variables (t_1 and t_2), shown in Figure 2, provides an overview of the information captured by the PLS model on the joint **X-Y** data blocks. The first two LVs are shown because they explain approximately 80% of the modeled variability of the Y variables (i.e. 39,5% out of the 52,7% of the 5 LVs). Some information can be contained in the three other latent variables (around 20%). The results presented in this paper focus on the first two LVs for brevity and also because most of the important information is carried in these two LVs. The 82 markers within this plot correspond to the projection of each multivariate observation onto the plane formed by the first two principal components. Weekly observations projected within a similar region of the t_1-t_2 score plot show similar patterns in their data structure, hence similar in the latent variable space (i.e. similar combinations of raw material properties, recipes, process conditions and anode properties) whereas those falling in distinct regions are different.

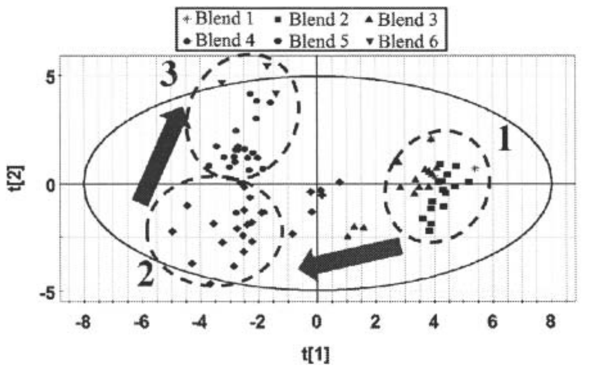

Figure 2: Scatter plot of the first two latent variables of the PLS model.

Three distinct clusters are observed in Figure 2. Group 1 represents baked anodes produced from December 29, 2008 to July 13, 2009, group 2 represents anodes manufactured between September 14, 2009 and January 18, 2010, and group 3 between January 25, 2010 and July 26, 2010. The different causes leading to process movement in the latent variable space will be analyzed. One of the tools that can be used for interrogating the PLS model is the contribution plot, indicating which variables are associated with a movement in the t_1-t_2 latent variable space. Figure 3 represent changes from group 1 to group 2 and Figure 4 represents changes from group 2 to group 3.

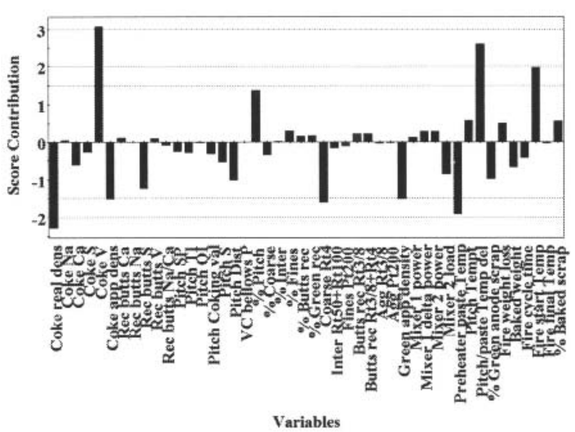

Figure 3: Contributors to changes from group 1 to group 2

The first transition between groups 1 and 2 is associated with several different factors. The first contribution comes from the sulfur and vanadium content of the coke and recycled butts. This change started with blend number 3 due to the change from coke D to coke E. The transition was accentuated by the change in pitch supplier which can be seen by the pitch coking value, quinoline and toluene insolubles. The pitch ratio in the anode was continuously raised during that period to compensate for the increase in QI content and to meet green weight set-point. There was also a step change in the pressure applied on the anode during vibrocompaction and it has been kept constant since then. Green apparent density (GAD) went up in correlation with the pitch ratio. There was a major change in the baked weight. This might be due to the pitch supplier change. The new pitch has a much higher QI and coking value which will lead to more binder cokefaction in the furnace, as discussed in [9, 10].

