JULIEN LAUZON-GAUTHIER

MULTIVARIATE LATENT VARIABLE MODELLING OF THE PRE-BAKED ANODE MANUFACTURING PROCESS USED IN ALUMINUM SMELTING

Modélisation multivariée par variables latentes du procédé de fabrication des anodes précuites utilisées pour la production d'aluminium primaire

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en Génie chimique
pour l'obtention du grade de Maître ès sciences (M. Sc.)

DÉPARTEMENT DE GÉNIE CHIMIQUE
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2011

Résumé

L'aluminium est fabriqué par un procédé électrolytique. La réaction consomme des anodes de carbone dont la qualité a une grande influence sur l'opération optimale du procédé. Cependant, leurs propriétés sont mesurées hebdomadairement sur moins de 1% de la production. L'objectif de ce projet est d'améliorer le contrôle de la qualité du procédé de fabrication des anodes par la prédiction de leurs propriétés. Une méthode de régression multivariée appelée projection sur structures latentes est utilisée pour relier les propriétés des matières premières et les paramètres d'opération du procédé aux propriétés des anodes cuites recueillies à l'Aluminerie Alcoa de Deschambault. Plusieurs modèles sont étudiés pour les propriétés physiques et la réactivité aux gaz qui expliquent 20% à 68% des variations de celles-ci. Considérant le niveau de bruit élevé des données industrielles, il est jugé qu'une portion significative de la variabilité est modélisée. De plus, l'interprétation de ces modèles est cohérente par rapport aux connaissances du procédé.

Abstract

Aluminum is manufactured by an electrolytic process. The reaction consumes carbon anodes. Anode quality has a great influence on the optimal operation of the reduction process. However, their properties are poorly characterized by weekly averages of anode sample laboratory analyses. The goal of this thesis is to improve quality control at the baked anode manufacturing plant by predicting anode properties. A multivariate latent variable regression method called Projection to Latent Structure (PLS) is used to relate the raw material and the manufacturing process data to the baked anode properties collected at the Alcoa Deschambault smelter. Several models are investigated for physical properties and gas reactivity. From 27% to 68% of the physical properties variance and 20% to 49% of the reactivity variations are captured. The models explained a significant amount of variability, considering that industrial data is typically very noisy. The interpretation of the models was found in agreement with process knowledge.

Acknowledgments

For the last two years I have spent a significant amount of time working on this project. I had the chance to be supported by many people and organizations. This work was made possible through the funding of Alcoa, the *Fonds de Recherche sur la Nature et les Technologies* (FQRNT) and the Aluminum Research Centre – REGAL.

First of all, thanks to Prof. Carl Duchesne. You have given me the opportunity to work on a great project. It has been a rewarding experience with all the travels and the opportunity to collaborate with the industry. Thanks for your incredible availability and support. I look forward to continue working with you for my Ph.D. Also, thanks to Jayson Tessier. You introduced me to the right people at Deschambault and this made my work so much easier. I'm grateful for your encouragement, comments and insight into this project.

I would like to thanks all the people at Deschambault and Alcoa who made this project possible, Isabelle, Francis-Joé, Yves, Réal, Pat, Bruno, Marc, and all of the others. Thanks for spending the time answering my numerous questions. Thanks to Chin Woo and Angelique Adams for the opportunity to attend the Best Practice pre-assessment. Special thanks to Katie Cantin, your meaningful critics of the models and your input in my understanding of the manufacturing process is greatly appreciated.

Thanks to my co-workers at the University. Myriam, thanks for instigating this idea that I could start a project on multivariate analysis. Also thanks to Ryan, Pierre-Marc and Massoud. Thanks to Michèle who helped me extract important data from the databases. Thanks to the REGAL and Mario Fafard for giving a lot of student access to support programs to make our research more interesting. Also, working along with the MACE^3 industrial research Chair was very stimulating.

Merci à mes parents de m'avoir donné la chance et encouragé à poursuivre des études universitaires qui m'ont permis d'accomplir ce projet. Merci de m'appuyer dans tout ce que j'entreprends. Finalement, merci à ma femme Marilou. Ton soutien, ta compréhension et surtout ton amour sont très importants pour moi.

Merci

Table of contents

Table list

Figure list

Chapter 1 Introduction

Aluminum is the third most abundant element on earth after oxygen and silicon and the most abundant metallic element. It constitutes approximately eight percent of the Earth's crust (Encyclopedia Britannica (a)). Due to its high chemical reactivity, the metallic state of aluminum does not occur in nature. It can be found combined to different minerals. The most common aluminum ore is bauxite, a mixture of aluminum hydroxide, aluminum oxide and other impurities. Aluminum is a young metal; industrial production began when Charles Martin Hall (1863-1914) and Louis-Paul Toussaint Héroult (1863-1914) almost simultaneously discovered that it was possible to produce aluminum through an electrolytic process in 1886 (Encyclopedia Britannica (b)). It is now known as the Hall-Héroult process in which an electric current dissociates alumina (Al_2O_3) into aluminum and oxygen. Alumina powder is dissolved into a molten mixture of cryolite (Na_3AlF_6) and aluminum fluoride (AlF_3) salts, commonly called the bath. Aluminum oxide is obtained from bauxite by the Bayer process (Grjotheim and Kvande 1993) and is the main raw material for modern aluminum smelter together with carbon and electricity. Alumina is fed almost continuously in the bath by dedicated point feeders and liquid aluminum is tapped on a regular basis to maintain a constant metal height in the cells. The Hall-Héroult process uses baked carbon anodes, the source of carbon for the electrolytic reaction. The oxygen produced by electrolysis reacts with the carbon to produce carbon dioxide. The global electrochemical reaction is:

$$2Al_2O_{3(diss)} + 3C_{(s)} = 4Al_{(l)} + 3CO_{2(g)} \quad (1)$$

The reaction occurs in reduction cells, also called pots. The Figure 1 presents a sketch of a modern pre-baked cell. They are made of a steel shell lined by insulating materials. Liquid aluminum is produced in the bath layer and settles at the bottom of the cells with the molten bath floating on top of it. Carbon anodes are immersed a certain depth into the bath and are covered by a layer of cover material (i.e. mixture of alumina and crushed bath) which forms a hard protective crust. The active cathode is the bath-metal interface. There are two types

of aluminum reduction technologies currently in use around the world: the Söderberg and the prebaked cells. They differ by the type of anodes used during the operation of the pots. The Söderberg anodes are formed continuously and self baked in the cells. This technology is being slowly replaced by the prebaked technology due to its higher environmental impact and lower productivity. The prebaked technology uses prefabricated anodes that are formed and baked in a dedicated manufacturing plant. As they are continuously consumed, anodes are lowered in the pots to keep the anode cathode distance (ACD) constant and they are replaced periodically when about 2/3 of the original anode is consumed. The residual anodes called butts are reused in the anode manufacturing process.

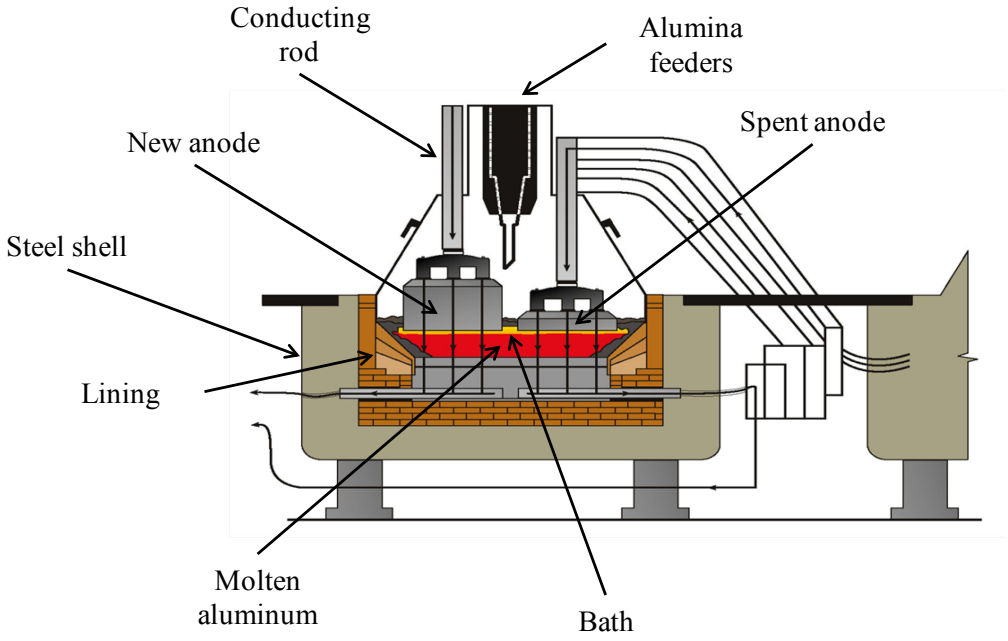


Figure 1 – Cross section of a prebaked reduction cell technology (Courtesy of Alcoa)

Prebaked anodes are made of calcined petroleum coke aggregates, coal tar pitch and recycled anode butts (i.e. crushed residual anodes after consumption in the cells). The materials are mixed together and moulded into blocks called green anodes. These blocks are then baked in a furnace up to a temperature of 1050-1150°C. The baked anode is later assembled to an anode rod (the rod is made of aluminum or copper, the tripod and stubs are made of steel). An anode assembly is presented in Figure 2.

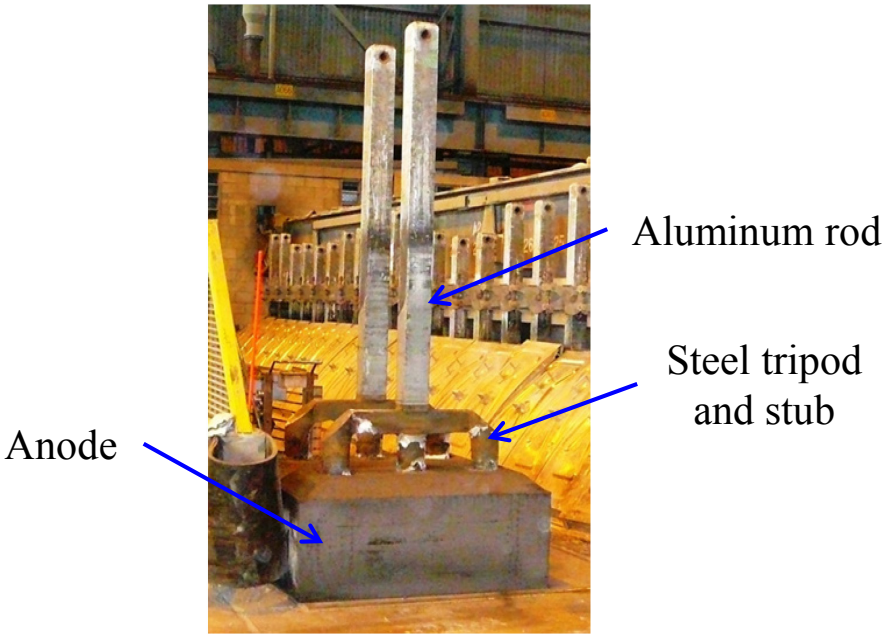


Figure 2 – Anode assembly (Courtesy of Alcoa)

As prebaked anodes are used in the pots, they are consumed by the electrolytic reaction (equation (1)). Spent anodes (i.e. butts) are removed from the pots and reused in the green anode mixture together with fresh coke and pitch.

Baked anode properties are evaluated using core samples analyzed in the laboratory. The number of samples taken during production is typically less than 1% of the weekly anode production. This is not enough to provide a representative measure of the variability of the anode population. Sinclair and Sadler (2009) discuss the issues of different sampling strategies. Furthermore, laboratory results used to monitor process performance are available 4 to 6 weeks after the anodes have been produced, that is, well after the anodes have been set in the pots due to sampling and analysis delay. When a deterioration of anode properties is detected, it is often too late for applying remedial control actions to avoid affecting the performance of reduction cells.

The goal of this thesis is to improve quality control at the baked anode manufacturing plant by predicting anode properties immediately after baking, thus saving the 4-6 weeks delay to obtain anode quality information. This is achieved through developing a soft-sensor based on all the data available at the anode manufacturing plant. A multivariate latent variable modeling approach, Projection to Latent Structure (PLS) in this case, is used to relate the

raw material and the manufacturing process data to the baked anode properties collected at the Alcoa Deschambault smelter. This latent variable model allows for predicting anode properties (i.e. quality attributes) and can also be used for process monitoring.

This thesis is organized as follows. Chapter 2 and Chapter 3 provide background information on the anode raw materials and the manufacturing process, and on the multivariate latent variables methods, respectively. Chapter 4 discusses the data collected at the Alcoa Deschambault smelter and how it was pre-processed. Chapter 5 focuses on building, interpreting and demonstrating the predictive ability of the PLS models. Finally, some conclusions are drawn and future work is discussed.

Chapter 2 Carbon materials and manufacturing process

2.1 Anode manufacturing process

This chapter describes the manufacturing process of prebaked anodes, the traditional quality control strategy based on core sampling and the routine laboratory analyses performed for assessing anode quality. It also discusses in details the effect of carbon materials properties and process parameters on anode properties and their performance in the smelting process. Figure 3 presents the typical steps for the production of anodes (Fischer et al. 1995). The flowsheet is described in details in the following sections.

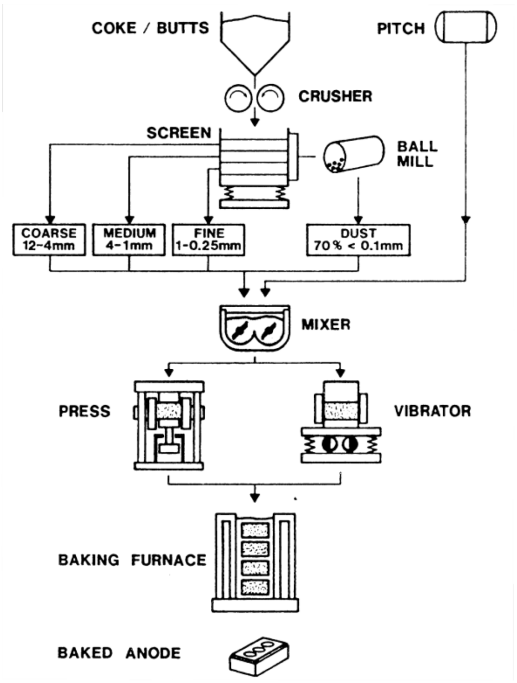


Figure 3 – Anode manufacturing process flowsheet (Fischer et al. 1995)

The anode raw materials consist of calcined petroleum coke, coal tar pitch and recycled anode butts. They are stored in silos at the plant. In the past solid pitch was used, but now most facilities store their pitch in liquid state for health and environmental issues. The coke and butts (after cleaning and crushing) are classified and grinded to the desired particle size

distribution. This is the dry aggregate preparation step. The dry aggregate is composed of the coke particles of different sizes (e.g. coarse, intermediate and fine) and butt particles which also contain the baked scrap anodes. The typical particle size for each fraction is given in Table 1 (Jones 1986).

Table 1 – Typical dry aggregate particle size (Jones 1986)

Dry aggregate		Particle size
Coke	Coarse	-¼ in/+28 mesh
	Intermediate	-28/+100 mesh
	Fine	-100 mesh
Butts & baked scrap		max 1 in

The dry aggregate is then pre-heated. The green scrap anodes are added during this step. It is followed by the mixing of the dry aggregate with the liquid pitch. The green anodes are formed either using a press or a vibro-compactor. The baking of the green anodes is performed in an anode baking furnace up to a temperature of 1050-1200°C (Fischer et al. 1993) for about two weeks. Anodes are made of approximately 65 % of coke, 20% of butts and 15% of pitch in weight (Fischer and Perruchoud 1985). Figure 4 shows the various processing steps and how each of them affects baked anode quality (Fischer et al. 1995). The rest of this chapter explains in details some of these relationships.

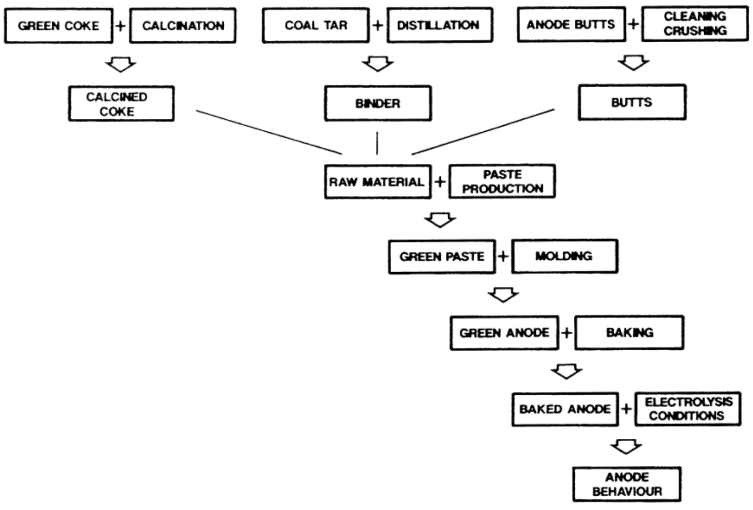


Figure 4 – Relationships between raw materials, process conditions and anode quality (Fischer et al. 1995)

Anode quality is defined by a number of laboratory measurements obtained from core sample taken from baked anodes. These quality attributes (listed in Table 2) are grouped into four categories: physical properties, mechanical properties, reactivity and chemical composition (e.g. contaminants). Details on the laboratory analysis are available in Fischer et al. (1995). For the reactivity analysis, the three following properties are measured: the percentage lost is the material consumed, the dust is the material that is detached from the sample and the non reacted part is called the residue.

Table 2 – Anode properties typically measured from core samples

Properties		**Unit**
Air permeability		nPm
Apparent density		kg/dm^3
Thermal conductivity		W/mK
Electrical resistivity		μohm*cm
Flexural strengh		MPa
Fracture energy		J/m^2
Coefficient of thermal expansion		K^{-1}
Compressive strength		MPa
Young's modulus		GPa
Real density		kg/dm^3
Cristalite size L_c		nm
Ash content		%
CO_2 reactivity	CO_2 loss CRL	%
	CO_2 dust CRD	%
	CO_2 residue CRR	%
Air reactivity	Air loss ARL	%
	Air dust ARD	%
	Air residue ARR	%
Chemical impurities	Sulphur S	%
	Vanadium V	ppm
	Nickel Ni	ppm
	Silicon Si	ppm
	Iron Fe	ppm
	Aluminium Al	ppm
	Sodium Na	ppm
	Calcium Ca	ppm

Anode consumption is often used in the literature to measure the performance of baked anode in the reduction cells. Since it is difficult to relate anode behaviour to individual properties, net carbon consumption is used. The theoretical carbon consumption of the Hall-Héroult process is 0.333 kg C/ kg Al but the operational consumption varies between 0.4 - 0.5 kg C/ kg Al (Grjotheim and Kvande 1993). The excess consumption is related to carbon attack by air and by CO_2 (Boudouard reaction) and also by mechanical failure due to thermal shock and anode cracking (Jones 1986). There are two different modes for gas reactivity. The first is the direct reaction with air or CO_2. The second is the preferential oxidation of the binder matrix which causes some anodes particles to disintegrate and fall in the pots. This phenomenon is called dusting and leads to operational problems in the potroom and to higher carbon overconsumption since this carbon material is not used for metal production.

Some fundamental relationships can be established between the various anode quality variables. Hence, anode properties are typically correlated with one another to some extent. An overview of known relationships is provided below (Fischer et al. 1995).

- A high baked apparent density can be linked to a longer anode life in the pots,
- Apparent density and air permeability are closely related. When density increases, permeability usually decreases,
- A decrease in air permeability also leads to a lower air and CO_2 reactivity rates,
- The apparent density can also be linked to coke bulk density, anode recipe and mixing condition,
- Electrical resistivity and thermal conductivity are closely related and both are strongly influenced by coke quality and by the final baking temperature,
- Flexural strength (FS), fracture energy (FE), compressive strength (CS), Young's modulus and coefficient of thermal expansion (CTE) are all important mechanical properties for anode resistance to thermal shock and anode cracking. Coke structure influences the coefficient of thermal expansion. The higher the CTE, the higher the

other mechanical properties need to be to avoid anode failure in the pots. Mechanical properties are influenced by the coke properties and the manufacturing process. There is a positive relationship between FS, FE, CS and Young's modulus (Perruchoud et al. 2004). The spatial distribution of the properties throughout the anode is also an important aspect of anode quality (i.e. homogeneity).

There are numerous elements that influence anode quality: coke quality, pitch quality butts quality, dry aggregate blend, dust fineness control and addition, quality of mixing and low heat-up rate, final baking temperature and soaking time in the baking furnace. The rest of this chapter presents the most common knowledge on the effects of raw material properties and process parameters on anode quality.

2.2 Petroleum calcined coke

Calcined coke is manufactured from the residual heavy oil fractions of the petroleum refining industry. It is a low value by-product (i.e. waste) and therefore, refineries have no incentive to control and/or improve its quality. The following steps are required to transform heavy oil into coke: a delayed coking process yields the green coke and this process is followed by a calcining operation to produce the calcined coke of interest for the aluminum industry. Calcined coke quality is influenced by the calcining conditions and green coke quality which is influenced by crude oil quality, refining operation and delayed coking operation parameters (Fischer et al. 1995). Several papers describe the effects of oil quality, and process operation on green coke quality (e.g. (Fischer and Perruchoud 1985) and (Vitchus et al. 2001)).

The quantity and quality of calcined cokes available on the market varies quite a lot. This implies that carbon plants need to adapt to cokes having important differences in physical properties and chemical impurities from shipment to shipment (McClung and Ross 2000). These quality variations contribute to variations in anode quality. The critical chemical impurities are vanadium (V), sulfur (S) and sodium (Na). They affect air and CO_2 reactivity by acting as catalyst for the oxidation reactions (Fischer et al. 1995), (Grjotheim and Kvande 1993) and (Houston and Oye 1985). Air and CO_2 reactivity are two important

mechanisms for carbon overconsumption. The coke structure, porosity and density also have important effects on anode quality (Jones 1986).

Several articles investigate the effect of impurities on anode quality. They mainly affect the anode reactivity and aluminum purity but have little impact on the mechanical behaviour, thermal conductivity and resistivity of the baked anode. A review of the different literature associated to coke impurities is presented in Belitskus and Danka (1988). The overall effect of metallic impurities is a catalytic effect on the air and CO_2 reactivity of the anode. This increases the carbon consumption since the reacted carbon is not used to produce aluminum. Vanadium (V), Nickel (Ni) and Calcium (Ca) increase air reactivity and carbon consumption while CO_2 reactivity is mainly affected by sodium (Na) contamination (Perruchoud and Fischer 1991). Other impurities like iron (Fe) and silicon (Si) can lower the metal quality. They do not affect the cell operation but they reduce the economical value of the aluminum produced (i.e. metal grade).

The effect of sulfur is not as clearly understood. Indeed, the literature is divided into two groups. On one hand, some research groups conclude that sulfur increases reactivity. The article by Jones et al. (1979) demonstrates that an increase in sulfur content causes an increase in air reactivity, but a decrease in CO_2 reactivity. According to Jones (1986) and Barrilon and Pinoir (1977), sulfur increases the carbon consumption, which causes environmental concerns due to SO_2 emissions. On the other hand, some authors claim that increasing sulfur decreases the sensitivity of the anode to the attack by CO_2 and air (i.e. oxygen). An increase in sulfur decreases the sodium catalytic effect on anode reactivity by binding in an inactive form during baking, (Hume et al. 1993a). The effect of sulfur was also found by Hume et al. (1993b) to be an inhibitor of oxidation reaction. It is important to remember that the amount of each coke impurity is not independent and that usually, when the sulfur increases, all the other impurities increase in similar proportions (Houston and Oye 1985). Sulfur is considered as an inhibitor of the catalytic effect of the other impurities, especially for CO_2 reactivity (Gendron et al. 2008). There is also a slight desulfurization that can occur during baking for high sulfur coke. This can further increase the problem since the loss of sulfur increases the anode reactivity (Edwards, et al. 2007).

Coke structure is another important property that greatly influences anode properties. The typical good anode grade coke is called the sponge coke. This coke is a mixture of isotropic and anisotropic grains. Isotropic coke is denser, less porous, but less pure and more sensitive to thermal expansion. Anisotropic coke have less impurities and is less sensitive to thermal expansion but is more porous and have more accessible surface for oxidation reactions (Jones 1986). High isotropic coke also reduces the binding ability of pitch due to the low macroporosity (Edwards et al. 2009); pitch penetration within the coke pores is more difficult. The coke structure also influences the anode mechanical properties, thermal conductivity and resistivity. Since approximately 65% of the anode is composed of coke, its properties will have a strong influence on the anode quality. Obviously, the mixing and baking process have an impact on the final properties, but initial coke quality is important.