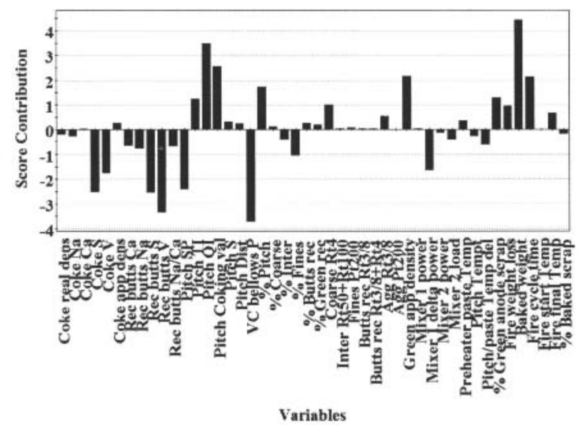

Figure 4: Contributors to changes from group 2 to group 3

Differences between group 2 and group 3 are also due to raw material variations. Coke real density in the new coke blend was lower and these affected GAD, pitch demand and baked weight. It should be noted that process operating conditions except from pitch ratio and vibrocompactor bellows pressure, were kept constant during the time spanned by the analysis. Thus it is not possible to capture the effect of most process variables since they were not changed during the time period considered.

Table VI: Average **Y** predicted for the three different operating groups (auto-scaled values)

Ypred-variable	Group 1	Group 2	Group 3
% Air dust	0,183	-0,623	0,245
% Air Rx	-0,666	0,975	-0,032
% CO_2 dust	-0,584	0,265	0,635
% CO_2 Rx	0,670	-0,289	-0,735
Flexu Strength	-0,952	0,725	0,479
Fracture Energy	-0,689	0,280	0,723
Thermal conduc	-0,749	0,575	0,340
Coeff Expansion	0,328	0,056	-0,584
Elec resistance	0,878	-1,064	-0,091
L_c	0,700	0,049	-0,992
Young Mod	0,486	-0,371	-0,444
App dens baked	-0,439	1,054	-0,450
Real dens baked	0,725	-0,406	-0,542

Averaged autoscaled predicted values for each group are listed in Table VI. Group 1 has the best CO_2 reactivity. Group 2 has the best air reactivity; this can be linked to the lower content of sulfur and vanadium during that period. The real baked density was higher in group 1 since coke real density has been gradually decreasing over time. L_c in baked anode seems to be highly correlated to coke real density. Baking final temperature was increased from group 1 to group 2, but the L_c has decreased. It is difficult to determine which group was better in terms of mechanical properties. Electrical resistance and baked apparent density were higher in group 2. This could be due to higher QI and coking value in the pitch used during that period. It is difficult to determine which of the 3 operational regions led to the best overall anode quality when looking at the quality variables one at a time. An important input would be to know which group had the best behavior in the potroom.

Prediction of anode properties

The second goal of this analysis is to predict anode quality variables. Figures 5 to 8 represent the predicted and measured values on time series graph for L_c, Electrical resistance, % CO_2 Rx and % Air dust. The operational groups described earlier are marked on the charts. The variance explained by the model for each of these variables is listed in Table IV. The prediction model can be useful for a faster detection of large deviation in anode properties. Usually, complete lab results of anode core sample are available only 4 to 6 weeks after sampling. Being able to instantly estimate anode properties fabricated during the week can lead to earlier process deviation detection and faster problem solving. Therefore, such a tool would enable process engineers and technical staff to operate the carbon plant and baking furnace closer to the optimum production state, with respect to anode properties. However, the laboratory analysis of core samples will always be needed to validate model predictions.