For coke blends, it is important to achieve a correct balance between particle shapes, density, strength and porosity. Blending cokes with very different properties can lead to poor anode quality and higher carbon consumption (Jones 1986). This also applies to coke and pitch blends. The cokefied pitch should have a similar structure as the petroleum coke (Jones 1986). The parameters affecting pitch coke structure will be discussed section 2.3.

The coke used for anode production should have:

- Low closed porosity and high density for long life in pots and low reactivity,
- Low electrical resistivity for low energy consumption,
- Low impurities to minimize metal contamination and air and CO_2 oxidation sensibility,
- Low coefficient of thermal expansion to minimize thermal shock problems.

2.3 Coal tar pitch

Coal tar pitch (CTP) is the binder used for making the baked anodes for the aluminum industry. This pitch is produced from coal tar through a distillation process. Coal tar is a by-product of the metallurgical coke production from coal. The majority of the coal tar

pitch produced around the world is obtained by a vacuum flash distillation procedure. The other manufacturing process is by heat treatment of coal tar to obtain a desired softening point. This second process leads to a lower quality pitch for anode manufacturing (Wombles and Baron 2006) since it creates a mesophase (i.e. secondary quinoline insoluble molecules) which reduces the pitch binding ability. Coal tar pitch quality is defined in the industry by the properties listed in Table 3. The range of pitch properties available in 2001 is also shown in Table 3, adapted from Perruchoud et al. (2003).

Table 3 – Typical properties for pitch in 2001, adapted from Perruchoud et al. (2003)

Properties	**Unit**	**Typical Range**
Softening point	°C	110 - 115
Insoluble in quinoline (QI)	%	6 - 16
Insoluble in toluene (TI)	%	26 - 34
Viscosity at 160 °C	mPas	1200 - 2000
Density in water	kg/dm3	1.31 - 1.33
Coking value	%	56 - 60
Distillation 0 - 270 °C	%	0.1 - 0.5
Distillation 0 - 360 °C	%	3 - 6
Ash content	%	0.1 - 0.3
S	%	0.4 - 0.6
Na	ppm	50 - 250
CI	ppm	50 - 150
Ca	ppm	20 - 100
Si	ppm	50 - 200
Fe	ppm	50 - 300
Zn	ppm	100 - 500
Pb	ppm	100 - 300

There are three major constituents in CTP, from the smallest molecules to the largest: the toluene soluble (TS), the β resin and the quinoline insoluble (QI) fractions. Figure 5 presents the three fractions of the CTP and how each constituent is measured.

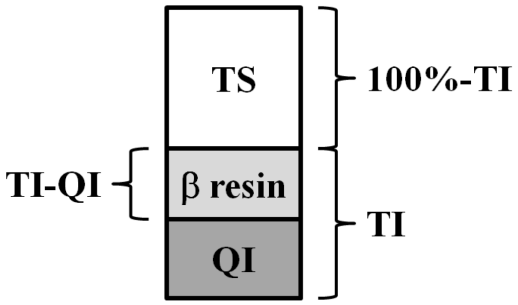


Figure 5 – Constituents of the coal tar pitch

The TS and β resin are measured from the toluene insoluble (TI) test. Toluene is a moderate solvent for pitch. The TS is the soluble fraction of the pitch in toluene. The β resin and QI fractions are contained in the toluene insolubles. Quinoline is a stronger solvent. The quinoline insoluble test is used to determine the quantity of solids and high molecular weight molecules in the pitch. The difference between TI and QI corresponds to the β resin.

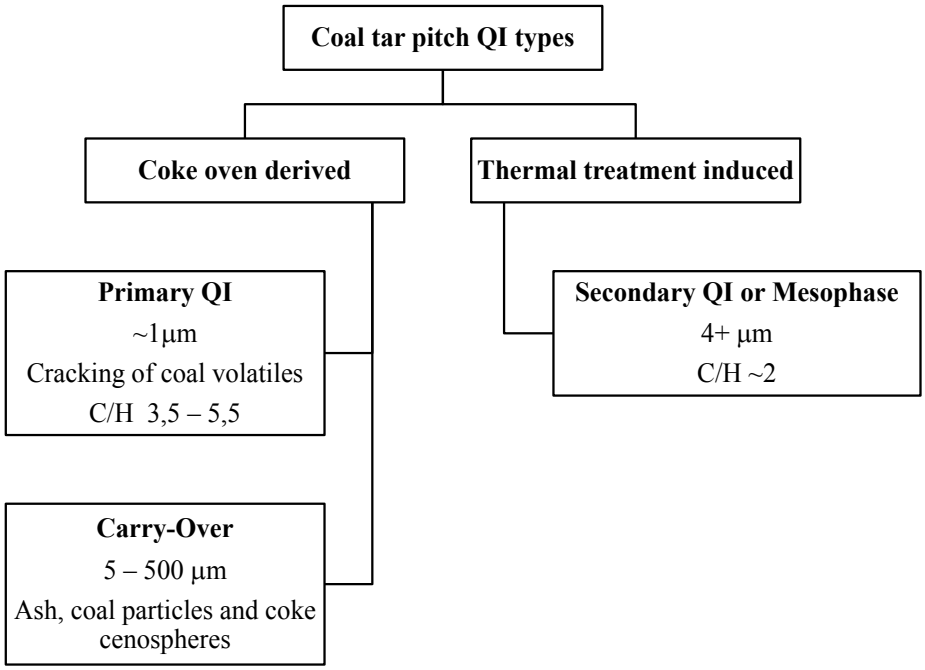


Figure 6 – Coal tar pitch QI types, adapted from Baron, et al. (2009)

The QI fraction can contain three types of molecules as seen in Figure 6, adapted from (Baron, et al. 2009). It is mainly composed of primary QI. These molecules are formed by the cracking of coal volatiles in the tunnel head of by-product coke oven. An increase in QI value decreases the wetting behaviour of the pitch. On the other hand higher QI content increases carbon product density, strength and coking value (i.e. amount of pitch coke produced during anode baking). The QI contribute to the binding effect of the pitch. Some suppliers apply a post-distillation heat treatment to the pitch. This leads to the formation of a mesophase, called secondary QI, which is undesirable for producing high quality anodes. The mesophase interferes with the binder mixing and inhibits carbon impregnation by the

pitch. Lower QI pitch with no mesophase wet coke surfaces more rapidly, thus they need less mixing time.

The β resin fraction contributes mainly to the wetting action between the coke filler and the pitch. It contains aromatics of high molecular weight molecules and as a high carbon yield, like the QI fraction, (McHenry 1992), (Golubic et al. 2010) and (Sorlie 2010).

The pitch binder becomes a graphite crystallite during baking. The pitch coke structure varies from a more disoriented and crosslinked to a more isotropic coke as the pitch QI increases. A moderate isotropic binder coke forms good physical and chemical bonds between the filler petroleum coke.

The toluene soluble (TS) fraction corresponds to the volatile content of coal tar pitch and acts as a softener. Thus it as a strong influence on the softening point of the pitch (SP). The higher the SP the fewer volatiles are contained in the binder pitch (Jones 1986). Most of the volatiles are lost by degassing during baking.

The properties of a good anode binder pitch are now discussed. First, it should have a good wetting (from the β resin) and binding (from the QI fraction) capacity during paste mixing. Secondly, the weight loss due to the degassing of volatiles during baking must be minimized. Also, the binder must provide the anodes with good mechanical properties (e.g. minimize thermal shock cracking), and high resistance to air and CO_2 oxidation.

Vacuum distilled pitch has less volatile than heat treated pitch for the same softening point. Also as the SP increases, the QI, β resin and coking value (C.V.) (e.g. the coke yield of the pitch) increase but the volatiles decrease. This situation can be advantageous during baking. With less volatile, there is less bake weight loss and also some potential for higher heat-up rate and increased baking furnace productivity (Turner 1993).

It is important to understand the correlation between the various pitch properties. Changing one property will affect the others. It is not possible to modify them one at a time and maximizing only one could lead to really bad pitch quality. This is illustrated in Figure 7 (Sorlie 2010) where the different relationships between the pitch properties are shown.

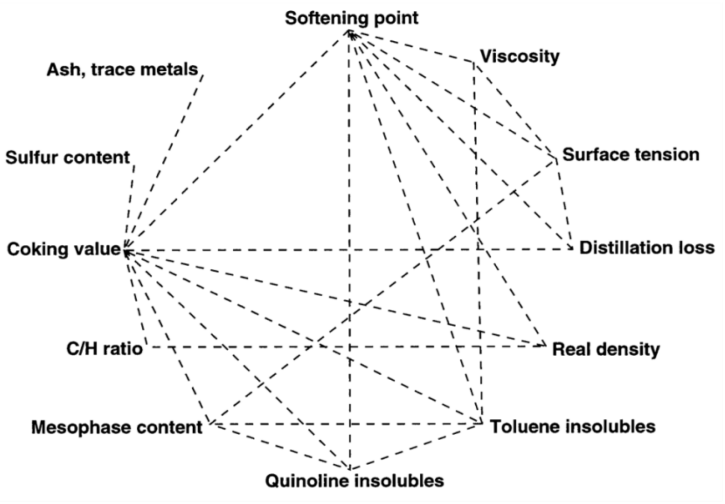


Figure 7 – Relationships between pitch properties (Sorlie 2010)

2.4 Recycle anode butts

Anode butts are the remaining part of the anode not consumed in the potroom. To avoid metal contamination from the steel stubs, the spent anodes are removed from the pots when approximately one-third is left unreacted. These butts are contaminated, thus a cleaning step follows the separation of the anode from the stub to remove anode cover material and frozen bath from the butt surfaces. After cleaning, the butts are crushed, screened to the desired size distribution and stored in silos for use in the production of fresh anodes. This reduces the amount of waste material. Butts constitute approximately 15-30% of the green anode formulation (Fischer and Perruchoud 1991).

The cleanliness of the butts is of great importance. It influences the anode mechanical strength, air permeability and CO_2 reactivity. Good quality butts are characterized by their hardness and low sodium (Na) content (Hulse 2000). They have properties similar to the baked anode. Hard butts increase anode apparent density, decrease air permeability and increase flexural strength. Finally, the presence of sodium in the form of bath particles increases the CO_2 reactivity (Fischer and Perruchoud 1991).

The sodium in the butts comes mainly from the contamination of the material by residual bath particles and by impregnation in the anode. The bath is composed of cryolite (Na_3AlF_6) and fluoride salts containing Na and Ca. The particles are therefore highly

concentrated in sodium which acts as a catalyst for the air and CO_2 oxidation reactions (Engvoll et al. 2001). Well cleaned butts contain a minimum amount of sodium. This is important to minimize its impact on anode quality. Sometime, during normal cell operation (e.g. controlling anode effect), the anodes are dipped in the metal. During that time, the impurities and the bath including sodium will soak up in the anode at a fast rate. Anode dipping can also occur when pots become unstable due to higher than normal metal movement and also sometimes due to operational errors when the operators are changing the anodes. It should be minimized as much as possible because even well cleaned dipped anode contains a high amount of Na.

2.5 Anode formulation (aggregate sizing)

The coke and butts particles included in the anode are called the dry aggregate. They are the filler of the anode, thus influencing its quality through porosity, packing and sizing (Hulse 2000). Coke and butts particles are prepared through screening and crushing. The coke is classified in three different fractions: coarse, medium and fines. The fines are produced in a ball mill. The dust collected throughout the paste plant is also directed in the fine fraction. The butts consist mainly of coarse material (Fischer et al. 1995). In the dry aggregate formulation, it is desired to use butt particles for the coarser size fraction and coke particles for the medium and fine fractions. Butts usually have a lower porosity than coke and keeping the particles bigger decreases the risk of sodium contamination in the filler matrix by high sodium butt fines. Also, coarser butts have a lower surface area compared to fine butts, thus limiting the effect of Na on reactivity. It is also important to maintain the size distribution of each fraction as constant as possible. The proportion and fineness of the dust needs to be in good control since it as a strong impact on pitch demand (Mannweiler and Keller 1994). The dust fineness is characterized using the Blaine number, which is a measure of the particle surface area. It is used for particles too small to be classified by sieve analysis. The most commonly used method is to measure the Blaine through the pressure-drop of a packed-bed of the sample. The drawback is that it measures only an average value of the size distribution. It can also be measured by laser diffraction which gives the distribution of the particle size in the dust. Green apparent density increases with the dust content but it increases the pitch demand. When coke is grinded to a

smaller particle size, its porosity decreases and its apparent density increases. There is an optimum balance between fines and pitch content as a high pitch ratio or high fines ratio anode can be difficult to process (Hulse 2000).

2.6 Paste pre-heating, mixing and forming

The dry aggregate blend is formulated using weight belts and is discharged in a pre-heating equipment. Green scrap is also added to the dry aggregate at this step. Maintaining an appropriate pre-heating temperature (ranging between 150 and 200°C (Hulse 2000)) is important when adding pitch to the dry aggregate blend. Pitch at a temperature ranging between 170 and 230°C (Hulse 2000) is incorporated in the dry aggregate mix when it enters the paste mixer. The goal of mixing is to combine the aggregate and the binder and evenly distribute the later component in the paste. A large temperature difference between the pitch and the dry aggregate when these are put in contact in the paste mixer can cause partial solidification of the pitch on the coke particles. This can prevent proper pitch penetration in the filler matrix. Mixing temperature is usually 50°C to 60°C (155°C to 180°C) above the pitch softening point. Anode quality generally increases with increased coke and pitch temperature up to the degassing temperature of pitch volatiles. The paste viscosity decreases with an increase in the temperature and this will improve the mixing, spreading and penetration of the binder matrix in the paste (Hulse 2000).

The pitch ratio in the paste is also of great importance. Under-pitched anodes will have insufficient mechanical properties leading to anode failure in the pots and high electrical resistivity due to a poor binding behaviour. Over-pitched anodes lead to slump formation (i.e. problems when forming the anodes), high weight loss, shrinkage and cracks formation during baking due to greater volatiles degassing, to packing material sticking also while baking and finally stub hole deformations (Hulse 2000, Mannweiler and Keller 1994). Pitch demand is a function of the fines and filler particle properties but also mixing temperature and duration. There is an optimum between mixing duration and temperature, dry aggregate pre-heating temperature and pitch ratio.

Anode forming is performed either by pressing or vibro-compaction. Pressed anodes are insensitive to paste viscosity as opposed to vibrated anode. The quality of pressed anodes

depends largely on raw material properties and recipe. Vibrated anode quality depends also on raw material quality but is more sensitive to anode forming process parameters (e.g. paste temperature during vibro-compaction) (Hulse 2000). If the temperature is too high, the paste viscosity will be too low and the anode could collapse when taken out of the mold and a low temperature causing high viscosity will lead to improper compaction.

2.7 Anode baking

Anode baking is done in an open ring baking furnace. Details of operation are explained in (Fischer et al. 1995). A section of the furnace is made of several pits (generally 6 or 7) where the anodes are staked vertically (e.g. 6 anodes large by 3 anodes high). The space between each pit (i.e. flue wall) is a cavity where natural gas and pitch volatiles are burned in order to supply heat to the anodes.

Anode baking aims essentially at developing the mechanical properties of the anodes. The heat-up temperature gradient, the final temperature and soaking time are the most important baking parameters (Mannweiler and Keller 1994). Also a minimal temperature gradient between the different positions within the furnace needs to be maintained (Fischer et al. 1993).

The heat-up temperature gradient controls the pitch volatiles degassing rate. Pitch devolatilization occurs in the 200-600°C range. The pitch volatiles are drawn and burned into the walls separating the pits of the baking furnace. This is a major energy input and it needs to be kept under control by the baking furnace process control system. The impact of a high temperature gradient is the emergence of internal or external cracks due to the internal pressure build-up. These cracks can compromise the mechanical properties of the anodes and increase in the risk of thermal shock when the anodes are set in the reduction cells. They also increase the electrical resistance of the anodes which has a negative impact on the energy consumption of the anode in the pots. A too high heat-up gradient can also increase the anode porosity which leads to higher air and CO_2 reactivity. The maximum temperature gradient is in the range 10-14°C/h (Fischer et al. 1993).

Final anode temperature is also an important parameter (i.e. the soaking temperature). It has a great influence on the crystallinity, measured by the L_C (i.e. measure of the crystallite length) of the anode. Thermal conductivity increases with the final baking temperature and L_C. The CO_2 reactivity residue (the amount of anode not reacted) increases with the final temperature. Air reactivity residue also follows this trend with the exception of anode with high sulfur coke. Desulfurization of the anode can occur at high temperature. This phenomenon tends to increase the porosity and increase the air reactivity of the anodes at high final baking temperature (Fischer et al. 1993). Temperatures above 1200°C do not improve anode quality (Jones 1986).

The temperature distribution within a pit of a baking furnace depends on its design, its maintenance and deterioration stage. It is important to maintain an even temperature distribution to obtain uniform anode quality throughout the entire anode population. A large range of anode quality will lead to operational problems in the potroom. A longer soaking time (i.e. time left at final baking temperature) can contribute to a better temperature gradient within the furnace and will also lead to better anode properties (Fischer et al. 1993). Soaking time should be at least 10 hours in the range of 1150 ± 50°C (Jones 1986).

2.8 Anode quality estimation

Some authors (Keller and Fischer 1982; Fischer et al. 1991) have developed models for the prediction of anode net carbon consumption (i.e. an anode performance indicator). These models are based on the baked anode properties measured from core samples as well as some pot design parameters and performance indicators. The models developed in this project could be used as input in these carbon consumption models to obtain estimates on an anode basis.

Other authors have investigated the relationships between some raw material properties and process parameters and baked anode properties, see Fischer et al. (1995) for an overview of those analyses. However, as opposed to the multivariate models developed in this thesis, most of past studies consider the effect of only a few variables using a one variable at a time approach. For example, Hume et al. (1993a) measured the influence of sodium contamination and sulfur content on the anodes CO_2 reactivity. Fischer and Perruchoud

(1991) studied the effect of butts properties (i.e. hardness and sodium contamination) on the anode apparent density, air permeability, flexural strength and air and CO_2 reactivity. Finally, Fischer et al. (1993) studied the effect of baking furnace parameters (i.e. heat-up rate, final baking temperature and soaking time) on the anode electrical resistivity, flexural strength, thermal conductivity, air and CO_2 reactivity and finally desulfurization (as a function of initial sulfur content).

The major difference between these studies and the project described in the thesis is that this work incorporates all the aspect of the manufacturing process form the raw materials to the baking furnace into a statistical model in order to predict baked anode properties.

Chapter 3 Multivariate statistical methods

This chapter presents the relevant statistical background information useful for understanding of the work presented in this thesis. The basic multivariate analysis methods are presented. These methods were developed in the field of chemometrics. Svante Wold defines this field of science as "How to get chemically relevant information out of measured chemical data, how to represent and display this information, and how to get such information into data" (Wold 1995). The goal of these methods is to extract the most useful information from the data. It has been extended to chemical process analysis and monitoring as well (Wise and Gallagher 1996) and (MacGregor and Kourti 1995). Two of the most used methods, Principal Component Analysis (PCA) and Projection to Latent Structures (PLS), also referred to as Partial Least Squares, are presented in the following sections together with a discussion on data scaling and on the selection of the number of latent variables to include in the models.

In this thesis, the following notation is used. Vectors are shown using bold lowercase characters (**lowercase**), matrices are represented by bold capital characters (**CAPITAL**) and the transpose operator is illustrated using uppercase capital T (e.g. $\mathbf{X}^T$ or $\mathbf{t}^T$).

3.1 Principal Component Analysis (PCA)

Principal Component Analysis is the basic multivariate data analysis approach. It is used to model and investigate multivariate datasets. Detailed tutorial and examples can be found in (Wold et al. 1987) and (Kourti 2005). Assume **X**, a data table made of *I* observations (or measurements) obtained from *J* different variables as illustrated in Figure 8. Most industrial datasets are very large, noisy, and the variables are typically highly collinear (e.g. **X** is not full rank). However, measuring hundreds to thousands of variables on a given process does not necessarily mean that a hundred independent events occurred on this process. In fact, process operation is usually driven by a much lower number of underlying independent events called lurking or latent variables (LV) (the **p**'s in Figure 8). These LVs cause the large number of process variables to vary together in certain directions (i.e. in a correlated

fashion). PCA is one of the basic methods for extracting these few latent variables capturing most of variance in a dataset. The projection of the dataset onto the lower dimensional space of A dimensions spanned by the latent variables can then be used to visualise and interpret the relationships between the variables and between the observations.

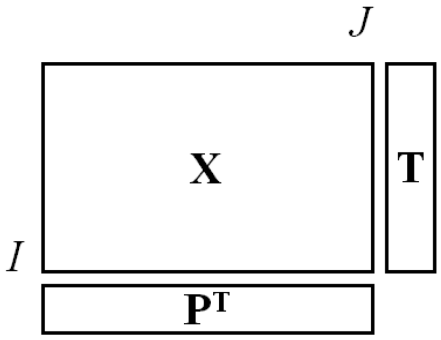


Figure 8 – Schematic representation of PCA

The first principal component is the linear combination of the J columns (variables) of $\mathbf{X}$, defined by the orthonormal vector $\mathbf{p_1}$, explaining the greatest amount of variance in the dataset. This is mathematically formulated as an eigenvector-eigenvalue problem with the following objective function:

$$\max_{\mathbf{p}_1} \left\{ \mathbf{p}_1^T \mathbf{X}^T \mathbf{X} \mathbf{p}_1 \right\} \text{ subject to } \mathbf{p}_1^T \mathbf{p}_1 = 1 \tag{2}$$

where the term with brackets represents the variance of the first latent variable $\mathbf{t_1}$ defined as the projection of $\mathbf{X}$ in the direction of $\mathbf{p_1}$:

$$\mathbf{t}_1 = \mathbf{X}\mathbf{p}_1 \tag{3}$$

This latent variable explains the most variance in $\mathbf{X}$ and it is removed from the dataset leaving the residual matrix $\mathbf{E_1}$:

$$\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1 \mathbf{p}_1^T \tag{4}$$

If the first component is not sufficient for explaining the variations in $\mathbf{X}$, a second PCA component can be added to the model. It corresponds to the linear combination of the J

variables explaining the greatest amount of variance not captured by the first component, (i.e. left in the residual matrix $\mathbf{E}_1$). The second component is the solution to the following eigen problem:

$$\max_{\mathbf{p}_2} \left\{ \mathbf{p}_2^T \mathbf{X}^T \mathbf{X} \mathbf{p}_2 \right\} \text{ subject to } \mathbf{p}_2^T \mathbf{p}_2 = 1 \text{ and } \mathbf{p}_1^T \mathbf{p}_2 = 0 \tag{5}$$

The additional constraint for this second component ensures that the latent variables are orthogonal to each other (e.g. they are independent). Additional components can be added sequentially to the PCA model using expression (5) until the desired number of latent variables (A) is computed. The maximum number of LVs is J, but for industrial data A is usually smaller than J ($A << J$) due to the highly collinear structure of the data. The final model has the following structure:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{6}$$

where the score and loading vectors are collected in the matrices $\mathbf{T}$ (IxA) and $\mathbf{P}$ (JxA) and the residuals are stored in matrix $\mathbf{E}$ (IxJ). In summary, PCA performs the eigenvector decomposition of $\mathbf{X}$. The $\mathbf{p}$ vectors are the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and the $\mathbf{t}$ vectors are the eigenvectors of $\mathbf{X}\mathbf{X}^T$.

For the numerical computation of the $\mathbf{p}$ and $\mathbf{t}$ vectors, the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm is used. This method is scaling dependent and this issue will be discussed in section 3.3. Figure 9 shows the algorithm details of NIPALS for PCA (Geladi and Kowalski 1986).

1. Set **t** to any column of **X**.
2. Start convergence loop.
   2.1. $\mathbf{p} = \mathbf{X}^T\mathbf{t}/(\mathbf{t}^T\mathbf{t})$
   2.2. $\mathbf{p} = \mathbf{p}/(\mathbf{p}^T\mathbf{p})^{½}$
   2.3. $\mathbf{t} = \mathbf{X}\mathbf{p}$
   2.4. Check for convergence of **t** and **p**. Continue to step 3 if converged.
3. $\mathbf{E} = \mathbf{X} - \mathbf{t}\mathbf{p}^T$
4. Store **p** and **t** as new columns in **P** and **T**.
5. Restart at step 1, replacing **X** by **E**.

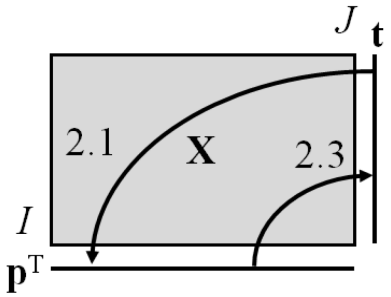


Figure 9 – NIPALS algorithm for PCA

When *J* is very large, which is normally the case with industrial data, this method is advantageous compared to eigenvector decomposition since it is often not necessary to compute all latent variables (in this case, $A<<J$).

3.2 Projection to Latent Structures (PLS)

Projection to Latent Structure is a multivariate regression method. Consider a second dataset **Y** of *H* variables and *I* observations (e.g. response variables such as product quality attributes) as seen in Figure 10. The PLS method is used to explore the relationships existing within and in between both datasets, **X** and **Y**. It can be seen as an extension of PCA, but for two sets of data.

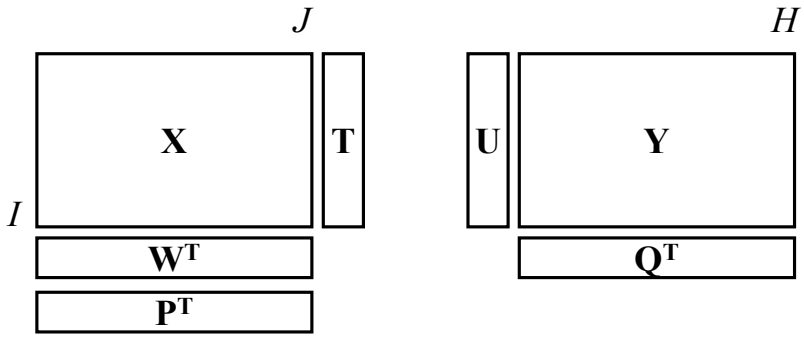


Figure 10 – Matrices of PLS

The basic assumption behind PLS is that variations in **X** and **Y** are linked by a common set of *A* latent variables **T** and **U** respectively (*I*×*A*). These latent variables are the directions in the **X** and **Y** space that maximize the covariance between these two datasets. Additional details and tutorials can be found in (Geladi and Kowalski 1986), (Höskuldsson 1988), (Burnham et al. 1996), (Burnham et al. 1999), (Wold et al. 2001), (Martens 2001), and (Kourti 2005).

Mathematically, the latent variables are computed as a set of linear combinations of the **X** descriptor variables defined by the weight vectors $\mathbf{w}_i$, i = 1,... , *A*, to maximize the squared covariance between **X** and **Y**. The solution to this problem is again formulated as an eigen problem with the following objective function:

$$\max_{\mathbf{w}_i} \left\{ \mathbf{w}_i \mathbf{X}^{\mathrm{T}} \mathbf{Y} \mathbf{Y}^{\mathrm{T}} \mathbf{X} \mathbf{w}_i^{\mathrm{T}} \right\} \text{ subject to } \mathbf{w}_i^{\mathrm{T}} \mathbf{w}_i = 1 \text{ and to } \mathbf{w}_i^{\mathrm{T}} \mathbf{w}_j = 0 \text{ for } i \neq j \tag{7}$$

As for PCA, the set of constraints ensure that the weight vectors $\mathbf{w_i}$ are orthonormal and that latent variables are orthogonal to each other (e.g. they are independent). The PLS model structure is described below, and is also shown schematically in Figure 10.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \tag{8}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\mathbf{T}} + \mathbf{F} \tag{9}$$

$$\begin{aligned} \mathbf{T} &= \mathbf{X}\mathbf{W}^{*} \\ \mathbf{W}^{*} &= \mathbf{W}\left(\mathbf{P}^{\mathbf{T}}\mathbf{W}\right)^{-1} \end{aligned} \tag{10}$$

T (*I*×*A*) is the common latent variable space (*A* dimensions) that models the relationship between **X** and **Y**. They are the combination of the **X** variables that are the most highly correlated with the data in **Y**. The weight of each variable in the common latent space is the matrix **W*** (*J*×*A*). The **P** (*J*×*A*) and **Q** (*H*×*A*) matrices contain the loading vectors relating the common latent variable space in each **X** and **Y** space respectively. **E** (*I*×*J*) and **F** (*I*×*H*) are the PLS model residuals. It was shown by (Höskuldsson 1988) that the vectors **w**, **q**, **t** and **u** are the eigenvectors of the following matrices $\mathbf{X}^{\mathbf{T}}\mathbf{Y}\mathbf{Y}^{\mathbf{T}}\mathbf{X}$, $\mathbf{Y}^{\mathbf{T}}\mathbf{X}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$, $\mathbf{X}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\mathbf{Y}^{\mathbf{T}}$ and $\mathbf{Y}\mathbf{Y}^{\mathbf{T}}\mathbf{X}\mathbf{X}^{\mathbf{T}}$ respectively.

The NIPALS algorithm was adapted to PLS by (Geladi and Kowalski 1986) and (Höskuldsson 1988) to compute the PLS latent variables sequentially. With this method, only the desired number of LV's are calculated. The algorithm is shown in Figure 11. The PLS vectors are also scaling dependent. This will be discussed with the selection of the number of latent variables in sections 3.3 and 3.4, respectively.

1. Set **u** to any column of **Y**.
2. Start convergence loop.
   2.1. $\mathbf{w} = \mathbf{X}^T\mathbf{u}/(\mathbf{u}^T\mathbf{u})$
   2.2. $\mathbf{w} = \mathbf{w}/(\mathbf{w}^T\mathbf{w})^{½}$
   2.3. $\mathbf{t} = \mathbf{X}\mathbf{w}$
   2.4. $\mathbf{q} = \mathbf{Y}^T\mathbf{t}/(\mathbf{t}^T\mathbf{t})$
   2.5. $\mathbf{u} = \mathbf{Y}\mathbf{q}/(\mathbf{q}^T\mathbf{q})$
   2.4. Check for convergence of **t** or **u**. Continue to step 3 if converged.
3. $\mathbf{p} = \mathbf{X}^T\mathbf{t}/(\mathbf{t}^T\mathbf{t})$
4. $\mathbf{E} = \mathbf{X} - \mathbf{t}\mathbf{p}^T$ and $\mathbf{F} = \mathbf{Y} - \mathbf{t}\mathbf{q}^T$
5. Store **w**, **p, t** and **u** as new columns in **W**, **P**, **T** and **U**.
6. Restart at step 1, replacing **X** by **E** and **Y** by **F**.

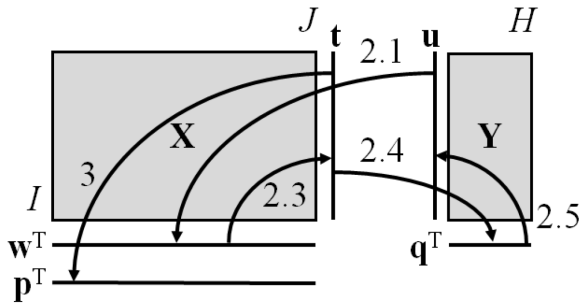


Figure 11 – NIPALS algorithm for PLS

3.3 Data scaling

Both PCA and PLS methods are sensitive to how the **X** and **Y** data matrices are scaled. When no prior knowledge is available on the relative importance of the variables, the common practice is to scale them to unit variance after applying mean-centering. This scaling procedure is applied to each variable (i.e. columns) of the **X** and **Y** data matrices. Consider a column vector ($\mathbf{x}_j$) of the **X** data matrix and its mean value ($\mathbf{x}_{j,\text{mean}}$) and standard deviation ($\mathbf{x}_{j,\text{std}}$). The scaled values ($\mathbf{x}_j$*) are obtained using the following equation (element by element division is assumed):

$$\mathbf{x}_j^* = \frac{\left(\mathbf{x}_j - \mathbf{x}_{j,\text{mean}}\right)}{\mathbf{x}_{j,\text{std}}} \tag{11}$$

This method is also called auto-scaling. Mean-centering allow the computation of the variations of the variables around there mean and scaling to unit variance gives equal importance to all the variable in the models as not all of them are measured in the same engineering units (Geladi and Kowalski 1986).

3.4 Number of latent variables (*A*)

Industrial data are typically highly collinear and noisy. Collinearity implies that a limited number of latent variables are needed to capture and explain most of the variations in a dataset (**X** and/or **Y**). The corruption of the data by noise means that carefulness must be used to model only the systematic variation (i.e. structured variations) and to ignore the noise. When the correct number of latent variables (*A*) is selected, the important information is stored in the loadings and weight matrices (**P**, **Q** and $\mathbf{W}^*$) and the irrelevant variations are left in the residuals (**E** and **F**). The most commonly used method for selecting the number of latent variable is cross-validation (Wold 1978), but other methods also exist to determine the model order (Nomikos and MacGregor 1995) and (Valle et al. 1999).

The cross-validation (CV) method keeps adding latent variables to the model until the latest component does not significantly improve predictions of X (PCA) or Y (PLS). For the cross-validation procedure, the *I* observations in **X** and/or **Y** are divided into *g* sub-groups of *n* observations (*I*=*gn*). Each sub-group is removed from the data once and only once and an *A* latent variable PCA or PLS model is built on the remaining *g*-1 sub-groups. Predictions are computed for the group left out of the analysis and the prediction error sum of squares (PRESS) is computed for this sub-group. PRESS(*a*) is the sum of the PRESS values for all g sub-groups for a model with *a* latent variables ($a = 1,2,\ldots A$). The model predictive ability is than evaluated with the predictive multiple correlation coefficient (Q^2_{CV}):

$$Q^2_{CV}(a) = 1 - \frac{\mathrm{PRESS}(a)}{\mathrm{SS}_r(a-1)} \tag{12}$$

where

$$\mathrm{PRESS(a)} = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\mathrm{y}_{i,j} - \hat{\mathrm{y}}_{i,j}(a)\right)^2 \tag{13}$$

and

$$\mathrm{SS_r(a\text{-}1)} = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\mathrm{y}_{i,j} - \hat{\mathrm{y}}_{i,j}(\mathrm{a\text{-}1})\right)^2 \tag{14}$$

In the above equations, I is the number of observations and J is the number of variables and a is the number of model components (a = 1,2,…A), and $\mathrm{SS}_{\mathrm{r}(a\text{-}1)}$ is the residual sum of squares of the model with a-1 latent variables. The number of principal components or latent variables (A) to keep in the model is selected to be the one achieving the highest predictive ability.

Selecting a too small number of latent variables leaves some structured information in the residuals. Selecting too many latent variables leads to overfitting and modeling of the random noise in the data.

Alternatively, one could use a separate validation dataset for computing predictive ability. While adding one LV at a time, it is possible to compute the PRESS on the validation set until the predictive ability starts to degrade due to overfitting. This approach with external data is the better way to validate a model, but a high number of observations are needed in order to split a dataset into a training set and a validation set.

3.5 Model interpretation tools

Aside from the model structure of PCA and PLS, which are powerful methods for process modeling, a number of tools can be use to help interpret the models and learn from the data. First, the score plot and loadings plot are used to interpret the relationship between the observations and the variables, respectively. A combination of two or three latent variables can be simultaneously visualized through these tools using 2D or 3D scatter plots. The use of these score plots will be illustrated later in the results section.

The DModX is the distance of an observation to its projection on the latent variable space (e.g. the residual). It is useful for detecting outliers because it highlights observations with a different correlation structure than the model (i.e. outliers in the space orthogonal to the LV space).

$$\mathrm{DModX}_i(a) = \sqrt{\frac{\sum_{j=1}^{J} e_{ij}^2(a)}{J - A}} \tag{15}$$

$\mathrm{DModX_i}(a)$ is computed from equation (15), where $e_{ij}(a)$ is the residual of observation i and variable j obtained with a model built using a latent variables.

$$e_{ij}(a) = \left(\mathrm{y}_{i,j} - \hat{\mathrm{y}}_{i,j}(a)\right) \tag{16}$$

The Hotelling's T^2 is the Mahalanobis distance of an observation to the center of the LV space. It can also be used for detecting outliers in the LV space.

$$\mathrm{T}_i^2 = \sum_{a=1}^{A} \frac{\boldsymbol{t}_{ia}^2}{s_{ta}^2} \tag{17}$$

where,

$$T_i^2 * \frac{J\left(J - A\right)}{A\left(J^2 - 1\right)} \rightarrow F\left(A, J - A\right) \tag{18}$$

approximately follows an F distribution with A and J-A degrees of freedom. In equation (17), T^2_i is the Hotelling's distance for an observation, $\mathbf{t}_{ia}$ is the score of an observation for the a latent variable and s^2_{ta} is the variance of $\mathbf{t}_a$.

An additional tool which can help identify important variables in a PLS model is the variable importance in projection (VIP) which is an indication of the importance of a variable in predicting the **Y** variables:

$$\mathrm{VIP}_{j,A} = \sqrt{\mathrm{J}\sum_{a=1}^{A} \mathrm{w}_{j,a}^2 \left(\mathrm{SSY}(a) / \mathrm{SSY}(tot)\right)} \tag{19}$$

where $w_{j,a}$ 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. Usually, variables with a VIP greater than 1 are considered important for the model (Eriksson et al. 2001).

Finally a useful interpretation tool is the contribution plot. It essentially consists of the difference in the values of a particular variable between two time points (or averaged over some time windows) weighted by the importance of that variable in the model given by the PLS model weights (**w***). It indicates which combination of variables contributes the most to a deviation in the score space (**T**) of a latent variable model. It is not a cause and effect relationship, but it is a good indicator of possible root causes. The calculation of the contributions is explained in Kourti (2005) and Weterhuis et al. (2000). The contribution of variable j, to the shift between two observations (k_1 and k_2) is computed using the expression below.

$$C_j = \sum_{a=1}^{A} \frac{\left[\left(x_{j,k_1} - x_{j,k_2}\right) \times w^*_{j,a}\right]^2}{s^2_{t_a}} \tag{20}$$

where $x_{j,k1}$ and $x_{j,k2}$ are the values of the j^{th} variable at time k_1 and k_2; $w_{j,a}$ is the weight associated to the j^{th} variable of the a^{th} latent variable and s^2_{ta} is the variance of the a^{th} score. Dividing by the score variance gives an equal importance of deviations in each LV. For contribution from a group of observations to another group, the difference in the mean value of the observation in each group for each variable is used.

Chapter 4 Data collection

This chapter describes how the data was collected and pre-treated prior to building the PLS models. The data used for this project were collected at the Alcoa Deschambault smelter from February 23, 2009 to January 2, 2011. All data presented in this thesis are auto-scaled. Data were collected from raw material laboratory analyses and suppliers certificate of analysis (COA), green mill process and baking furnace data historians, and core sample laboratory analysis. Most raw material data were stored as weekly averages, but some data were only available on a per shipment basis. The process parameters from the paste plant are available on a one minute basis. Baking furnace data are available in two different formats. The first consists of the data collected from the baking furnace at the end of each fire move and used by the plant personnel to monitor the baking process. These data were available for all the models computed in this work. The second set of baking furnace data is available for a subset of the original dataset with anodes manufactured from February 23, 2009 to May 21, 2010. It consists of the entire temperature profile during most of the baking cycle. Finally, core sample results were obtained from Deschambault's quality control laboratory database.

4.1 Coring strategy

There is a particular coring strategy used at Deschambault. Two anodes are cored in each section of the baking furnace. The position of the anodes in the oven is presented in Figure 12. The core samples are drilled from the top at the same position in each anode as shown in Figure 13. The core sample height is smaller than the anode height (anode not cored all the way to the bottom).

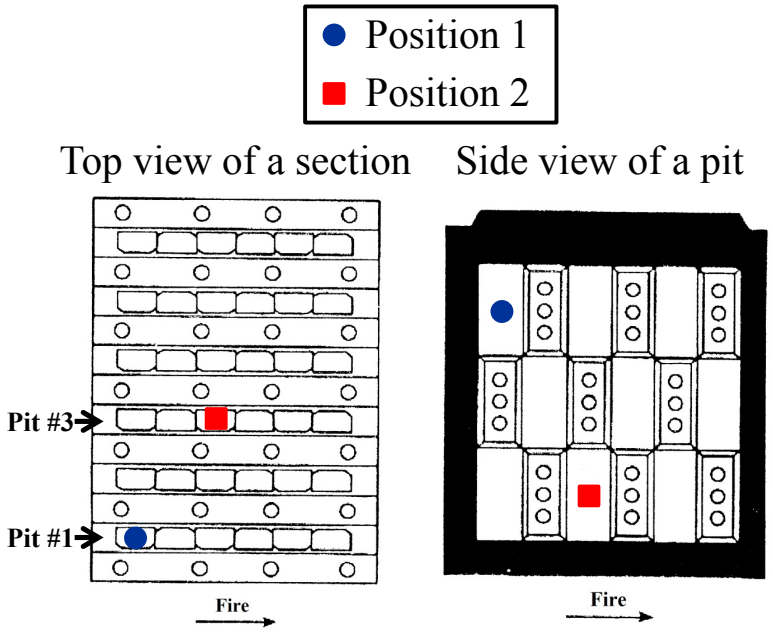


Figure 12 – Coring positions in the baking furnace (Courtesy of Alcoa)

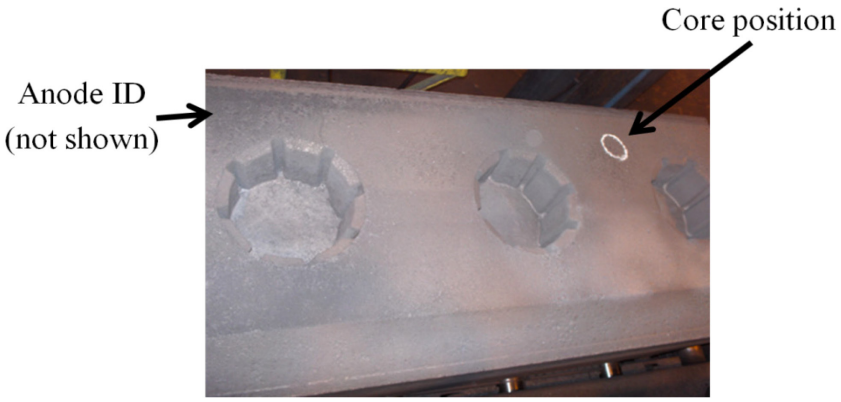


Figure 13 – Core sample position in the anode (Courtesy of Alcoa)

Seventeen anodes are chosen for coring each production week and are sent to the lab. Different tests are performed on different sets of anode based on a standardized test plan specific to the smelter. Physical properties and reactivity properties are measured using different core samples (i.e. different anodes) since the core samples are too small for performing all the tests. Therefore, separate PLS models are built for physical and reactivity properties. These models will be presented in Chapter 5.

4.2 Data synchronization

To build the models each measured variable need to be synchronized to a cored anode. This section describes how the traceability and synchronization of an anode was established at Deschambault's carbon plant. Each anode receives a unique sequential number. This number is applied by an automated mechanical punch when the green anode is formed in the vibro-compactors. This number can be tracked through the baking step. In the paste plant database, the time and date of the anode fabrication is stored along its identification number. This, together with the core sample ID given to each core by the lab, is the basis for the synchronization of the dataset. Figure 14 presents the synchronization scheme used in this project.

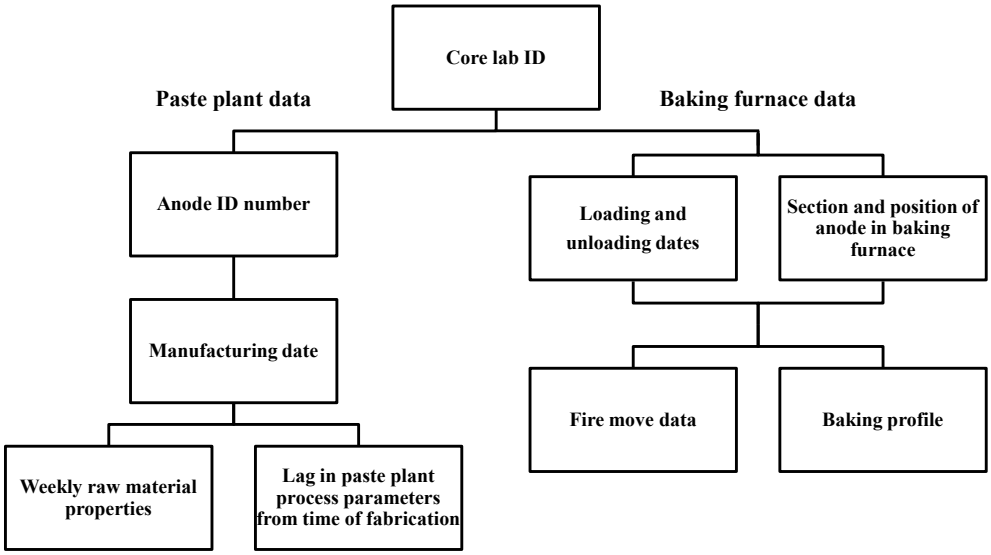


Figure 14 – Synchronization scheme

The synchronization starts with the core lab ID. With this number it is possible to obtain the anode ID number and the information about the baking furnace section, position, loading and unloading dates. From the anode ID number, the manufacturing date is retrieved and the synchronization of paste plant process and raw material data is possible. From the baking information, it is possible to extract fire move data and baking profiles.

All the data were extracted from Deschambault using their dedicated historian software and transferred in Excel spreadsheets. For the paste plant data and baking furnace profile, this represents enormous amount of data. All the computation for the synchronization of the

data was performed using the author's custom codes developed using Matlab® R2009B (The Mathworks Inc., Natick, MA, USA). For the statistical analysis of the data and model computation, SIMCA-P+ Version 11.0.0.0 (Umetrics Inc., Kinnelon, NJ, USA) was used.

Ultimately, the synchronization code allows the alignment of all the data into a matrix containing all the raw material properties (**Z**) and the process operating parameters (**X**) and a second matrix containing the desired anode quality variables (**Y**). This synchronization needs to be done for the physical property model and for the reactivity property model separately since these were measured from different anodes. The observations (anodes) used in the physical property model containing additional baking temperature profiles are stored in a third dataset and used in a separate PLS model. Figure 15 shows the different dataset for the three models. Details of the synchronization for every step of the manufacturing process are described in sections 4.2.1 to 4.2.3.

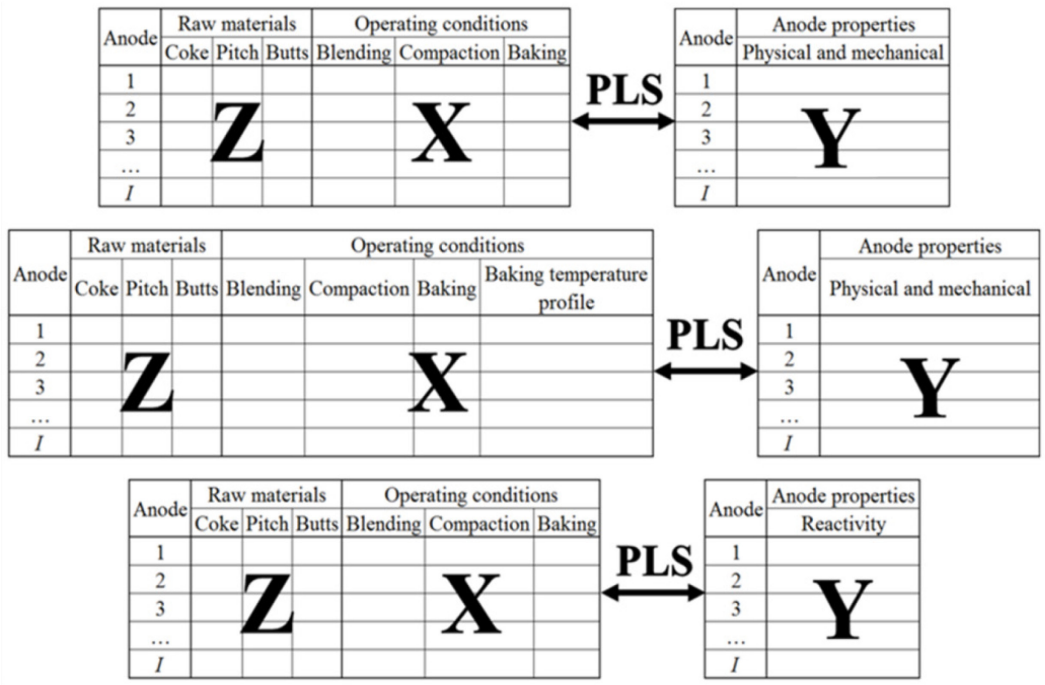


Figure 15 – Data matrices for: a) Physical property model, b) Physical property with baking profile information model and c) Reactivity property model

4.2.1 Raw material synchronization

The raw material synchronization is made on a weekly basis. During this thesis, Deschambault received coke from six different suppliers and pitch from two different 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 4. In this table, coke suppliers are identified by letters (A-E) and pitch suppliers with numbers (1-2).