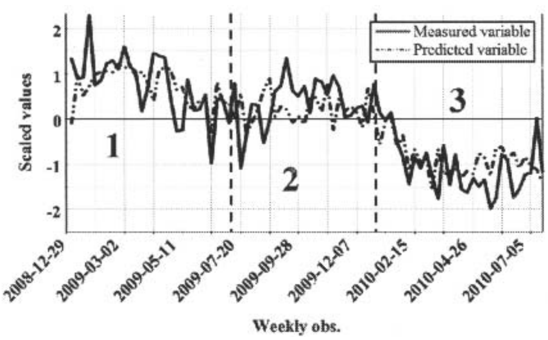

Figure 5: Measured and predicted values for L_c

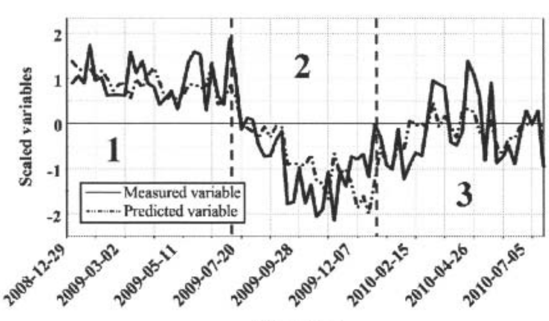

Figure 6: Measured and predicted values for Electrical resistance

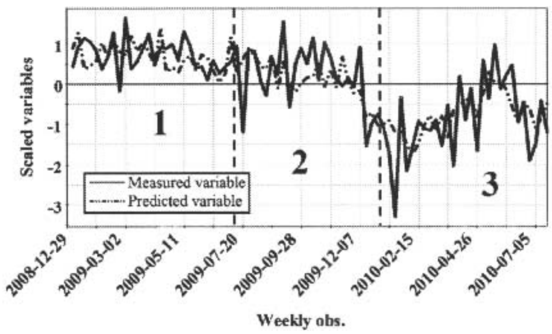

Figure 7: Measured and predicted values for % CO_2 Rx

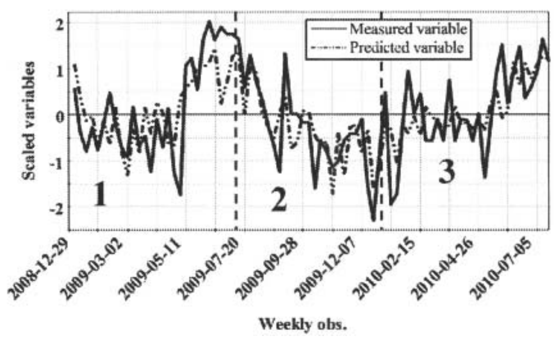

Figure 8: Measured and predicted values for % Air dust

Conclusions

The goals of this analysis were to investigate the causes of anode quality variability and the possibility to predict anode quality using raw material and process information. Using the projection to latent structure regression method, correlations between process data and anodes quality were modeled. Most of the variability affecting anode quality came from raw material variations (coke and pitch binder) mainly due to supplier changes. Since process conditions were kept constant, no variability could be captured. Variation within the same suppliers could be explored by performing separate analysis on each blend. It was also demonstrated that it is possible to predict a number of anode quality variables. Variance captured by the model (obtained by cross validation) ranged from 0,186 to 0,610. Variables with higher explained variance can be predicted on a weekly basis and at least 4 weeks before laboratory results. This can help in detecting problems, investigating them and taking corrective action much faster than when one has to wait for laboratory results.

Acknowledgments

This research was funded by the Fonds québécois de la recherche sur la nature et les technologies (FQRNT), through the Aluminum Research Centre-REGAL, by the National Science and Engineering Research Council (NSERC) and by Alcoa Inc. The authors would like to thank the Alcoa Deschambault anode plant technical staff for their help and Alcoa Inc. for permission to publish this paper.

References

1. Gendron, M., Whelan, S., Cantin, K., 2008, Coke blendind and fines circuit targeting at the alcoa deschambault smelter, *TMS Light Metals 2008*, pp. 861-864

2. Tessier, J., Duchesne, C., Tarcy, G.P., Gauthier, C., Dufour, G., 2008, Analysis of a potroom performance drift, from a multivariate point of view, TMS Light Metals 2008, pp. 319-324

3. Tessier, J., Zwirz, T.G., Tarcy, G.P., Manzini, R.A., 2009, Multivariate statistical process monitoring of reduction cells, TMS Light Metals 2009, pp. 305-310

4. Tessier, J., Doiron, P., 2010, Statistical investigation and modeling of bath level in hall-héroult cells, TMS Light Metals 2010, pp. 553-558

5. Majid, N.A.A., Young, B.R., Taylor, M.P., Chen, J.J.J., 2009, Detecting abnormalities in aluminium reduction cells based on process events using multi-way principal component analysis (MPCA), TMS Light Metals 2009, pp. 589-593

6. Kourti, T., MacGregor, J.F., Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemometrics and Intelligent Laboratory Systems, 1995, Vol. 28, pp. 3-21.

7. Chong, I.-G., Jun, C.H., 2005, Performance of some variable selection methods when multicollinearity is present, Chemometrics and Intelligent Laboratory Systems, Vol. 78, pp. 103-112.

8. Eriksson L., Johansson E., Kettaneh-Wold N., Trygg J., Wikstrom C., Wold S., 2006, Multivariate and Megavariate Data Analysis: Part I - Basic Principles and Applications. Umetrics, Umeå, Sweden, 307p.

9. Wombles, R. H., Baron, J. T. McKinney, S., 2009, Evaluation of the necessary amount of QI in binder pitch, TMS Light Metals 2009, pp. 913-916

10. Krupinski, K. C., Windfelder, J.J., 1992, Effects of tar quinoline insolubles on manufactured carbon performance, 1992 Ironmaking conference proceedings, pp. 487-492