Table 4 – Raw material blend

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

For pitch quality, a weekly average of all the material received was computed using the supplier's certificate of analysis (COA). The coke properties are measured at Deschambault's laboratory from sample gathered from unloading railcars. The butt samples were taken from the size distribution analysis sample collected every day at the paste plant. A composite of all butt samples is analyzed each week. The raw material properties are listed in Table 5.

Table 5 – Raw material property variables

Variable ID	Properties	Units
Coke real dens	Coke real density	g/cm^3
Coke Na	Coke sodium	ppm
Coke Ca	Coke calicum	ppm
Coke S	Coke sulphur	%
Coke V	Coke vanadium	ppm
Coke 28/48 app dens	Coke tapped bulk density -28 to +48 mesh fractions	g/cm^3
Coke Fe	Coke iron	ppm
Coke Si	Coke silicon	ppm
Coke Ni	Coke nickel	ppm
Coke ash	Coke ash	ppm
Pitch SP	Pitch softening point	°C
Pitch TS	Pitch toluene soluble fraction	%
Pitch Beta	Pitch b fraction	%
Pitch QI	Pitch quinoline insoluble fraction	%
Pitch B/QI	Pitch b/QI fractions ratio	---
Pitch CV	Pitch coking value	%
Pitch ash	Pitch ash	%
Pitch S	Pitch sulphur	%
Pitch dist	Pitch distillate at 355°C	%
Butts Al	Butts aluminium	ppm
Butts Ca	Butts calcium	ppm
Butts %F	Butts fluorine	%
Butts ash	Butts ash	%
Butts Fe	Butts iron	ppm
Butts Ni	Butts nickel	ppm
Butts Si	Butts silicon	ppm
Butts Na	Butts sodium	ppm
Butts S	Butts sulphur	%
Butts V	Butts vanadium	ppm
Butts Na/Ca	Butts sodium/calcium ratio	---

4.2.2 Paste plant data synchronization

The lag structure for the paste plant synchronization is presented in Figure 16. The values were determined after discussion with plant process engineers and operators based on their experience.

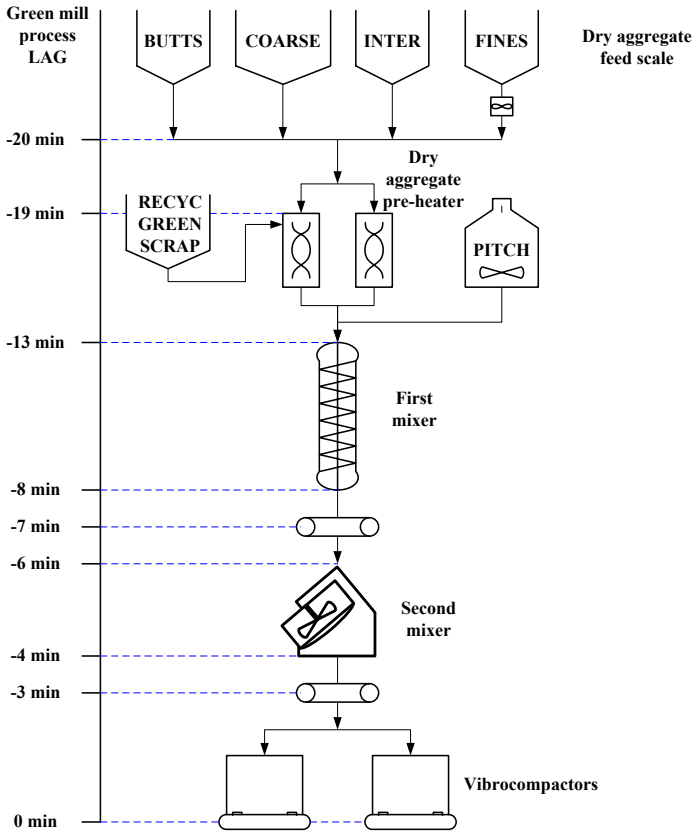


Figure 16 – Paste plant lag structure

Paste plant process data are available on a minute basis from the historians. The values for each variable are synchronized to the closest fabrication date (precise to the minutes) minus the lag for that variable. For the equipments having a residence time longer than one minute (e.g. pre-heater and mixers), the mean of all the data between the lag values were computed as follows:

$$\overline{\mathbf{x}}_{i,j} = \frac{\sum_{l=1}^{L} \mathbf{x}_{i,j,l}}{\text{nbr of lags}} \tag{21}$$

where L is the lag from the manufacturing time in minutes and $\mathbf{x}_{i,j,l}$ is the value of variable j at observation i and lag l. For example, the mean value of the first mixer for the i^{th} observation is calculated with lag from -8 minutes to -13 minutes because the residence time is 5 minutes.

The process variables used in the analysis are listed in Table 6.

Table 6 – Paste plant variables

Variable ID	Properties	Units
Coarse (tph)	Coke coarse mass feed rate	ton/h
Fines (tph)	Coke fines mass feed rate	ton/h
Inter. (tph)	Coke intermediate mass feed rate	ton/h
Butts (tph)	Butts mass feed rate	ton/h
Pitch (tph)	Pitch mass feed rate	ton/h
Green recyc (tph)	Green recycle mass feed rate	ton/h
Dry agg (tph)	Dry aggregate mass feed rate	ton/h
Paste (tph)	Total paste mass feed rate	ton/h
Coarse %	Coke coarse ratio	%
Fines %	Coke fines ratio	%
Inter. %	Coke intermediate ratio	%
Butts %	Butts ratio	%
Pitch %	Pitch ratio	%
Green recyc %	Green recycle ratio	%
Fines rot valve speed	Fines rotating valve feeder speed	rpm
Agg pre-heater T	Dry aggregate temperature after pre-heater	°C
Agg pre-heater_1 current	Dry aggregate pre-heater 1 current draw	Amp
Agg pre-heater_2 current	Dry aggregate pre-heater 2 current draw	Amp
MX1 KW mean	Paste mixer 1 mean power	kW/ton
MX1 KW max	Paste mixer 1 minimum power	kW/ton
MX1 KW min	Paste mixer 1 maximum power	kW/ton
MX1 therm oil T	Paste mixer 1 thermal oil temperature	°C
MX1 P	Paste mixer 1 pressure	psi
MX1 current	Paste mixer 1 current draw	Amp
Paste T between MX	Paste temperature after mixer 1	°C
MX2 KW mean pan 1	Paste mixer 2 pan 1 mean power	kW
MX2 KW mean pan 2	Paste mixer 2 pan 2 mean power	kW
MX2 KW mean rotor	Paste mixer 2 rotor mean power	kW
MX2 paste weight	Paste mixer 2 paste load	kg
MX2 dump gate pos	Paste mixer 2 dump gate opening	%
Paste T after MX2	Paste temperature after mixer 2	°C
Anode type (dim)	Anode dimensions	---
VC vib time	Vibrocompation vibration time	s
Green anode height	Green anode height	mm
VC bellows P	Vibrocompator bellows pressure	psi
PP mean ext T	Paste plant mean external temparature 3 hours around anode manufaturing time	°C
Core state	Qualitative quality of the anoe core	---

In Table 6, the ratios of the paste constituent are calculated based on total paste feed rate instead of on a dry aggregate feed rate basis. The size distribution of the different fractions and the dry aggregate is measured every day. For the coarse (~¼in to +30 mesh), intermediate (+50 mesh to +100 mesh) and butt fractions a sample is grabbed once every 12 hour shifts. The size distributions measured during the shift when the anode was produced was kept in the models. For the fines (- 100 mesh) and the dry aggregate, three samples are analysed during each operator shifts. The mean value of the size distribution analysis is stored in the datasets. The size distribution variables used in the analysis are listed in Table 7 where Rt stands for the material retained and Pt is an abbreviation for material that passed a given mesh size.

Table 7 – Size distribution variables

Variable ID	Properties	Units
Butts Rt3/8+Rt4	Butts Rt3/8+Rt4 fractions	%
Coarse Rt4	Coke coarse Rt4 fraction	%
Coarse Rt8	Coke coarse Rt8 fraction	%
Inter Rt50+Rt100	Coke intermidiate Rt50+Rt100 fraction	%
Fines Pt200	Coke fines Pt200 fraction	%
Agg Rt3/8	Dry aggregate Rt3/8 fraction	%
Agg Rt4@Rt30	Dry aggregate Rt4@Rt30 fractions	%
Agg Rt50+Rt100	Dry aggregate Rt50+Rt100 fractions	%
Agg Rt200+Pt200	Dry aggregate Rt200+Pt200 fractions	%
Agg Pt200	Dry aggregate Pt200 fraction	%

4.2.3 Baking furnace data synchronization

For the baking furnace, the first set of data is the fire move data. These are the maximum temperature obtained during one cycle for each stage of the baking process (e.g. max temperature under the first burner bridge, fire cycle time, mean baked weight for a section, etc. (Table 8)). For the temperature, the data are taken at different position in the oven section. They are described in Figure 17. Position A and B are used to measure external anode temperature during the three fire cycles and position C is used to measure the starting temperature which is the external temperature of the anode after the devolatilization of the pitch.

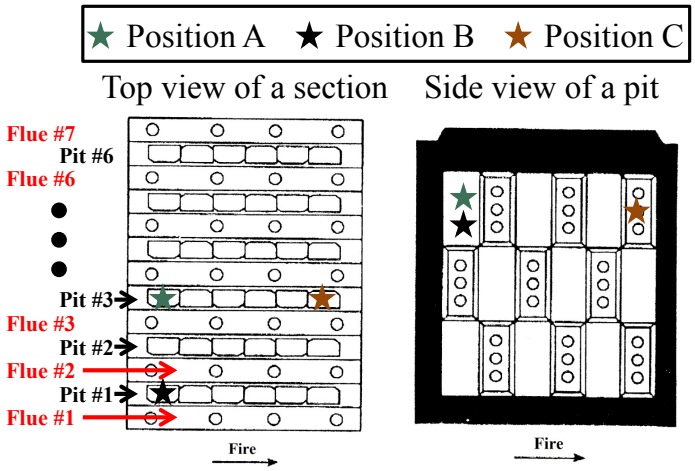


Figure 17 – Position of anode temperature measurements (Courtesy of Alcoa)

Table 8 – Fire move data

Variable ID	Properties	Units
Oven	Oven (1 or 2)	---
Fire	Fire (2 per oven)	---
Pit position	Position 1 or 2	---
Fire cycle T	Duration of fire cycle time	hr
BF pit starting T	Baking furnace temparature after burning of the pitch volatiles (position C)	°C
BB1 pit max T pos A	First burner bridge maximum temperature (position A)	°C
BB2 pit max T pos A	Second burner bridge maximum temperature (position A)	°C
BB3 pit max T pos A	Third burner bridge maximum temperature (position A)	°C
BB3 pit max T pos B	Third burner bridge maximum temperature (position B)	°C
BB1 flue 3 max T	First burner bridge maximum temperature (flue #3)	°C
BB3 flue 3 max T	Third burner bridge maximum temperature (flue #3)	°C
BB3 flue 3 T set point	Third burner bridge temperature set point (flue #3)	°C
BF mean ext T	Baking furnace mean external temparature during the baking cycle	°C
BF min ext T	Baking furnace minimum external temparature during the baking cycle	°C
BF max ext T	Baking furnace maximum external temparature during the baking cycle	°C
BF ext T var	Baking furnace external temparature variance during the baking cycle	°C

For the second set of data, information is extracted from the baking profile of a section in the baking furnace. An example of profile is presented in Figure 18. The anode temperature is measured at the position A and B as described above. The flue wall temperature is calculated as the mean value between the two flues on each side of the core anode (e.g. flues 1 and 2 for anode 1 and flues 3 and 4 for anode 2).

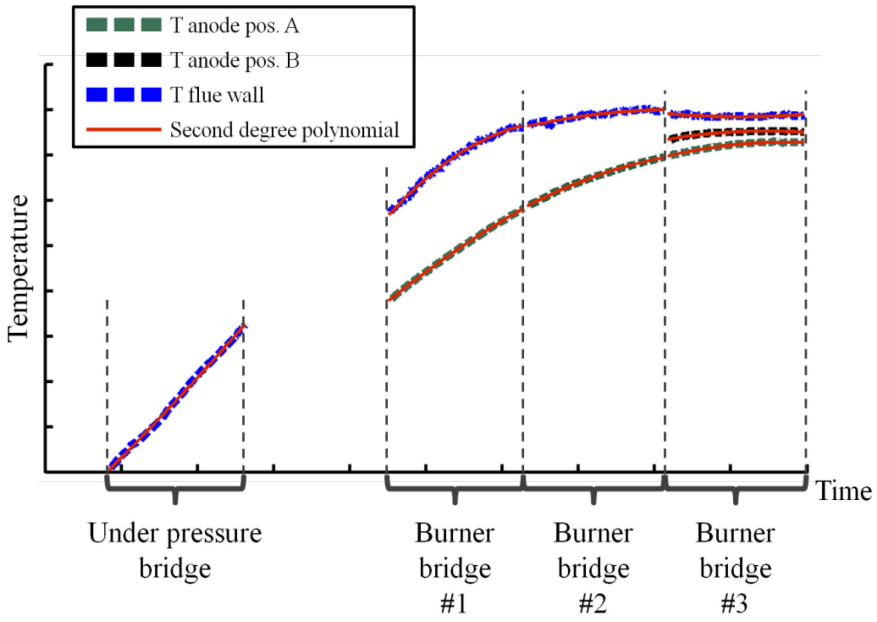


Figure 18 – Temperature profiles available for a section of the baking furnace and the second order polynomial fit.

To reduce the amount of data used in the models, the parameters of second order polynomial equations fitted on the profiles (i.e. A, B and C) were used instead of the full trajectory data. Two other methods could have been used instead of the polynomials. First, a PCA model on the trajectories could have been computed and the scores (**T**) could have been used as new **X** variables in the PLS model. Secondly, the entire trajectories (3 way array) could have been unfolded observation-wise and all the data points added as new variables in the PLS model (i.e. multiway method). For simplicity and due to time restriction, the polynomial fitting was used in this project.

$$f(t) = A + BT + CT^2 \tag{22}$$

In the previous equation, t is the time, T the temperature and A, B and C the coefficient of the polynomial. The pressure profile measured at the under pressure bridge (UPB) was also collected. A first order polynomial was used to fit this profile.

$$f(t) = A + BT \tag{23}$$

Figure 19 shows the profiles for the UPB.

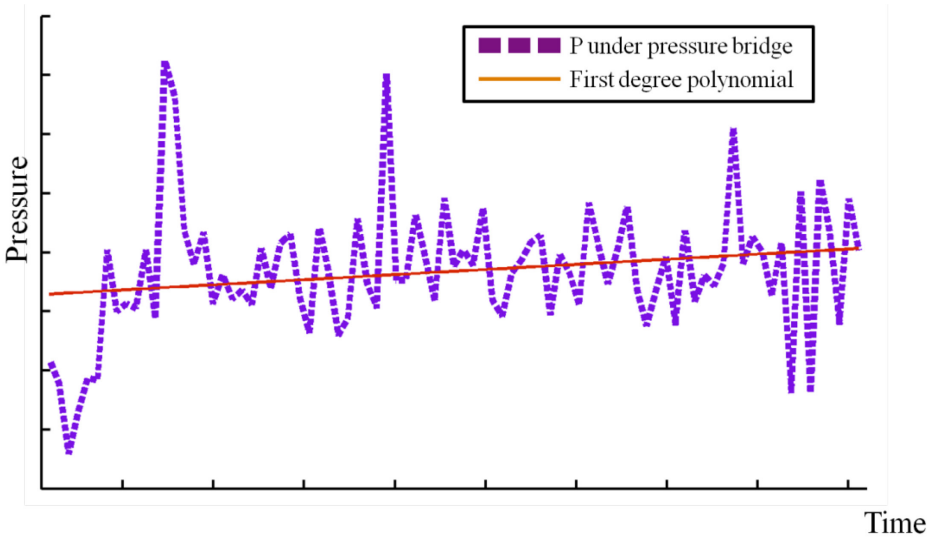


Figure 19 – Under pressure bridge pressure profile and the first order polynomial fit.

A first order polynomial was used for the pressure profile because these profiles are almost flat during the baking cycle. The minimum and maximum values as well as the variance of pressure during each cycle were also incorporated in the data matrix in addition to the parameters of the first order polynomial fit. However, the high frequency variations that appear on the figure above are due to the scale of the pressure axis (not shown for confidentiality reasons) which is very small. For this profile the mean value of the flue pressure on each side of the cored anode pit was computed at each time interval.

Table 9 – Baking furnace profiles

Variable ID	Profile
BB1 A T	Burner bridge 1 anode A temperature
BB2 A T	Burner bridge 2 anode A temperature
BB3 A T	Burner bridge 3 anode A temperature
BB3 B T	Burner bridge 3 anode B temperature
UPB flue T	Under pressure bridge flue temerature
BB1 flue T	Burner bridge 1 flue temperature
BB2 flue T	Burner bridge 2 flue temperature
BB3 flue T	Burner bridge 3 flue temperature
UPB P	Under pressure bridge pressure

For each of the profiles listed in Table 9, the polynomial coefficients are computed with a number of statistical variables listed in Table 10. This method is used to extract as much

information as possible for the different profiles and to compare them for all the cored anodes.

Table 10 – Baking profile information computed from the polynomial fits

Variable ID	Baking furnace profile information
0	A, polynomial constant
1	B, 1st degree coefficient
2	C, 2nd degree coefficient (for the temperature profiles only)
R^2	Fit of the polynomial (R^2)
min	Minimum value
max	Maximum value
mean	Mean value

Chapter 5 Results and discussion

The results obtained for the various models are presented and discussed in this chapter. The databases obtained using different sets of anode properties **Y** (physical properties and reactivities) were divided in two subsets: a training and an external validation sets. Approximately 2/3 of the data were used for training the models. The observations included in the training and validation sets were selected as follows. First, numbers 1 to 3 were randomly assigned to each observation. Then, observations labelled 1 and 2 were selected for training the models and those labelled 3 were kept for validation.

5.1 Physical property models

For physical properties, three models were investigated and the main difference between them is the information included in the process data matrix **X**. The first model only accounts for transport delay (i.e. residence time) between/within the various process equipments, and includes the raw material properties (listed in Table 5), the paste plant operating conditions (listed in Table 6 and Table 7) and basic information about the baking furnace (listed in Table 8). The second model uses the same process variables as the first, but the **X** matrix was expanded with lags of certain variables collected from the paste plant (variables measured at a higher frequency in Table 6) to account for additional process dynamics and uncertainties in the estimation of the various transport delays. Finally, the last physical model includes baking profile information and considers only transport delays from the paste plant (i.e. as for the first model).

A comparison between the first two models is proposed in order to verify whether including additional dynamics is necessary for predicting baked anode properties. They contain the same observations except for two anodes that had outliers in the lags and these two were excluded from the second model. The same procedure as described in section 4.2.2 was used to synchronize the data. For the second model, lags of selected paste plant variables were included to span a time window of -5 minutes to +2 minutes (sampled every minute) around their synchronized time (i.e. that of the first model). That is, the lags of the process variables correspond to the following time points [k-5 k-4 k-3 k-2 k-1 k k+1 k+2]

where k corresponds to the synchronized time (in minutes) of the first model considering only transport delays. Therefore, the selected paste plant variables appear at eight different times in the model spanning a time period of 7 minutes which is greater than vibro-compaction cycle time. The third model (i.e. with baking profile) will be discussed in more details in a separate section since it was built using a smaller subset of the original data.

Table 11 presents a summary of the results obtained for the first three models. The number of observations, X- and Y-variables, latent variables (LVs), as well as overall statistics showing the model performance in fit and validation are shown in this Table.

Table 11 – Summary of the results of physical models

Physical model	**Transport delay only**	**Dynamics (with lags)**	**Baking profile**
Number of training observations	438	436	302
Number of validation observations	238	238	---
Number of **X** variables	96	313	159
Number of **Y** variables	10	10	10
Number of latent variables	11	14	11
R^2X_{train} (%)	54,6	74,9	49,9
R^2Y_{train} (%)	45,5	44,8	51,8
Q^2Y_{train} (%)	28,7	24,7	23,5
R^2Y_{valid} (%)	42,9	40,5	---

The R^2 statistics are the cumulative multiple correlation coefficients quantifying the fit of the model (i.e. the variance explained by the model) for both the **X** and **Y** data. The R^2X is the fraction of variance of **X** (process variables) used by the model for predicting **Y**. The R^2Y is the fraction of variance of the **Y** (anode quality) explained by the model. The indices "train" and "valid" mean that the R^2 statistics were computed on the training and validation datasets, respectively. The Q^2Y statistics is the variance of **Y** predicted by the model through the cross-validation procedure (used for selecting the number of latent variables in each model). The R^2Y and Q^2Y statistics shown in Table 11 were computed based on all Y-variables (i.e. overall statistics). For the baking profile model there is no validation set since the number of observation is too small. The statistics for each individual Y-variable will be presented later for selected models.

The variance of the anode quality explained by the models (i.e. R^2Y) is similar for the first two models. It is not possible to compare the model with baking profiles for now since the model is not computed on the same set of observations. Section 5.1.2 discusses this model further. The Q^2 is 4% smaller for the lagged model compared to the transport delay model and the fit on the validation set is similar. Therefore the lags do not seem to improve the model predictive ability compared to the transport delay model. This is probably due to the fact that the dynamic in the paste plant is much slower than 5-10 minutes when operated in steady-state. The results of both models are presented in Table 12 for each anode properties.

Table 12 – Results of the first two physical models (transport delay and lags) for each physical properties

	Transport delay model		**Lags model**	
Variable ID	**R^2Y_{train} (%)**	**R^2Y_{valid} (%)**	**R^2Y_{train} (%)**	**R^2Y_{valid} (%)**
Green apparent density	63,80	60,56	63,77	58,13
Green weight	53,34	44,59	52,86	41,65
Baked weight	59,82	53,30	60,88	55,04
Thermal conductivity	27,31	30,56	28,37	32,14
Baked apparent density	40,80	41,95	40,54	40,51
Real density	42,08	46,93	43,12	45,30
Compressive strength	28,76	28,39	29,81	23,66
L_C	67,74	61,51	63,28	55,57
Young's modulus	27,37	22,40	26,95	21,07
Electrical resistivity	43,75	38,65	38,18	31,86

The **Y** fit ranges from 27,31% to 67,74% for the transport delay model and from 28,37% to 63,77% for the lags model. Usually, higher R^2 values are expected using statistical regression methods. However, it should be kept in mind that these results are arising from industrial process data and not from laboratory measurements or design of experiments. 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 these are considered as good results. Except for thermal conductivity, compressive strength and young's modulus, the model explains a significant amount of the variance of anode properties. Finally, except for the baked weight and real

density, the model with no lag is slightly superior or equal to the lagged model. For these reasons and because of the high number of variable in the lagged model, only the model without lag will be further discussed.

5.1.1 Physical transport delay model

The rest of this section will focus on performance assessment and interpretation of the physical property model considering transport delays only (first model). Table 13 displays the details of the variance explained for each of the 11 latent variables selected through the cross-validation procedure based on the training dataset. But the model was than validated with the validation dataset. The discussion will focus on the first three LVs because they represent most of the total variance explained by the model (i.e. 26 out of 45,5%). The acronym "cum" means the cumulative variance explained up the designated latent variable. The Q^2Y for the last three variables is decreasing because LVs are added as long as at least one **Y** variable Q^2 increases, even if the total Q2Y for this component is less than zero. The reader is referred to (Eriksson et al. 2001) for more information on how the CV procedure is done in the SIMCA-P+ software.

Table 13 – Physical transport delay model LV details

Latent variable (*A*)	**R^2Y (%)**	**$R^2Y_{(cum)}$ (%)**	**Q^2Y (%)**	**$Q^2Y_{(cum)}$ (%)**
LV 1	12,89	12,89	11,97	11,97
LV 2	8,86	21,75	8,96	19,86
LV 3	4,21	25,97	3,33	22,53
LV 4	7,01	32,97	6,36	27,45
LV 5	2,58	35,56	1,97	28,88
LV 6	3,72	39,27	2,88	30,93
LV 7	1,27	40,54	0,48	31,26
LV 8	1,32	41,85	0,19	31,39
LV 9	1,31	43,16	-1,67	30,25
LV 10	1,27	44,44	-0,83	29,67
LV 11	1,04	45,48	-1,34	28,73

First the model is checked for outliers using the Hotelling's T^2 (Figure 20) and Distance to model or DModX plots (Figure 21) computed based on all model dimensions (11 LVs).

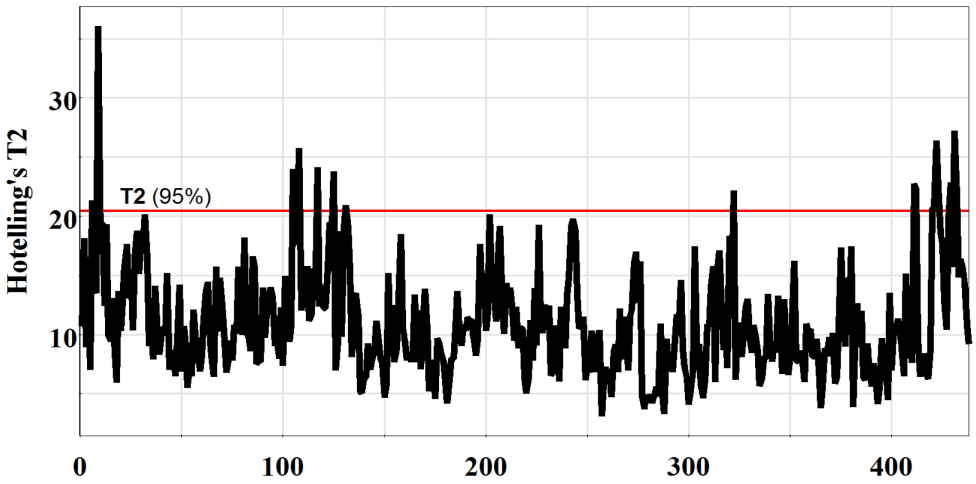


Figure 20 – Physical transport delay model Hotelling's T^2

The T^2 statistic is an indication of the distance of an observation to the origin of the A-dimensional latent variable space of the PLS model. It is also an indication of the leverage of a given observation on the regression model. The red line represents the approximate 95% confidence interval on the T^2 statistics. Nothing special can be observed in Figure 20. There are a just a small number of observations above the 95% T^2 distance and no significant excursion above the line can be observed.

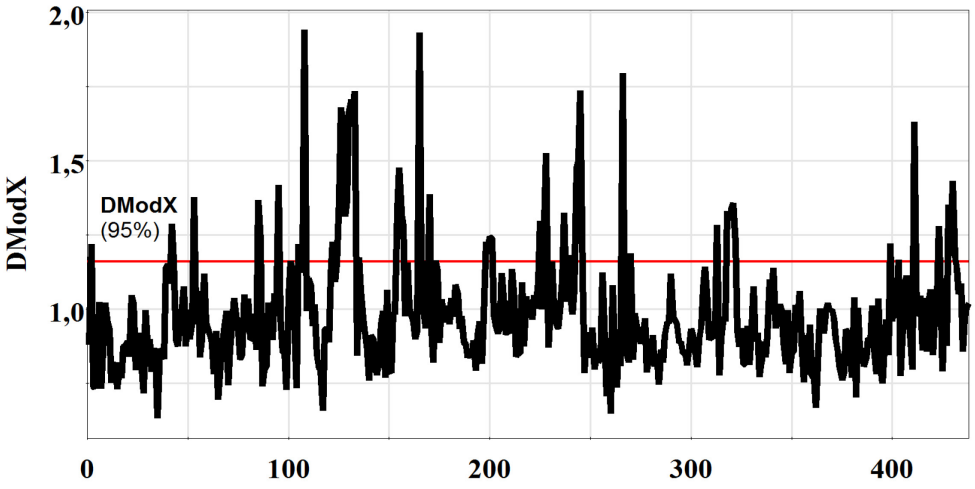
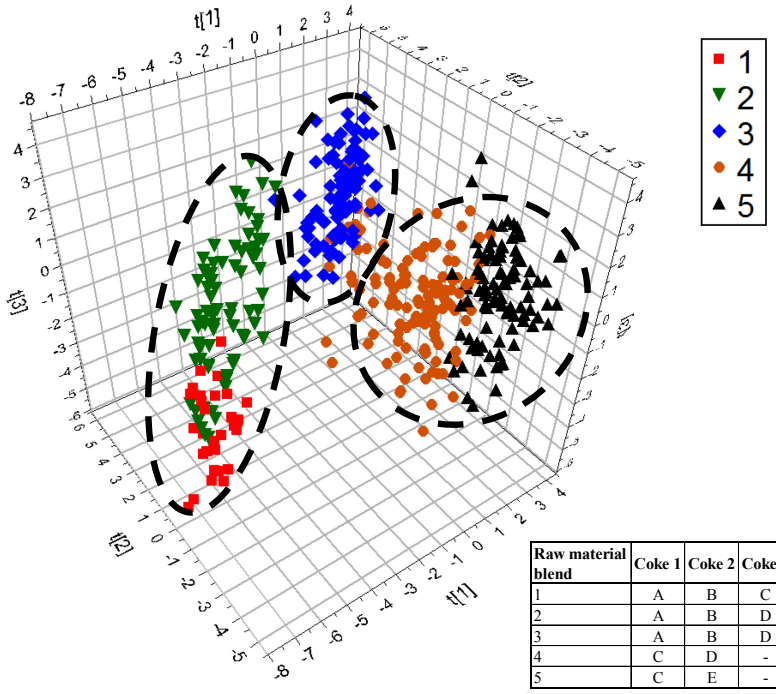


Figure 21 – Physical transport delay model DModX

The DModX represents the model residuals (i.e. distance of an observation off the lower dimensional latent variable space). The red solid line is the critical distance to the model with an approximate 95% confidence level. High DModX values indicate a different

correlation structure between the variables for these observations compared to the model. For this model, the number of high DModX observation is normal, it is expected to have a certain number of observations above the confidence interval (in this case 5%). And there is also no long and distinguishable disturbance.

The score plot of the first three LVs of the model are shown in Figure 22 (3D graph). The observations are colored according to the raw material blend used for its fabrication (Table 4).



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

Figure 22 – 3D score plot of first three LVs of the physical transport delay model

Each point in this plot is the result of a linear combination of all **X** variables according to their weight in the model. It is in fact the projection of the multivariate observations in a 96 dimensional space onto the hyperplane defined by the first three latent variables explaining most of the covariance between process data and anode properties. Observations (anodes) projected within a similar region of the score plot show similar patterns in their data structure; hence they are 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. The score plot is useful to investigate the

relationships between the observations. Figure 22 explicitly shows that raw material variations have a significant impact on anode quality. Three main clusters can be observed. It is possible to use contribution plots (defined in section 3.5) to investigate the combination of variables that are associated with these changes in the LV space. Figure 23 to Figure 26 show the contribution of each X-variable in the transition between blend 1-2, blend 2-3, blend 3-4 and blend 4-5 respectively. Each of these figures displays the combinations of variables associated with the shift between the center (mean) of each cluster. In these graphs, a positive contribution means that the mean value of a variable increased from the initial cluster to the final cluster and a negative contribution indicates a decrease in the variable value.

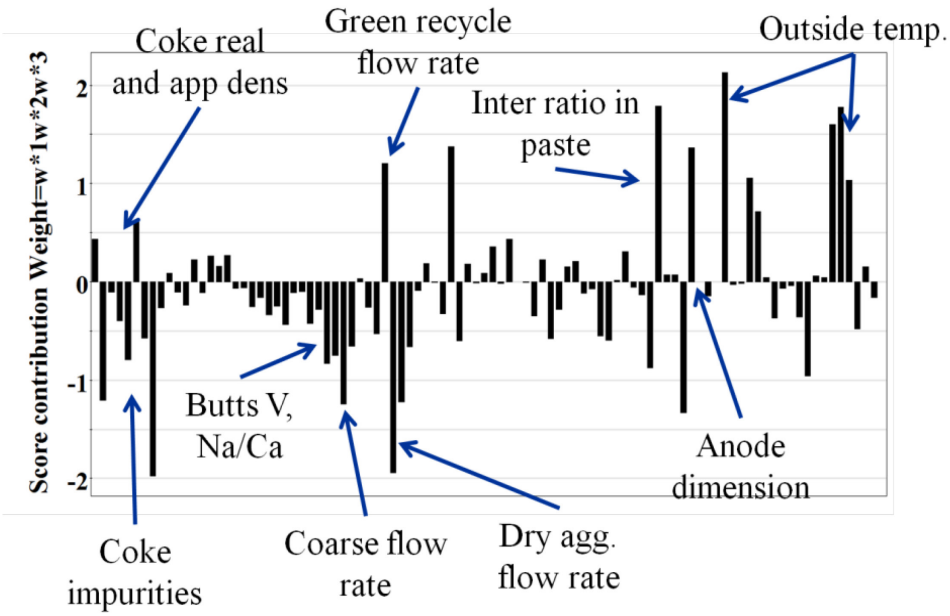


Figure 23 – Change in the variables from blend 1 to blend 2 for the physical transport delay model (contribution plot)

In Figure 23, the combination of variable that contributed to the change in anode quality (in the latent variable space) when the transition between blends 1-2 occurred can be interpreted. First, a small increase in coke density accompanied by a decrease in the coke and butts impurities was observed. This is expected from the coke supplier change. Since only one out of three coke suppliers changed, the properties of the coke blend did not change much and it is possible to see in Figure 22 that the two blends fall within the same region in the latent variable space of the model (i.e. both clusters are close to each other).

Changes were also made to process variables during this transition. Paste production rate was reduced (i.e. dry aggregate flow rate), and the balance between the coarse and intermediate size fractions in the paste was altered (less coarse, more intermediate). During the same period, anode dimensions were increased, and outside temperature increased (seasonal changes).

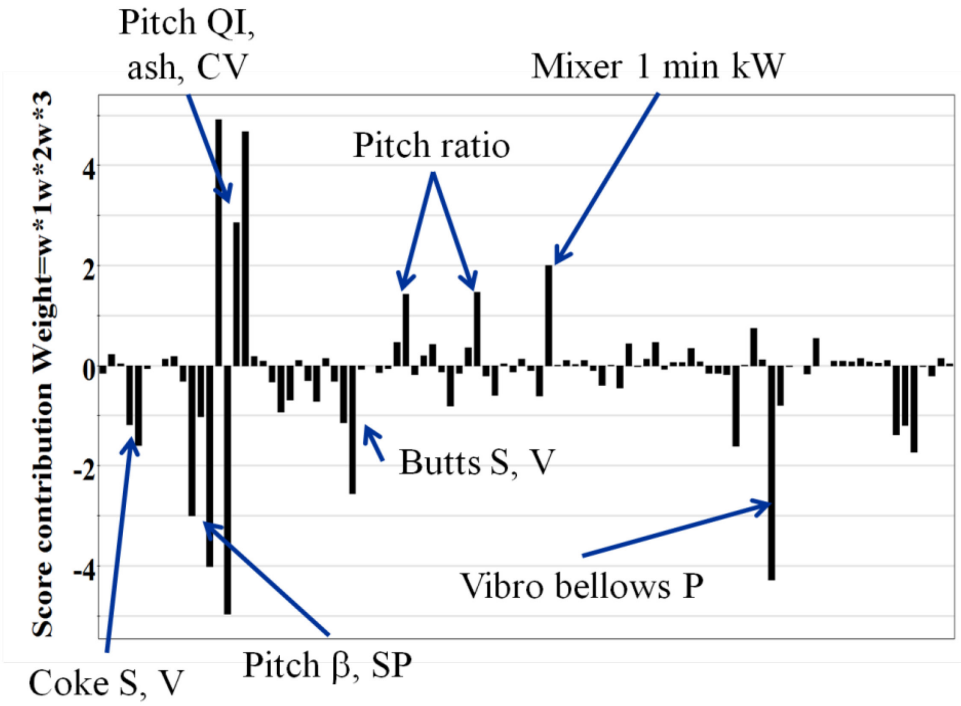


Figure 24 – Change in the variables from blend 2 to blend 3 for the physical transport delay model (contribution plot)

In Figure 24, the transition between blends 2-3 can be interpreted. The major difference between these blend is due to the pitch supplier change. The new pitch had almost double QI content, a higher coking value (CV) as well as lower softening point (SP) and β fraction content. The plant had to adjust the pitch quantity in the paste to compensate for the higher pitch demand due to the high QI. The bellows pressure was also lowered during the same period (i.e. step change). Deschambault had an anode underpitching problem related to some process operating conditions that was interfering with their usual method of pitch dosing. This high QI pitch and the higher binder ratio in the paste contributed to a higher mixer energy requirement. One of the three coke suppliers for this blend was supplying a

low sulfur coke for this period. This explains the lower vanadium and sulfur content in the coke and butts.

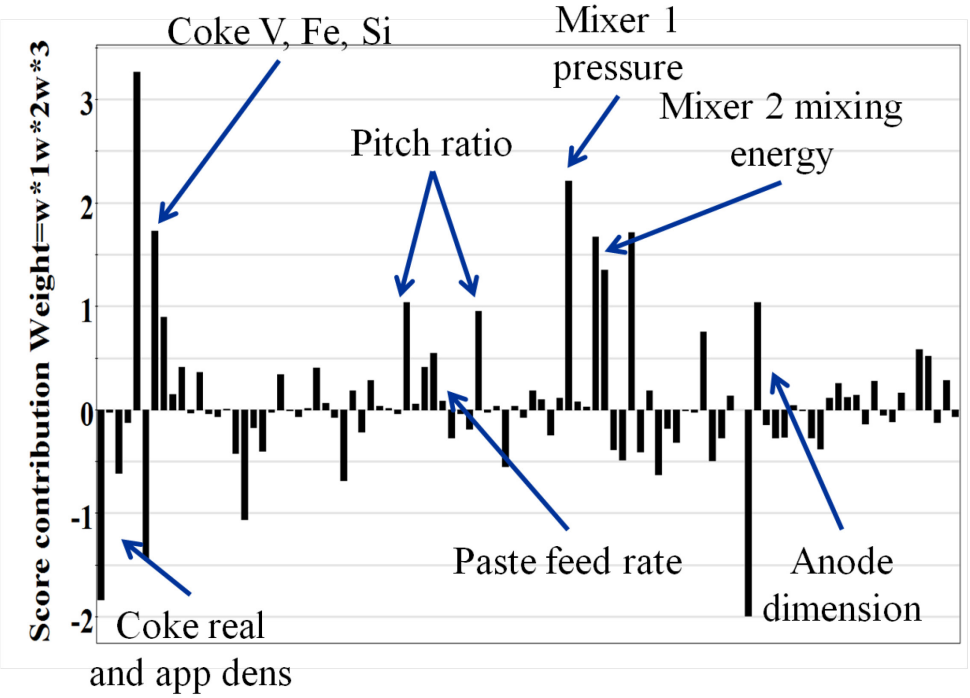


Figure 25 – Change in the variables from blend 3 to blend 4 for the physical transport delay model (contribution plot)

Another important shift occurred when transitioning between blends 3-4 when two cokes used in blend 3 were replaced by one new coke in blend 4. The variables contributing the most to this change are shown in Figure 25. Coke density (apparent and real) was lower for the new coke used in blend 4. The coke impurities were also higher. The pitch ratio in the anode was again adjusted for the new coke-pitch combination (i.e. higher pitch QI) and the lower bellows pressure (adjusted during the production of blend 3, see Figure 24). Mixer 1 pressure and mixer 2 energy increased when processing blend 4 compared to blend 3. It is not clear if these changes in pressure and mixing energy are due to raw material properties variations or process modifications or both. The feed rate of the paste is higher due to bigger anodes and the higher throughput of the paste which can cause higher mixing energy. On the other hand, coke density has also a good probability of influencing mixing properties.

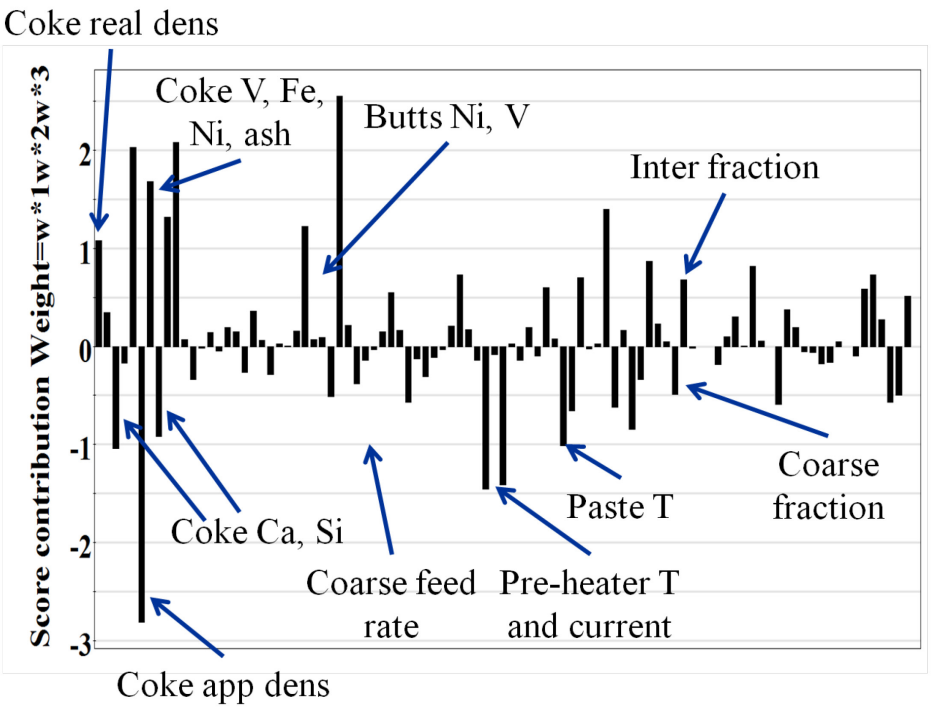


Figure 26 – Change in the variables from blend 4 to blend 5 for the physical transport delay model (contribution plot)

Another coke supplier transition occurred between blends 4-5 and Figure 26 shows the variables associated with the change. The new coke source had a higher real density while the apparent density was lower. Some coke impurities were higher while others showed the opposite behaviour. Process operation also changed since both the dry aggregate and paste temperatures were lower when processing blend 5 compared to blend 4. The coarse/intermediate size ratio was also different in blend 5.

The use of score plots and contributions plots allowed a quick and simple visualisation of the available data and the changes that occurred in raw materials and operation over time. Further interpretation of the model will investigated the relationships between the two groups of variables (**X** and **Y**) as extracted by the PLS model. This can be achieved using the loading plots. Figure 27 and Figure 28 present the loading plots for two pairs of latent variables: 1-2 and 1-3 respectively. These figures present the loadings of the variables belonging to both the **X** and **Y** groups. They allow interpreting the correlation structure (i.e. relationships) between the regressor variables **X** (raw material properties, paste plant and baking furnace) and the anode properties **Y**. Variables falling close to each other in the

loading plots are positively correlated (i.e. their loadings have the same sign) while those having loadings of opposite signs are negatively correlated. The farther away from the origin (i.e. the large is the absolute value of their loadings), the stronger is their influence in the model and their correlation with **Y**. The loadings of the first two latent variables are shown in Figure 27 and the loadings of LV1 and LV3 are shown in Figure 28. Only the first three are shown because they capture most of the variance in the data. It is important to note that the relationships described in this analysis are correlations only since the data were not collected from a design of experiment approach where independent changes of process variables are made. The analysis is rather performed on happenstance data collected during production.

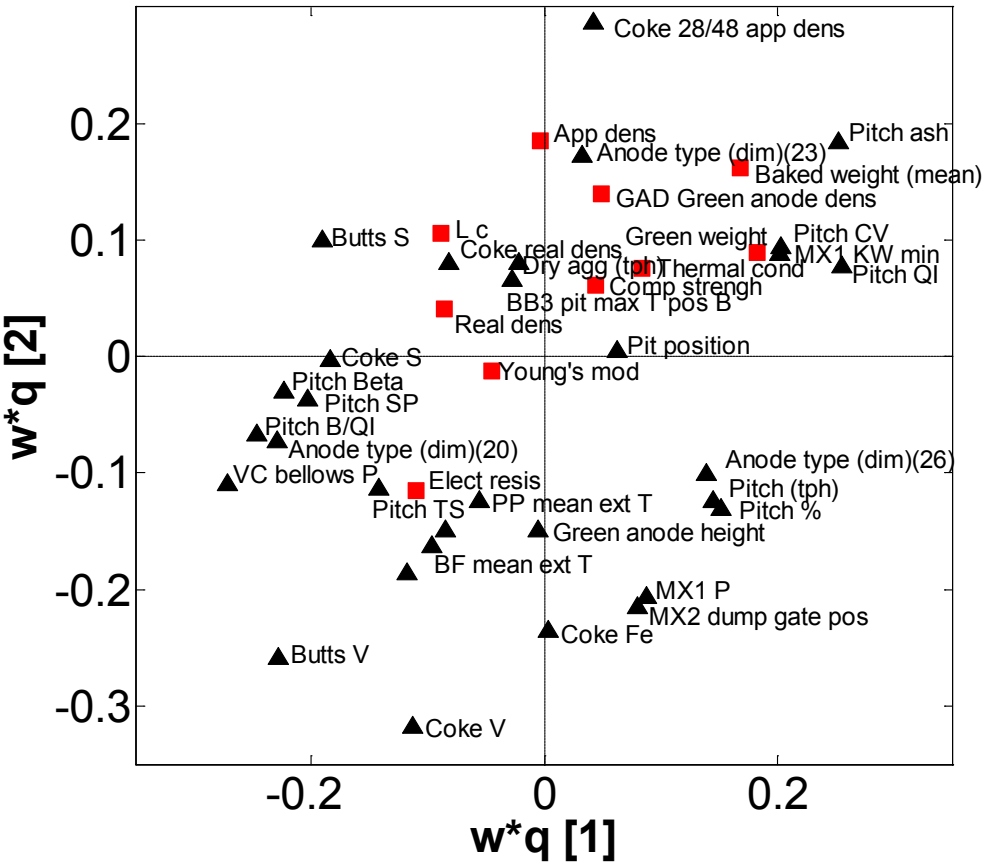


Figure 27 – Loadings of the first two LVs for the physical transport delay model showing variables with VIP greater than 1.0

In Figure 27, only the most important variables in the model, those having a VIP > 1 (see section 3.5), are shown together with their names. This makes the figure clearer and the

interpretation easier. The first LV (horizontal axis) is mainly driven by raw material variations, and especially pitch quality variations, since most coke and pitch properties have strong loading values in LV1 (i.e. along the horizontal axis). This is expected since the quality of the pitch changed drastically from one supplier to the other. The second LV (vertical axis) is also affected by some raw material variations (different from those extracted in LV1), but also by paste plant process parameters, and particularly by paste recipe variations. The variations in pitch ratio, anode dimensions and in the size distribution of the dry aggregate are captured by LV2. The increase in anode dimension (length) modelled using a binary variable (dim 20 = 0 and dim 23 = 1) is correlated to baked weight as expected. Green anode apparent density and baked apparent density are positively correlated with coke apparent density. They are correlated with pitch ratio in the first LV but not in the second LV. Coke apparent density and anode apparent density are inversely correlated to electrical resistivity because of the lower porosity (i.e. higher density). This property is negatively correlated with thermal conductivity, which is expected since they are both influenced by the density and microstructure of the anode. It is also inversely correlated with compressive strength in both LVs. Electrical resistivity is negatively correlated with L_C and real density in the first LV but positively in the second LV. It was expected that the two are inversely correlated since a more crystalline structure decreases the electrical resistance. The different behaviour in the two LVs suggests that several phenomena affect the electrical resistivity. It is also positively correlated with pitch toluene soluble (TS), butts and coke impurities. L_C and anode real density are positively correlated to coke real density. They are inversely correlated to pitch ratio, paste flow rate, MX1 pressure and MX2 dump gate opening. This variable is probably related to residence time of the paste in the process and thus mixing quality. Finally, the anode apparent density is correlated with dry aggregate temperature, coke bulk tapped density, aggregate size distribution and green anode density.

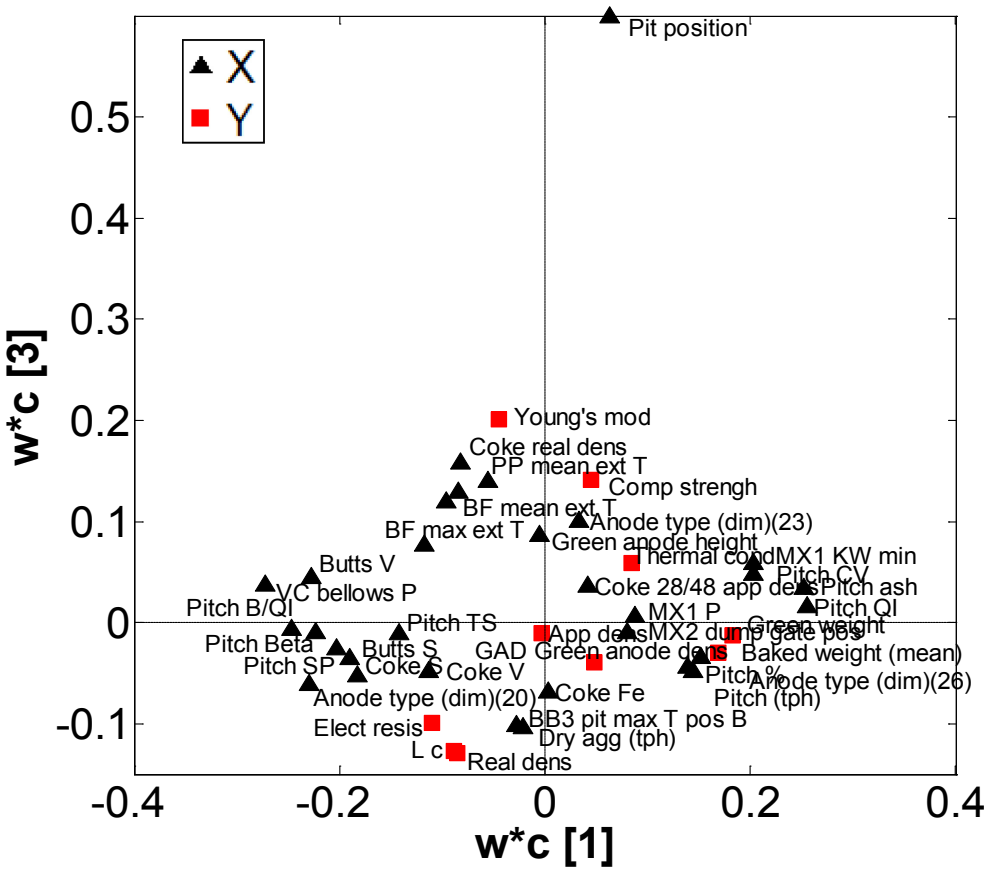


Figure 28 – Loadings LV1 and LV3 for physical transport delay model

The loading plot for LV3 vs LV1 is presented in Figure 28. This plot was selected to demonstrate the importance of the anode position in the baking furnace. The third latent variable is dominated by this variable. The pit position is an indication of the final anode temperature. There are no direct measurements of final baking temperature, but the hottest anode is located in position 1 and the coldest anode in position 2 of a furnace section. The pit position is correlated to electrical resistivity, anode real density and L_C. Pit position is also modelled using a binary variable with a value of 0 for the hottest anode (position 1) and a value of 1 for the coldest (position 2). Based on this definition, when the pit position variable increases, this means that the anode baking temperature decreases. In the figure, there is a negative correlation between pit position and the variables listed above. It is in fact a positive relationship. When the baking temperature increases, the electrical resistivity, the real density and the L_C increase.

Table 14 – VIP for physical transport delay model

Rank	X variable ID	VIP	Process
1	Pit position	2,850	Baking
2	Green anode height	2,291	Paste plant
3	Butts V	1,544	Raw material (butts)
4	VC bellows P	1,470	Paste plant
5	Pitch ash	1,456	Raw material (pitch)
6	Coke V	1,425	Raw material (coke)
7	Pitch QI	1,355	Raw material (pitch)
8	Anode type (dim)(20)	1,313	Paste plant
9	Pitch B/QI	1,312	Raw material (pitch)
10	Coke 28/48 app dens	1,273	Raw material (coke)
11	BF max ext T	1,239	Baking
12	Butts S	1,239	Raw material (butts)
13	Pitch Beta	1,215	Raw material (pitch)
14	BF mean ext T	1,214	Baking
15	MX1 P	1,197	Paste plant

The VIP is a metric used to quantify the importance of a variable in the model for predicting the Y-variables. Generally, a VIP greater than 1 is considered important (Eriksson, et al. 2001). Table 14 provides a list of the 15 most influential variables (i.e. highest VIPs) in the model without lag. A list of all VIP for this model is available in Table 27 in Appendix B. Pit position is the most important variable and this indicates the baking step as a strong influence on the model. This parameter is quantified using a binary variable taking into account the fact that the baking process is different from one position to another. These differences have an important impact, related to the importance of the local differences in the furnace on the anode properties. But this effect was not captured by the fire move information through max flue and anode temperature. Current anode information is not adequate to characterize the baking temperature distribution since only the pit position variable is listed as important in the model. Most of the high VIP variables are coke or pitch properties. It is expected since the raw material variations have a significant impact on anode quality. There are only a few paste plant operating condition variables and no size distribution variables with VIP > 1. This could be due to the limited feedback/feedforward control actions implemented on the process to counteract variations in raw material (except for the pitch ratio and bellows pressure).

The predictive ability of the model for each variable is now investigated. The statistics of the model for each of them are listed in Table 12. Figure 29 displays the measured values against the predicted values for each observation of the validation dataset (e.g. data not used for the calibration of the model) for each of the physical properties with transport delay model. In these figures, the black triangle is the measured value and the red square is the predicted value. All results are scaled to preserve confidentiality of the data. The goodness of fit or variance explained by the model on the validation set is displayed in each figure. It can be observed that most of the low frequency trends are well captured. Most of the unexplained variance is due to high frequency variations. This is consistent with the data available. Coke properties are only available on a weekly basis and are a composite of samples taken from unloading rail cars. The samples are taken before the coke is added on top of the silos and not at the bottom. There are some uncertainties in the residence time of the coke depending on the severity of the segregation and the amount of material in the silos. Furthermore the samples are taken from very large batches of coke and the uncertainties in the laboratory quality measurements are unknown. Therefore it is believed that coke properties vary at a higher frequency than what is currently measured on a weekly basis, hence a certain percentage of the variance in **Y** cannot be explained using the currently available **X** data. The reasons for lower R^2 values were explained earlier in this section, but looking at Figure 29, it is possible to classify the performance of the model. Good prediction obtained for a) green apparent density, b) green weight, c) baked weight, e) baked apparent density and f) baked real density. The L_C results seems widely spread, but this is due to a bimodal distribution from the different coring position. This phenomenon is explained later in this chapter. Fair results are obtained for j) electrical resistivity and d) thermal conductivity. For the electrical resistivity particularly (Figure 29 j)), the variance of the variable has increased considerably from observations of the beginning of 2010 to the end of the dataset. This means that a new, unmeasured source of variation, may have affected the process in this period.

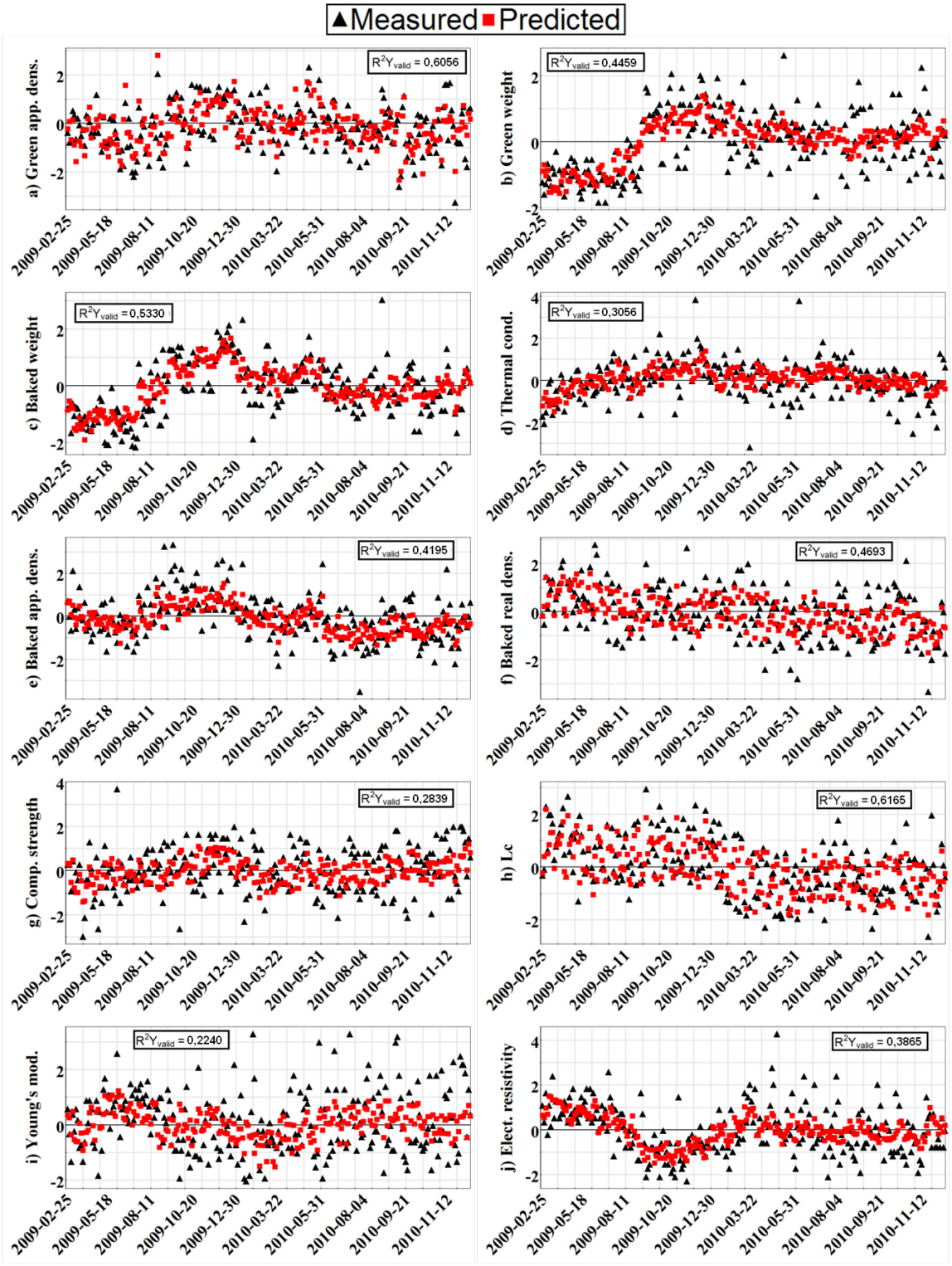


Figure 29 – Comparison of measured and predicted values (validation set) for: a) GAD, b) Green weight, c) Baked weight, d) Thermal conductivity, e) Baked app. density, f) Baked real density, g) Compressive strength, h) L_C, i) Young's modulus and j) Electrical resistivity

Poor predictability is obtained for g) compressive strength and i) Young's modulus. If coke mechanical properties and more detailed information on baking temperature would be available, this could help improve the model predictive ability for those variables. It is unclear at this point whether the lack of fit of the model for these variables is due to important measurements that are missing in the database or to a higher level of uncertainties in measuring the mechanical properties.

Overall, this model is considered good since it captures the important low frequency variation in anode quality. For the green and baked weight, the initial increase is due to an increase in anode length but the subsequent decrease is due to raw material variations.

As can be observed in Figure 29 h), the mean crystallite size L_C as a widespread distribution. This is due to the difference in final baking temperature. Precise final temperature for each anode is unknown, but they are different for each position in the pit. The difference in L_C is shown in Figure 30 where the pit position is identified by different colors. Anodes in position 1 (i.e. cored at the hottest position) have systematically higher L_C values (ANOVA P-value of 0). Since the coring position represents the coldest and hottest anodes in the pit, the real crystallite size distribution for anodes located at other position within the pits of the baking furnace should fall between these two extreme values (when manufactured with the same raw materials).

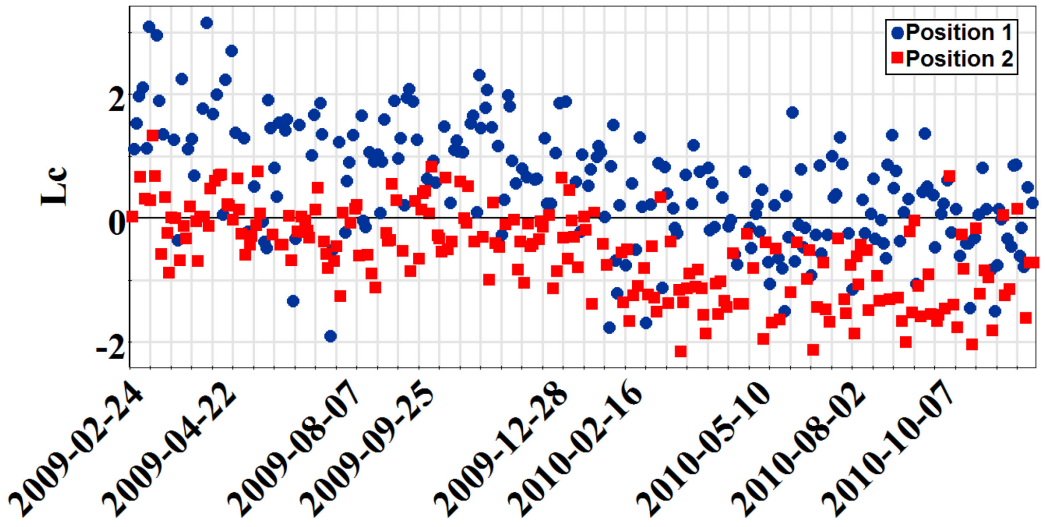


Figure 30 – L_C values colored according to the two pit position used for coring

The same trend can be observed for baked real density (ANOVA P-value of $3{,}83e^{-5}$) as shown in Figure 31. The real density is influenced by the baking process and also certainly by coke initial real density since approximately 65% of the anode is composed of coke.

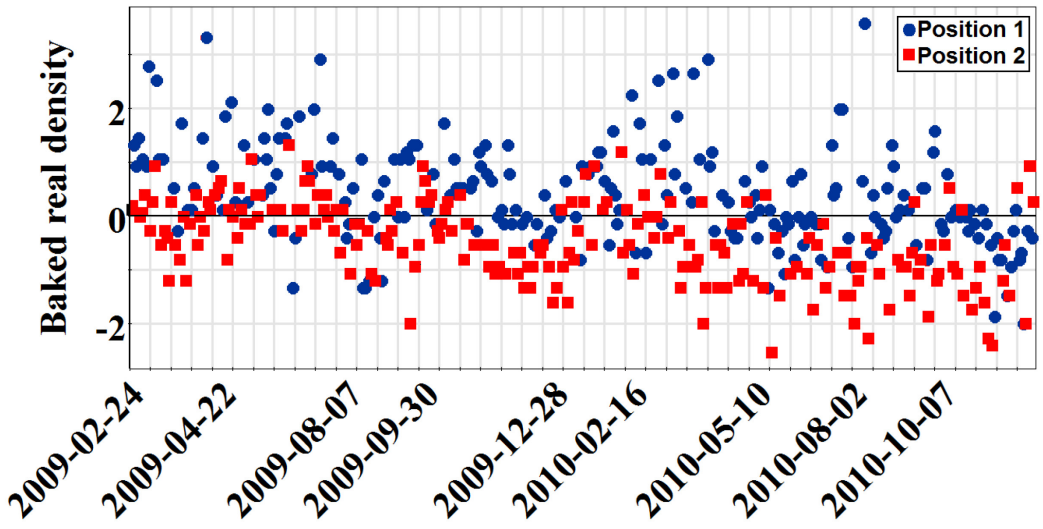


Figure 31 – Baked real density colored according to pit position used for coring

A final comment to conclude this section is that raw material variations are of great importance for explaining the variability in anode properties quality as shown in Figure 22. The position of the anodes in the baking furnace (i.e. pit position) was the most important factor in the physical property model. It is probably due to anode final baking temperature and soaking time, but these are not measured for each position. The addition of baking furnace profile did not improve the prediction ability of the model but this needs to be further investigated.

5.1.2 Model with baking furnace data

It is observed that baking is an important step in anode manufacturing and that pit position has a strong influence on anode quality. The data from the fire move dataset (described in section 4.2.3) does not seem to have a strong influence on the physical model discussed in section 5.1.1. For this reason baking profile data were retrieved from another data historian for a certain number of anodes. These profiles were pre-processed according to the procedure explained in section 4.2.3 and this new information (Table 9 and Table 10) was added to the transport delay model (i.e. into the original **X** matrix). A comparison between

the two models is provided in Table 15. Both models were computed based on the same reduced set of observations (i.e. anodes). The first model is computed with the fire move data (first physical model but with a different subset of observations) and the second with the baking profile data.

Table 15 – Model with baking profile comparison

Physical model	**Fire move**	**Baking profile**
Number of observations	302	302
Number of **X** variables	82	159
Number of **Y** variables	10	10
Number of latent variables	11	11
R^2X_{train} (%)	59,3	49,9
R^2Y_{train} (%)	51,2	51,8
Q^2Y_{train} (%)	30,8	23,5

Unfortunately, adding baking profile information does not significantly improve the predictive ability of the model. The flue wall temperature profiles might not capture the information related to the spatial distribution of the temperature of each anode in the pits. The variance explained only increase from 51,2 to 51,8%. For this reason, this model will not be discussed further in this thesis. However, it is recommended that the use of baking profiles be investigated in future work since other ways of analysing the profile data could improve the model and could help obtaining a clearer interpretation of the data and models.

5.2 Model for the reactivity properties

For the anode reactivity model, a new synchronized dataset was collected and organized. It was not possible to use the same observations as for the previously described models since physical properties and reactivities are not obtained based on the same core samples (i.e. not from the same baked anode). The same X-variables were used as regressors but the CO_2 and air reactivity measurements obtained from core samples were used as responses variables (**Y**). The reactivity model incorporates all six anode reactivity properties: CO_2 and air lost, dust and residue. The model statistics are listed in

Table 16 and Table 17.

Table 16 – Reactivity model overall statistics

Statistics	**Reactivity model**
Number of training observations	260
Number of validation observations	165
Number of **X** variables	93
Number of **Y** variables	6
Number of latent variables	3
Overall R^2X_{train} (%)	27,3
Overall R^2Y_{train} (%)	29,1
Overall Q^2Y_{train} (%)	15,3
Overall R^2Y_{valid} (%)	19,1

Table 17 – Reactivity model statistics for each variable

Variable ID	**R^2Y_{train} (%)**	**Q^2Y_{train} (%)**	**R^2Y_{valid} (%)**
CRD CO_2 dust	20,00	9,80	13,56
CRL CO_2 lost	28,26	16,28	18,86
CRR CO_2 residue	26,18	14,00	18,96
ARD Air dust	26,83	12,45	15,95
ARL Air lost	36,23	19,66	24,31
ARR Air residue	37,02	19,27	22,83

The model explains only 20% to 37% of the anode reactivity variability. This could be explained by the raw material data quality measured at a too low frequency (e.g. weekly measurements) and also by the lack of information of the microstructure of the anode which can have a strong impact of the reactivity. The model is checked for outliers using the Hotelling's T^2 (Figure 32) and Distance to model or DModX plots (Figure 33) computed based on all three LVs selected by cross-validation. It is possible to observe a slight deviation of the first few observations on the Hotelling's figure. But no major outliers are detected.

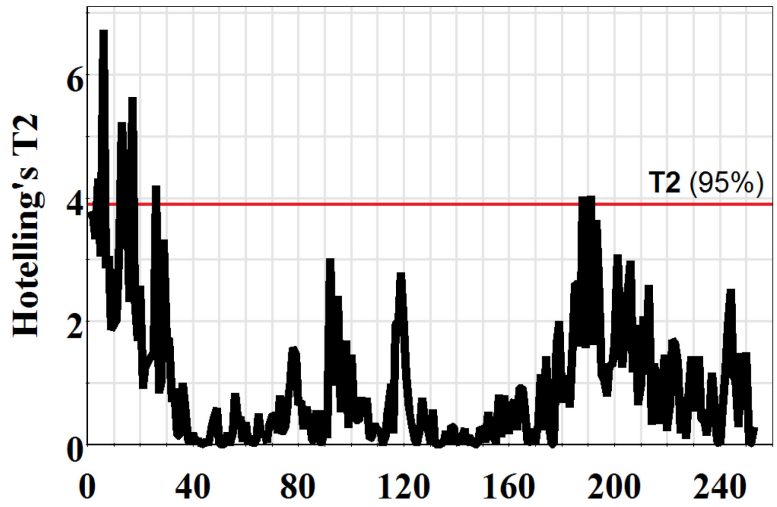


Figure 32 – Reactivity model Hotelling's T^2

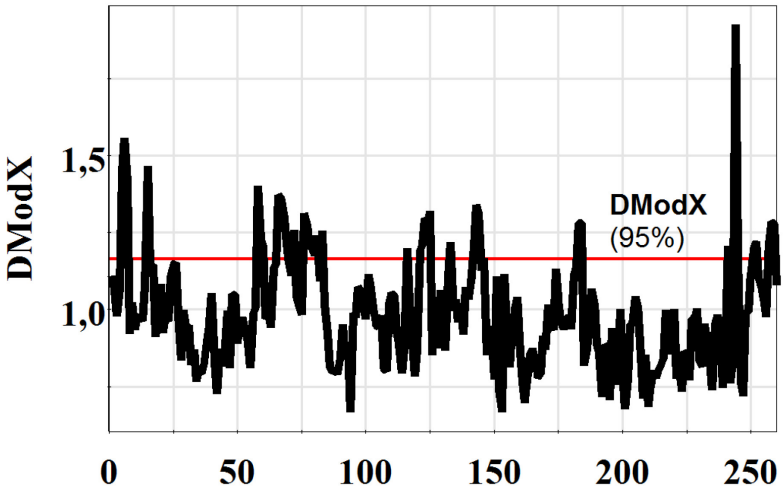
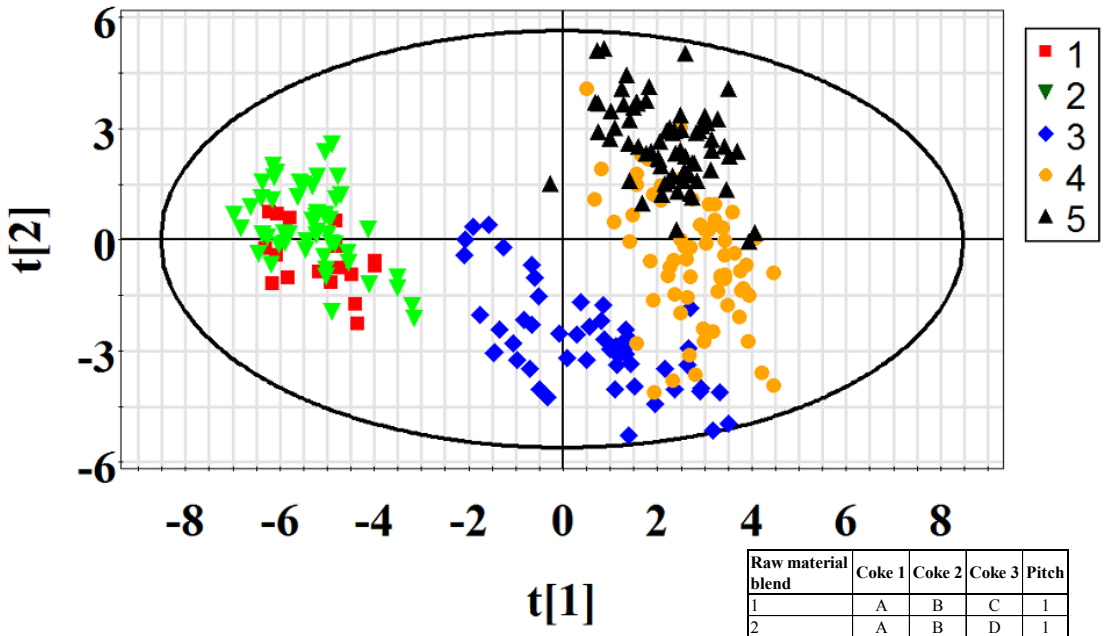


Figure 33 – Reactivity model DModX

A score plot of the first two latent variables is shown in Figure 34. The color code as well as the numbers 1-5 represent the same raw material blends as for the physical property model (see Table 4). Once again, it is possible to observe the impact of raw material variations due to supplier change.



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

Figure 34 – Reactivity model score plot (LV1-LV2)

The correlation between the variables can be interpreted using the loading plots for LV1 and LV2 shown in Figure 35. In this figure, it is possible to observe the relationship between the reactivity variables and the raw material properties and process parameters. Only the variables and their labels for those with a VIP > 1 are displayed. First, the residues are inversely correlated with the lost and dust variables. This is expected since the combination of carbon lost and dust is equal to the initial sample weight minus the residue. One interesting relationship is that air reactivity and CO_2 reactivity seems inversely correlated in the first component, but not in the second component. The air reactivity residue is positively correlated with calcium, coke apparent density and pitch QI and inversely correlated with vanadium, outside temperature, mixing energy and final baking temperature (i.e. correlated to pit position). The CO_2 reactivity residue is correlated with butts impurities (sodium content, sodium to calcium ratio, and vanadium), Pitch β fraction, pitch softening point, fines % and sulfur and is inversely correlated with pitch %, pitch QI, MX2 mixing energy and coke apparent density. Coke impurities, especially vanadium sulfur and the Ca/Na ratio seem to have an impact on both reactivities. Air residue (ARR) is strongly correlated to anode position in the oven, thus it is correlated to final baking

temperature. This could be the result of the desulfurization of the anode binder that occurs at higher temperature. There is a strong correlation between the air dust and air lost and the outside temperature. It could be due to seasonal temperature variations or simply that coke changes occur with season changes creating a fortuitous correlation with outside temperature.

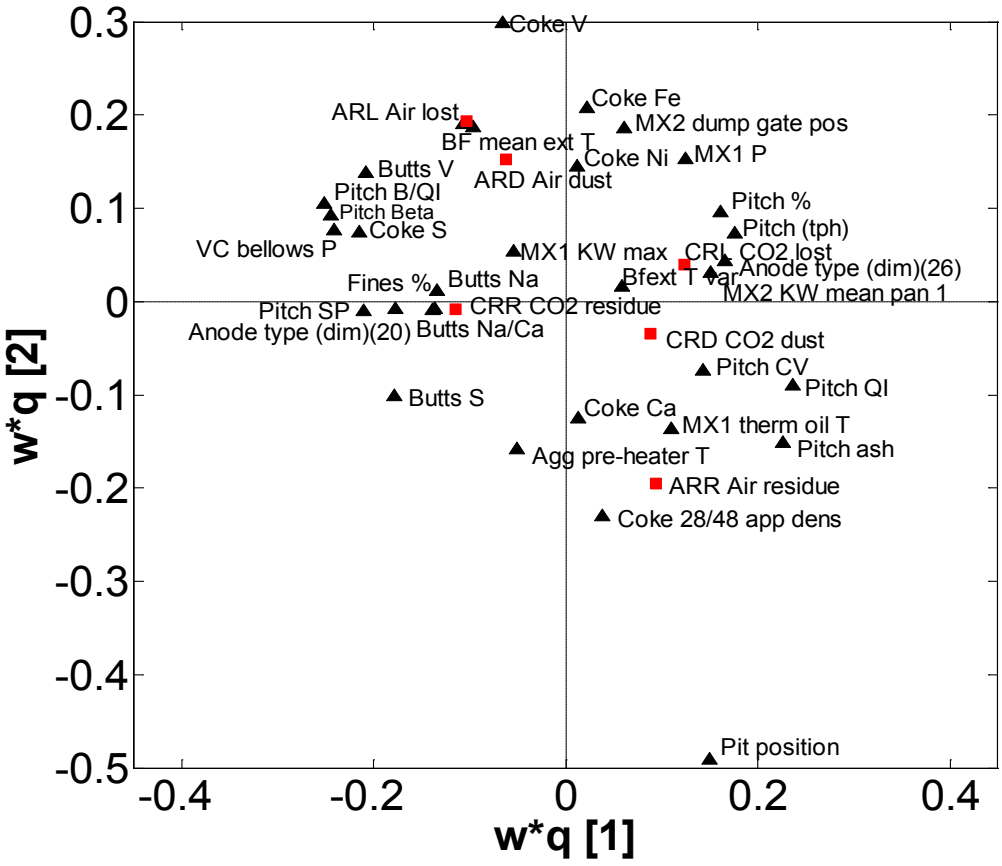


Figure 35 – Loadings of the first two LVs for the reactivity model showing variables having a VIP greater than 1.0

The 15 most influential variables (highest VIPs) for the reactivity model are listed in Table 18. A list of all VIP for this model is available in Table 28 in Appendix B. When compared to the VIPs of the physical model, more process variable have a greater influence on the model (i.e. have a VIP > 1). Pit position is the most important variable for prediction and as can be seen from the loading plot, it has a major influence on air reactivity. Binder impurities from coke and butts are also important contributors. Almost all of the pitch variables have a VIP > 1 and finally, the mixing operating conditions appear to have an impact on anode reactivity.

Table 18 – VIP for reactivity model

Rank	X variable ID	VIP	Process
1	Pit position	3,060	Baking
2	Butts V	1,797	Raw material (butts)
3	Coke 28/48	1,605	Raw material (coke)
4	Pitch B/QI	1,578	Raw material (pitch)
5	Coke V	1,544	Raw material (coke)
6	Pitch QI	1,523	Raw material (pitch)
7	Pitch Beta	1,511	Raw material (pitch)
8	Pitch SP	1,509	Raw material (pitch)
9	VC bellows	1,499	Paste plant
10	MX2 KW mea	1,491	Paste plant
11	Pitch ash	1,449	Raw material (pitch)
12	MX2 dump g	1,415	Paste plant
13	Butts S	1,400	Raw material (butts)
14	Pitch %	1,378	Paste plant
15	Butts Na/C	1,347	Raw material (butts)

The prediction for all six reactivity variables is presented in Figure 36. Once again, it is possible to observe that the low frequency variations are captured by the model. The total variance explained is low for each of the variables for this model, but it can give an indication of major deviations. In an attempt to improve the quality of the predictions, two separate models for CO_2 residue and air residue were computed. The statistics for these models are presented in Table 19.

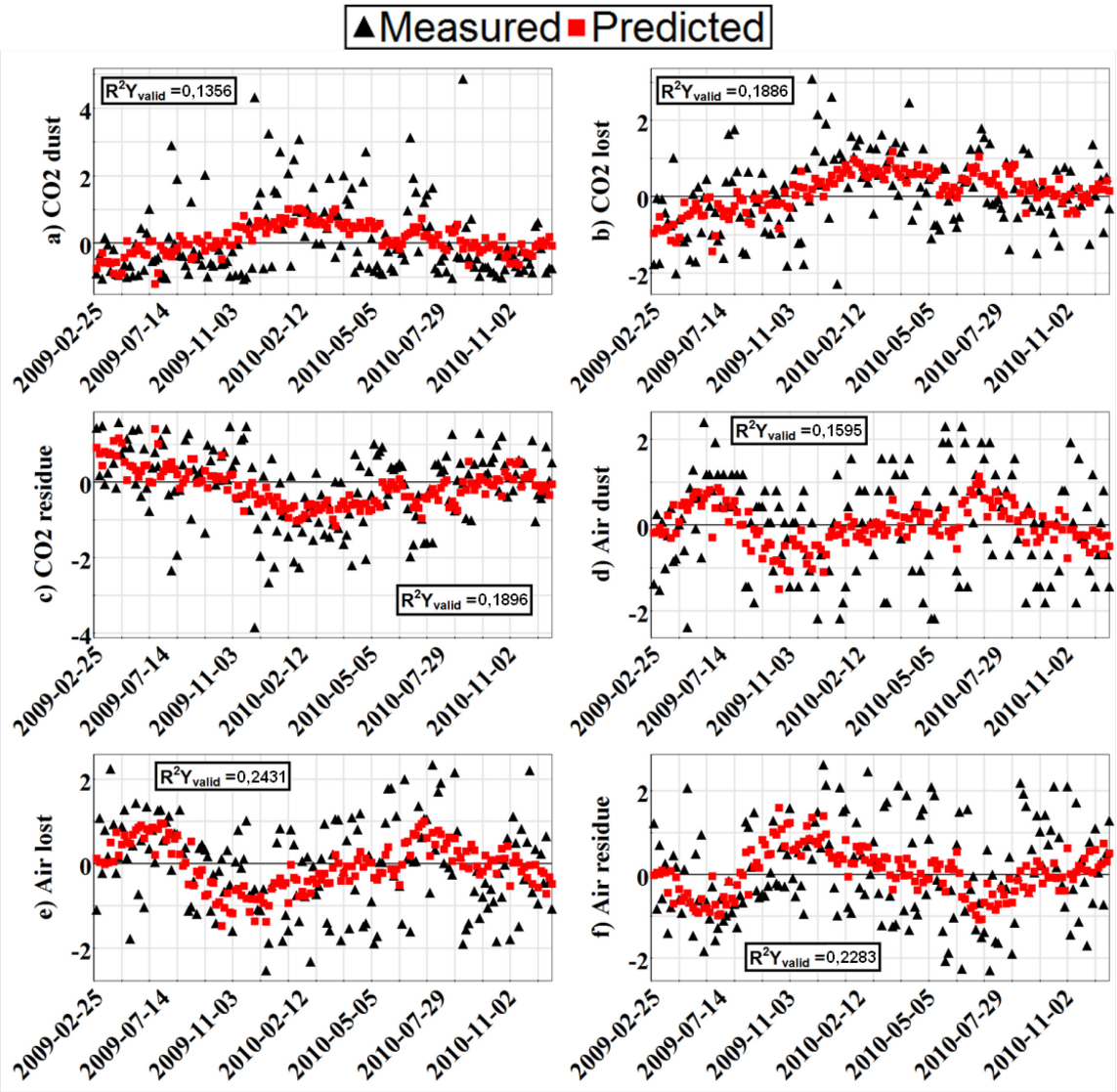


Figure 36 – Comparison of measured and predicted values (validation set) for: a) CRD, b) CRL, c) CRR, d) ARD, e) ARL, f) ARR

Table 19 – Residue models statistics

Statistics	**CRR model**	**ARR model**
Number of training observations	260	259
Number of validation observations	165	163
Number of **X** variables	93	93
Number of **Y** variables	1	1
Number of latent variables	2	3
R^2X_{train} (%)	21,4	26,3
R^2Y_{train} (%)	29,6	48,6
Q^2Y_{train} (%)	13,4	23,0
R^2Y_{valid} (%)	18,4	29,7

The variance explained for the CO_2 reactivity model is low. The air reactivity model performance is more acceptable close to 50% of variance explained. The better performance can be explained by the fact that ARR and CRR are not influenced by the same variables (i.e. the latent variable space associated with each reactivity is different). The coefficient of correlation between air residue and CO_2 residue is only -3,03%. When both the CO_2 and air reactivity are grouped in the same model, it tries to maximize the covariance for both group of variable at the same time. When the variables are not correlated, better performances can be obtained by creating separate models for these variables.

Figure 37 compares the R^2 of the training set and the predictive ability of the different reactivity models. For the CO_2 residue, the gain on explained variance is not significant and the overall model could be used for this variable. The air reactivity residue model is better than the overall model. For the ARR, the residue model explains 31% more variance than the overall model. This model should be used for ARR estimation from the raw material and process variables.

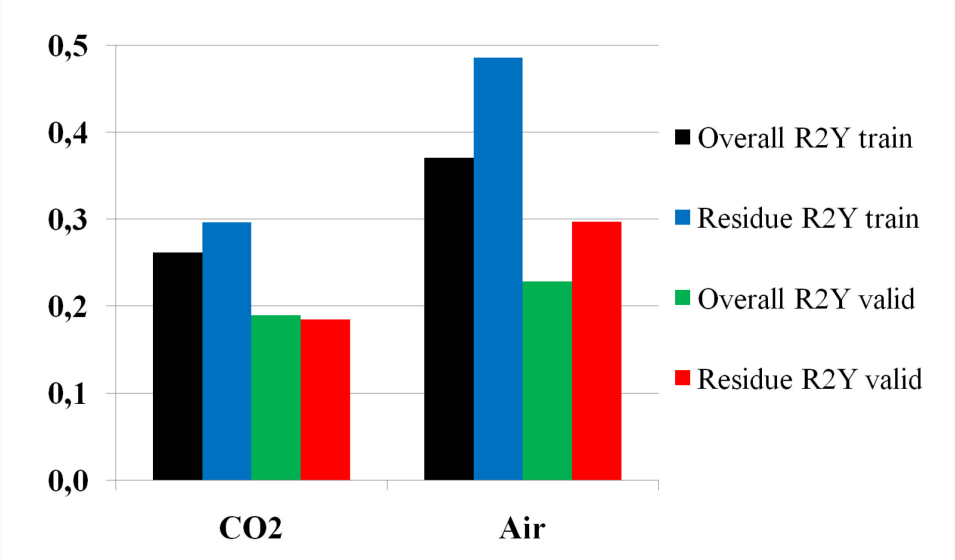


Figure 37 – Residue and overall reactivity models comparison

Table 20 compares the ten most important VIP for each residue model. It is possible to observe that there are differences between the two models. CRR is strongly correlated with mixer energy and most pitch variables as well as vanadium and calcium. For the ARR, the pit position (i.e. baking) is the most important variable, followed by vanadium and sulfur. Pitch properties are also important, and the external temperature appears correlated to variance in air residue. The later correlation could be due to coincidence with raw material changes.

Table 20 – Residue models VIP comparison

	CRR model		ARR model	
Rank	**Variable ID**	**VIP**	**Variable ID**	**VIP**
1	MX2 KW mean pan 1	2,033	Pit position	3,977
2	Pitch SP	1,824	Coke V	1,695
3	Butts V	1,767	Butts V	1,684
4	Pitch B/QI	1,692	Pitch ash	1,653
5	Butts S	1,666	Pitch B/QI	1,636
6	Pitch QI	1,665	Pitch QI	1,600
7	Pitch Beta	1,662	VC bellows P	1,599
8	VC bellows P	1,660	Pitch Beta	1,523
9	Pitch %	1,617	Coke S	1,512
10	Coke Ca	1,602	BF max ext T	1,430

The ARR predictions for the overall and residue model are compared in Figure 38. It can be observed that there is more variability captured by the ARR residue model. The low frequency trends are still predicted, but more high frequency variations are also captured by this model.

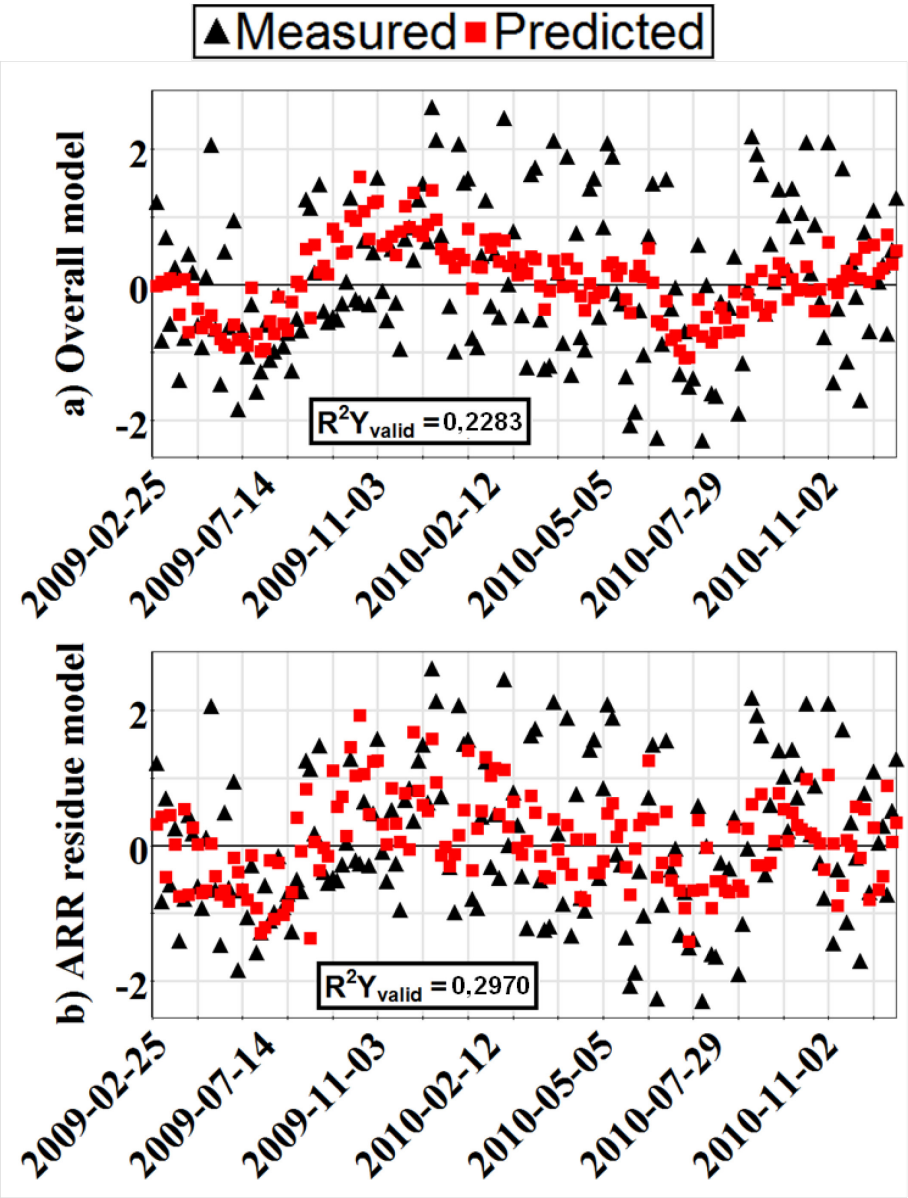


Figure 38 – ARR models prediction comparison: a) Overall model and b) ARR residue model

Conclusion

The manufacturing of prebaked anode for primary aluminum production is a complex multi-step and multivariate process affected by many sources of variations: raw material variability, paste processing and baking operating conditions. The ultimate goal is to manufacture anodes that perform well in the reduction cells, with low resistivity and long life to maximize metal production and minimize power consumption.

The main objective of this thesis was to investigate the different sources of anode quality variability and to improve quality control of the baked anode by predicting its properties. These properties are only available for the few anodes sampled each week. The prediction of anode properties could give an estimation of the quality for every anode produced. In regards of the industrial and multivariate nature of the data, a multivariate regression method is used to model the process. Due to anode core laboratory analysis limitations, two different classes of model are built: the physical properties and the reactivity properties. Different models are investigated for each of these two model class. Raw material properties and process operation condition variations are studied according to anode quality variation. Anode quality is assumed to be defined by its core sample properties.

The first step of the project was to extract and synchronize all the data, and then PLS models were computed. The first model is the physical property model. The variables included are (in model performance order) : L_C, green apparent density, baked weight, green weight, electrical resistivity, baked real density, baked apparent density, compressive strength, Young's modulus and thermal conductivity. The variance explained for these variables ranges from 27% to 68% of fit on the training data and 21% to 66% fit on the validation data. These are good results except for compressive strength, Young's modulus and thermal conductivity. It is demonstrated that high frequency variations are not captured by the model, but low frequency trends are well explained.

A second model with lagged variables from the paste plant was computed to check for a possible lack-of-fit due to potentially missing dynamic elements in the model in addition to transport delays between and mean residence times within the process equipments. The performance of the latter model did not significantly improved and the weight of each of

the lagged variables was similar for each time lag. In normal operation, the dynamics of the process is probably slower than the eight minutes lag studied in this work.

It was observed from the first physical property model that the fire move data do not have a significant weight in the model. However, it is known that the baking step has a strong influence on the anode properties. To try to capture some of the effects of baking temperature variations, profile data was extracted and added in a third physical property model. Temperature profile data were fitted using polynomials in order to reduce the amount of data to include in the model. Temperature profile polynomial coefficients and other statistics of the curves were used as new **X** variables in addition to the other raw material and process variables. This model did not significantly improve the model performance either.

The second set of models computed is related to reactivity properties. The CO_2 and air oxidation are major contributors to carbon overconsumption and they need to be minimized. The dust, lost and residue of both gas reactivities were used all together as the **Y** variables in the first modelling attempt (i.e. overall model). The variance explained ranged from 20% to 37% for the validation set (14% to 24% for the prediction set) depending on the Y-variables. The results for this model were worse than the results for physical properties. The impurities are well measured in the raw materials, but not the microstructure of the coke and its mechanical properties. This could be an important source of variability that is not measured. Although the explained variance for reactivites was lower, the low frequency trends were again captured by the models. High frequency variations for reactivities seem broader than for physical properties. In an attempt to improve residue predictions, two models were built for CO_2 and air reactivity separately. The CO_2 residue model has an explained variance of 30% (18% in validation). It is only a 13% improvement to the overall model. The air residue model explains 49% (30% in validation) of the variance of that property. This is a 31% improvement to the overall model. The two variables are modeled separately because they are almost independent (correlation coefficient of -3%). The VIP shows that they are not driven by the same combinations of variables. Both are influenced by catalytic impurities, but air residue is more influence by baking variations and CO_2 residue is more influence by mixing energy.

In conclusion, the major sources of variability are coming from the raw materials. Few process conditions have been changed during this analysis. This could be a reason why the baking process data do not have a significant impact on the models. Overall, a significant amount of variance is captured by the models. Augmenting the set of raw material properties, improving the quality of these data, and introducing variations in the most important process manipulated variables could improve the model performance, both in terms of interpretability and quality of predictions. These predictions of anode properties should be used for implementing early corrective actions on process manipulated variables in order to attenuate the impact of raw material changes. Hence, the models developed in this thesis enable the implementation of monitoring and feedback/feedforward quality control schemes at the carbon plant.

Alcoa Deschambault is a modern smelter with a modern anode plant. Data collection and storage is up to date compared to other smelters in the industry. Despite this ability, this work as pointed out some missing information from the raw material and process. A considerable amount of variability is unexplained and some additional data and modifications to data collection methods could improve this situation:

- Coke properties measured on few sample and from rail cars. This could be measured at a higher frequency and from the bottom of the silos. Also more information, in addition to impurities and apparent density are important. Microstructure, shape and mechanical properties could improve the models.

- Butts properties consist only in chemical compositions. Some physical properties like density and hardness could contribute to the models.

- More information on the fines like the Blaine number could help measured the dust fineness effects on the anode quality.

- Being able to perform some design of experiments on the paste plant process operating conditions could help capture some variability from the process, and assess potential feedback/feedforward control schemes for reducing variability

- More work should be done on the baking profile, important information are probably hidden in them.

The next step in this analysis could be the development of online sensors for coke properties and paste properties based on microscopy or image analysis. Also the models could be trained on a selection of good anodes and bad anodes to help map the raw material and process condition combinations that led to good anode to see if it is possible to build an optimal latent process space and be able to detect when the process deviates from this latent space.

Bibliography

Baron, John T., Stacey A. McKinney, Robert H. Wombles. “Coal Tar Pitch - Past, Present, and Future.” *Light Metals.* TMS, 2009. 935-939.

Barrilon, E., J. Pinoir. “Use of High-Sulphur Cokes in the Production of the Prebakes Anodes.” *Light Metals.* The Metallurgical Society of AIME, 1977. 2889-299.

Belitskus, David, Daniel J. Danka. “A Comprehensive Determination of Effects of Calcined Petroleum Coke Properties on Aluminium Reduction Cell Anode Propeties.” *Light Metals.* TMS, 1988. 429-442.

Burnham, Alison J., John F. MacGregor, and Román Viveros. “Latent variable multivariate regression modeling.” *Chemometrics and Intelligent Laboratory Systems* 48 (1999): 167-180.

Burnham, Alison J., Roman Viveros, John F. MacGregor. *Frameworks for Latent Variable Multivariate Regression* 10 (1996): 31-45.

Edwards, Les Charles, Keith J Neyrey, Lorentz Petter Lossius. “A Review of Coke and Anode Desulfurization.” *Light Metals 2007.* TMS, 2007. 895-900.

Edwards, Les, Vogt, Franz, Robinette ,Mike, Love, Ric, Ross, Anthony, McClung, Marilou, Roush, RJ, Morgan, William. “Use of Shot Coke as an Anode Raw Material.” *Light Metal.* TMS, 2009. 985-990.

Encyclopedia Britannica (a). http://www.britannica.com/EBchecked/topic/17944/aluminum-Al (accessed 02 14, 2011).

Encyclopedia Britannica (b). http://www.britannica.com/EBchecked/topic/18071/aluminum-processing (accessed February 14, 2011).

Engvoll, Marianne Aanvik, Harald A. Oye, Morten Sorlie. “Influence of Bath Contaminations on Anode Reactivity.” *Light Metals.* TMS, 2001. 661-667.

Eriksson, L., E. Johansson, N. Kettaneh-Wold, S. Wold. *Multi- and Megavariate Data Analysis Principles and Applications.* Umea, Sweden: Umetrics AB, 2001.

Fischer, Werner K., and Raymond C. Perruchoud. “Interdependence Between Properties of Anode Butts and Quality of Prebaked Anodes.” *Light Metals.* TMS, 1991. 721-724.

Fischer, Werner K., Raymond C. Perruchoud. “Influence of Coke Calcining Parametters on Petroleum Coke Quality.” *Light Metals.* TMS, 1985. 811-824.

Fischer, Werner K., Keller, Felix, Raymond C. Perruchoud. “Interdependence between Anode Net Consumption and Pot Design, Pot operating Parameters and Anode Properties.” *Light Metals.* TMS, 1991. 681-686.

Fischer, Werner K., Felix Keller, Raymond C. Perruchoud, S. Oderbolz. “Baking parameters and the resulting anode quality.” *Light Metals.* TMS, 1993. 683-694.

Fischer, Werner K., Ulrich Mannweiler, Felix Keller, Raymond C. Perruchoud, and Urs Bühler. *Anodes for the Aluminum Industry.* 1st Edition. Sierre, Switzerland: R&D Carbon, 1995.

Geladi, Paul, Bruce R. Kowalski. “Partial Least-Squares Regression: A Tutorial.” *Analytica Chimica Acta* 185 (1986): 1-17.

Gendron, Michel, Stephen Whelan, Katie Cantin. “Coke Blending and Fines Circuit Targeting at the Alcoa Deschambault Smelter.” *Light Metals.* TMS, 2008. 861-864.

Golubic, Thomas A., Stacey A. McKinney, Robert H. Wombles. “High Softening Point Coal Tar Pitch as Anode Binder Pitch.” *Light Metals.* TMS, 2010. 909-911.

Grjotheim, Kai, Halvor Kvande. *Introduction to aluminium electrolysis : understanding the Hall-Hérloult process.* 2nd Edition. Düsseldorf, Germany: Aluminium-Verlag, 1993.

Höskuldsson, Agnar. “PLS Regression Methods.” *Journal of Chemometrics* 2 (1988): 211-228.

Houston, Geoffrey. J., H. A. Oye. “Reactivity Testing of Anode Carbon Materials.” *Light metals.* Metallurgical Society of AIME. Light Metals Committee, 1985. 885-899.

Hulse, Kristine Louise. *Anode Manufacture: Raw Materials, Formulation and Processing Parameters.* Sierre: R&D Carbon Ltd., 2000.

Hume, Sheralyn M., W. K. Fischer, R. C. Perruchoud, Barry J. Welch. “A model for petroleum core reactivity.” *Light metals.* TMS, 1993b. 525-534.

Hume, Sheralyn M., W. K. Fischer, R. C. Perruchoud, James B. Metson, R. Terry Baker. “Influence of Petroleum Coke Sulphur Content on the Sodium Sensitivity of Carbon Anodes.” *Light Metals.* TMS, 1993a. 535-542.

Jones, S. S., R. D. Hildebrandt, M. C. Hedlund. “Influence of High-Sulfur Cokes on Anode Performance.” *Light Metals.* The Metallurgical Society of AIME, 1979. 553-574.

Jones, Samuel S. “Anode-Carbon Usage in the Aluminum Industry.” Dans *Petroleum-Derived Carbons*, de John D. Bacha, John W. Newman J. L. White, 234-250. Washington, DC: American Chemical Society, 1986.

Keller, Felix, Fischer, Werner K.. “Development of Anode Quality Criteria by Statistical Evaluation of Operational Results in the Electrolysis.” *Light Metals.* TMS, 1982. 729-740.

Kourti, Theodora. “Application of Latent Lariable Methods to Process Control and Multivariate Statistical Process Control in Industry.” *International Journal of Adaptive Control and Signal Processing*, 2005: 213-246.

Kourti, Theodora. “Process analysis and abnormal situation detection: from theory to practice.” *Control Systems Magazine* (IEEE) 22, n° 5 (2002): 10-25.

MacGregor, John F., Theodora Kourti. “Statistical process control of multivariate processes.” *Control Engineering Pratice*, 1995, éd. 3.

Mannweiler, Ulrich, and Felix Keller. “The Design of a new Anode Technology for the Aluminium Industry.” *JOM* 46 (1994): 15-21.

Martens, Harald. “Reliable and relevant modelling of real world data: a personal account of the development of PLS Regression.” *Chemometrics and Intelligent Laboratory Systems* 58 (2001): 85-95.

McClung, Marilou, J. Anthony Ross. “A Method to Correlate Raw Material Properties to Baked Anode Core Performance.” *Light Metal.* TMS, 2000. 481-486.

McHenry, E.R. “Industrial Pitch Quality of the Future.” *Proceedings of the Fourth Australasian Aluminium Smelter Workshop.* 1992. 192-232.

Nomikos, Paul, John F. MacGregor. “Multivariate SPC Charts for Monitoring Batch Processes.” *Technometrics* 37, n° 1 (1995): 41-59.

Perruchoud, Raymond C., Werner K. Fischer. “Determination of the Sodium Sensitivity of Petroleum Coke.” *Light Metals.* TMS, 1991. 581-584.

Perruchoud, Raymond C., Markus W. Meier, Werner K. Fischer. “Survey on Worlwide Prebaked Anode Quality.” *Light Metals.* TMS, 2004. 573-578.

Perruchoud, Raymond C., Werner K. Fischer, Jorge Brown, Markus W. Meier. "Worldwide Pitch Qulaity for Prebaked Anodes." *Light Metals.* TMS, 2003. 509-518.

Sinclair, Keith A., Sadler, Barry A., "Which Strategy to Use When Sampling Anodes for Coring and Analysis? – Start With how the Data Will be Used.' *Light Metals.* TMS, 2009. 1037-1041.

Sorlie, Morten. *Pitch Quality (As seen from a users side)* (21 September 2010).

Turner, N. R. "Recent Trends in Binder Pitches for Reduction Anodes." *JOM*, 1993.

Valle, Sergio, Weihua Li, S. Joe Qin. "Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods." *Industrial & Engineering Chemistry Research* 38 (1999): 4389-4401.

Vitchus, Bernie, Frank Cannova, Howard Childs. "Calcined Coke from Crude Oil to Customer Silo." *Light Metals.* TMS, 2001. 589-596.

Westerhuis, Johan A., Theodora Kourti, John F. MacGregor. "Analysis of Multiblock and Hierarchical PCA and PLS Models. *Journal of Chemometrics* 12 (1998): 301-321.

Weterhuis, Johan A., Stephen P. Gurden, Age K. Smilde. "Generalized Contribution Plots in Multivariate Statistical Process Monitoring." *Chemometrics and Intelligent Laboratory Systems* 51 (2000): 95-114.

Wise, Barry M., Neal B. Gallagher. "The process chemometrics approach to process monitoring and fault detection." *Journal of Process Control*, 1996: 329-348.

Wold, Svante. "Chemometrics; what do we mean with it, and what do we want from it?" *Chemometrics and Intelligent Laboratory Systems*, 1995: 109-115.

Wold, Svante. "Cross-Validatory Estimation of the Number of Components in Factor a Principal Components Models." *Technometrics* 20, n° 4 (1978): 397-405.

Wold, Svante, Johan Trygg, Anders Berglund, Henrik Antti. "Some Recent Developments in PLS Modeling." *Chemometrics and Intelligent Laboratory Systems* 58 (2001): 131-150.

Wold, Svante, Kim Esbensen, Paul Geladi. "Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems*, 1987: 37-52.

Wold, Svante, Michael Sjöstöm, Lennart Eriksson. "PLS-regression: a basic tool of chemometrics." *Chemometrics and Intelligent Laboratory systems* 58 (2001): 109-130.

Wombles, Robert H., John Thomas Baron. "Laboratory anode comparision of chinese modified pitch and vacuum distilled pitch." *Light Metals.* TMS, 2006. 535-540.

Appendix A Models variable lists

The appendix includes lists of all variable included in each of the model discussed.

Table 21 – **X** variables for the physical transport delay model

#	X variable ID	#	X variable ID	#	X variable ID
1	Coke real dens	33	Inter. (tph)	65	Inter Rt50+Rt100
2	Coke Na	34	Butts (tph)	66	Fines Pt200
3	Coke Ca	35	Pitch (tph)	67	Agg Rt3/8
4	Coke S	36	Green recyc (tph)	68	Agg Rt4@Rt30
5	Coke V	37	Dry agg (tph)	69	Agg Rt50+Rt100
6	Coke 28/48 app dens	38	Paste (tph)	70	Agg Rt200+Pt200
7	Coke Fe	39	Coarse %	71	Agg Pt200
8	Coke Si	40	Fines %	72	Anode type (dim)(20)
9	Coke Ni	41	Inter. %	73	Anode type (dim)(23)
10	Coke ash	42	Butts %	74	Anode type (dim)(26)
11	Pitch SP	43	Pitch %	75	Green anode height
12	Pitch TS	44	Green recyc %	76	VC bellows P
13	Pitch Beta	45	Fines rot valve speed	77	PP mean ext T
14	Pitch QI	46	Agg pre-heater T	78	Oven
15	Pitch B/QI	47	Agg pre-heater_1 current	79	Fire
16	Pitch CV	48	Agg pre-heater_2 current	80	Pit position
17	Pitch ash	49	MX1 KW mean	81	Fire cycle T
18	Pitch S	50	MX1 KW max	82	BF pit starting T
19	Pitch dist	51	MX1 KW min	83	BB1 pit max T pos A
20	Butts Al	52	MX1 therm oil T	84	BB2 pit max T pos A
21	Butts Ca	53	MX1 P	85	BB3 pit max T pos A
22	Butts %F	54	MX1 current	86	BB3 pit max T pos B
23	Butts ash	55	Paste T between MX	87	BB1 flue 3 max T
24	Butts Fe	56	MX2 KW mean pan 1	88	BB3 flue 3 max T
25	Butts Ni	57	MX2 KW mean pan 2	89	BB3 flue 3 T set point
26	Butts Si	58	MX2 KW mean rotor	90	BF mean ext T
27	Butts Na	59	MX2 paste weight	91	BF min ext T
28	Butts S	60	MX2 dump gate pos	92	BF max ext T
29	Butts V	61	Paste T after MX2	93	BF ext T var
30	Butts Na/Ca	62	Butts Rt3/8+Rt4	94	Core state(1)
31	Coarse (tph)	63	Coarse Rt4	95	Core state(2)
32	Fines (tph)	64	Coarse Rt8	96	Core state(3)

Table 22 – **X** variables for the physical with lags model

#	X variable ID	#	X variable ID	#	X variable ID	#	X variable ID	#	X variable ID	#	X variable ID
1	Coke real dens	36	-5 Green recyc (tph)	81	-4 MX1 KW max	126	-2 Inter. (tph)	171	-1 Agg pre-heater_1 current	216	Paste T after MX2	261	+2 Green recyc %	306	BB3 Flue 3 T set point
2	Coke Na	37	-5 Dry agg (tph)	82	-4 MX1 KW min	127	-2 Butts (tph)	172	-1 Agg pre-heater_2 current	217	+1 Coarse (tph)	262	+2 Fines rot valve speed	307	BF mean ext T
3	Coke Ca	38	-5 Paste (tph)	83	-4 MX1 therm oil T	128	-2 Pitch (tph)	173	-1 MX1 KW mean	218	+1 Fines (tph)	263	+2 Agg pre-heater T	308	BF min ext T
4	Coke S	39	-5 Coarse %	84	-4 MX1 P	129	-2 Green recyc (tph)	174	-1 MX1 KW max	219	+1 Inter. (tph)	264	+2 Agg pre-heater_1 current	309	BF max ext T
5	Coke V	40	-5 Fines %	85	-4 MX1 current	130	-2 Dry agg (tph)	175	-1 MX1 KW min	220	+1 Butts (tph)	265	+2 Agg pre-heater_2 current	310	BF ext T var
6	Coke 28/48 app dens	41	-5 Inter. %	86	-4 Paste T between MX	131	-2 Paste (tph)	176	-1 MX1 therm oil T	221	+1 Pitch (tph)	266	+2 MX1 KW mean	311	Core state (1)
7	Coke Fe	42	-5 Butts %	87	-4 MX2 KW mean pan 1	132	-2 Coarse %	177	-1 MX1 P	222	+1 Green recyc (tph)	267	+2 MX1 KW max	312	Core state (2)
8	Coke Si	43	-5 Pitch %	88	-4 MX2 KW mean pan 2	133	-2 Fines %	178	-1 MX1 current	223	+1 Dry agg (tph)	268	+2 MX1 KW min	313	Core state (3)
9	Coke Ni	44	-5 Green recyc %	89	-4 MX2 KW mean rotor	134	-2 Inter. %	179	-1 Paste T between MX	224	+1 Paste (tph)	269	+2 MX1 therm oil T		
10	Coke ash	45	-5 Fines rot valve speed	90	-4 MX2 paste weight	135	-2 Butts %	180	-1 MX2 KW mean pan 1	225	+1 Coarse %	270	+2 MX1 P		
11	Pitch SP	46	-5 Agg pre-heater T	91	-4 MX2 dump gate pos	136	-2 Pitch %	181	-1 MX2 KW mean pan 2	226	+1 Fines %	271	+2 MX1 current		
12	Pitch TS	47	-5 Agg pre-heater_1 current	92	-4 Paste T after MX2	137	-2 Green recyc %	182	-1 MX2 KW mean rotor	227	+1 Inter. %	272	+2 Paste T between MX		
13	Pitch Beta	48	-5 Agg pre-heater_2 current	93	-3 Coarse (tph)	138	-2 Fines rot valve speed	183	-1 MX2 paste weight	228	+1 Butts %	273	+2 MX2 KW mean pan 1		
14	Pitch QI	49	-5 MX1 KW mean	94	-3 Fines (tph)	139	-2 Agg pre-heater T	184	-1 MX2 dump gate pos	229	+1 Pitch %	274	+2 MX2 KW mean pan 2		
15	Pitch B/QI	50	-5 MX1 KW max	95	-3 Inter. (tph)	140	-2 Agg pre-heater_1 current	185	-1 Paste T after MX2	230	+1 Green recyc %	275	+2 MX2 KW mean rotor		
16	Pitch CV	51	-5 MX1 KW min	96	-3 Butts (tph)	141	-2 Agg pre-heater_2 current	186	Coarse (tph)	231	+1 Fines rot valve speed	276	+2 MX2 paste weight		
17	Pitch ash	52	-5 MX1 therm oil T	97	-3 Pitch (tph)	142	-2 MX1 KW mean	187	Fines (tph)	232	+1 Agg pre-heater T	277	+2 MX2 dump gate pos		
18	Pitch S	53	-5 MX1 P	98	-3 Green recyc (tph)	143	-2 MX1 KW max	188	Inter. (tph)	233	+1 Agg pre-heater_1 current	278	+2 Paste T after MX2		
19	Pitch dist	54	-5 MX1 current	99	-3 Dry agg (tph)	144	-2 MX1 KW min	189	Butts (tph)	234	+1 Agg pre-heater_2 current	279	Butts Rt3/8+Rt4		
20	Butts Al	55	-5 Paste T between MX	100	-3 Paste (tph)	145	-2 MX1 therm oil T	190	Pitch (tph)	235	+1 MX1 KW mean	280	Coarse Rt4		
21	Butts Ca	56	-5 MX2 KW mean pan 1	101	-3 Coarse %	146	-2 MX1 P	191	Green recyc (tph)	236	+1 MX1 KW max	281	Coarse Rt8		
22	Butts %F	57	-5 MX2 KW mean pan 2	102	-3 Fines %	147	-2 MX1 current	192	Dry agg (tph)	237	+1 MX1 KW min	282	Inter Rt50+Rt100		
23	Butts ash	58	-5 MX2 KW mean rotor	103	-3 Inter. %	148	-2 Paste T between MX	193	Paste (tph)	238	+1 MX1 therm oil T	283	Fines Pt200		
24	Butts Fe	59	-5 MX2 paste weight	104	-3 Butts %	149	-2 MX2 KW mean pan 1	194	Coarse %	239	+1 MX1 P	284	Agg Rt3/8		
25	Butts Ni	60	-5 MX2 dump gate pos	105	-3 Pitch %	150	-2 MX2 KW mean pan 2	195	Fines %	240	+1 MX1 current	285	Agg Rt4@Rt30		
26	Butts Si	61	-5 Paste T after MX2	106	-3 Green recyc %	151	-2 MX2 KW mean rotor	196	Inter. %	241	+1 Paste T between MX	286	Agg Rt50+Rt100		
27	Butts Na	62	-4 Coarse (tph)	107	-3 Fines rot valve speed	152	-2 MX2 paste weight	197	Butts %	242	+1 MX2 KW mean pan 1	287	Agg Rt200+Pt200		
28	Butts S	63	-4 Fines (tph)	108	-3 Agg pre-heater T	153	-2 MX2 dump gate pos	198	Pitch %	243	+1 MX2 KW mean pan 2	288	Agg Pt200		
29	Butts V	64	-4 Inter. (tph)	109	-3 Agg pre-heater_1 current	154	-2 Paste T after MX2	199	Green recyc %	244	+1 MX2 KW mean rotor	289	Anode type (dim)(20)		
30	Butts Na/Ca	65	-4 Butts (tph)	110	-3 Agg pre-heater_2 current	155	-1 Coarse (tph)	200	Fines rot valve speed	245	+1 MX2 paste weight	290	Anode type (dim)(23)		
31	-5 Coarse (tph)	66	-4 Pitch (tph)	111	-3 MX1 KW mean	156	-1 Fines (tph)	201	Agg pre-heater T	246	+1 MX2 dump gate pos	291	Anode type (dim)(26)		
32	-5 Fines (tph)	67	-4 Green recyc (tph)	112	-3 MX1 KW max	157	-1 Inter. (tph)	202	Agg pre-heater_1 current	247	+1 Paste T after MX2	292	Green anode height		
33	-5 Inter. (tph)	68	-4 Dry agg (tph)	113	-3 MX1 KW min	158	-1 Butts (tph)	203	Agg pre-heater_2 current	248	+2 Coarse (tph)	293	VC bellows P		
34	-5 Butts (tph)	69	-4 Paste (tph)	114	-3 MX1 therm oil T	159	-1 Pitch (tph)	204	MX1 KW mean	249	+2 Fines (tph)	294	PP mean ext T		
35	-5 Pitch (tph)	70	-4 Coarse %	115	-3 MX1 P	160	-1 Green recyc (tph)	205	MX1 KW max	250	+2 Inter. (tph)	295	Oven		
36	-5 Green recyc (tph)	71	-4 Fines %	116	-3 MX1 current	161	-1 Dry agg (tph)	206	MX1 KW min	251	+2 Butts (tph)	296	Fire		
37	-5 Dry agg (tph)	72	-4 Inter. %	117	-3 Paste T between MX	162	-1 Paste (tph)	207	MX1 therm oil T	252	+2 Pitch (tph)	297	Pit position		
38	-5 Paste (tph)	73	-4 Butts %	118	-3 MX2 KW mean pan 1	163	-1 Coarse %	208	MX1 P	253	+2 Green recyc (tph)	298	Fire cycle T		
39	-5 Coarse %	74	-4 Pitch %	119	-3 MX2 KW mean pan 2	164	-1 Fines %	209	MX1 current	254	+2 Dry agg (tph)	299	BF pit starting T		
40	-5 Fines %	75	-4 Green recyc %	120	-3 MX2 KW mean rotor	165	-1 Inter. %	210	Paste T between MX	255	+2 Paste (tph)	300	BB1 pit max T pos A		
41	-5 Inter. %	76	-4 Fines rot valve speed	121	-3 MX2 paste weight	166	-1 Butts %	211	MX2 KW mean pan 1	256	+2 Coarse %	301	BB2 pit max T pos A		
42	-5 Butts %	77	-4 Agg pre-heater T	122	-3 MX2 dump gate pos	167	-1 Pitch %	212	MX2 KW mean pan 2	257	+2 Fines %	302	BB3 pit max T pos A		
43	-5 Pitch %	78	-4 Agg pre-heater_1 current	123	-3 Paste T after MX2	168	-1 Green recyc %	213	MX2 KW mean rotor	258	+2 Inter. %	303	BB3 pit max T pos B		
44	-5 Green recyc %	79	-4 Agg pre-heater_2 current	124	-2 Coarse (tph)	169	-1 Fines rot valve speed	214	MX2 paste weight	259	+2 Butts %	304	BB1 flue 3 max T		
45	-5 Fines rot valve speed	80	-4 MX1 KW mean	125	-2 Fines (tph)	170	-1 Agg pre-heater T	215	MX2 dump gate pos	260	+2 Pitch %	305	BB3 flue 3 max T		

Table 23 – **X** variables for the physical model with baking profile

#	X variable ID	#	X variable ID	#	X variable ID	#	X variable ID	#	X variable ID
1	Coke real dens	36	Green recyc (tph)	71	Agg Pt200	106	BB2 A T 2	141	BB2 flue T 2
2	Coke Na	37	Dry agg (tph)	72	Anode type (dim)(20)	107	BB2 A T R2	142	BB2 flue T R2
3	Coke Ca	38	Paste (tph)	73	Anode type (dim)(23)	108	BB2 A T min	143	BB2 flue T min
4	Coke S	39	Coarse %	74	Anode type (dim)(26)	109	BB2 A T max	144	BB2 flue T max
5	Coke V	40	Fines %	75	Green anode height	110	BB2 A T mean	145	BB2 flue T mean
6	Coke 28/48 app dens	41	Inter. %	76	VC bellows P	111	BB3 A T 0	146	BB3 flue T 0
7	Coke Fe	42	Butts %	77	PP mean ext T	112	BB3 A T 1	147	BB3 flue T 1
8	Coke Si	43	Pitch %	78	Oven	113	BB3 A T 2	148	BB3 flue T 2
9	Coke Ni	44	Green recyc %	79	Fire	114	BB3 A T R2	149	BB3 flue T R2
10	Coke ash	45	Fines rot valve speed	80	Pit position	115	BB3 A T min	150	BB3 flue T min
11	Pitch SP	46	Agg pre-heater T	81	Fire cycle T	116	BB3 A T max	151	BB3 flue T max
12	Pitch TS	47	Agg pre-heater_1 current	82	BF pit starting T	117	BB3 A T mean	152	BB3 flue T mean
13	Pitch Beta	48	Agg pre-heater_2 current	83	BB1 pit max T pos A	118	BB3 B T 0	153	UPB P 0
14	Pitch QI	49	MX1 KW mean	84	BB2 pit max T pos A	119	BB3 B T 1	154	UPB P 1
15	Pitch B/QI	50	MX1 KW max	85	BB3 pit max T pos A	120	BB3 B T 2	155	UPB P R2
16	Pitch CV	51	MX1 KW min	86	BB3 pit max T pos B	121	BB3 B T R2	156	UPB P min
17	Pitch ash	52	MX1 therm oil T	87	BB1 flue 3 max T	122	BB3 B T min	157	UPB P max
18	Pitch S	53	MX1 P	88	BB3 flue 3 max T	123	BB3 B T max	158	UPB P mean
19	Pitch dist	54	MX1 current	89	BB3 flue 3 T set point	124	BB3 B T mean	159	UPB P var
20	Butts Al	55	Paste T between MX	90	BF mean ext T	125	UPB flue T 0		
21	Butts Ca	56	MX2 KW mean pan 1	91	BF min ext T	126	UPB flue T 1		
22	Butts %F	57	MX2 KW mean pan 2	92	BF max ext T	127	UPB flue T 2		
23	Butts ash	58	MX2 KW mean rotor	93	BF ext T var	128	UPB flue T R2		
24	Butts Fe	59	MX2 paste weight	94	Core state(1)	129	UPB flue T min		
25	Butts Ni	60	MX2 dump gate pos	95	Core state(2)	130	UPB flue T max		
26	Butts Si	61	Paste T after MX2	96	Core state(3)	131	UPB flue T mean		
27	Butts Na	62	Butts Rt3/8+Rt4	97	BB1 A T 0	132	BB1 flue T 0		
28	Butts S	63	Coarse Rt4	98	BB1 A T 1	133	BB1 flue T 1		
29	Butts V	64	Coarse Rt8	99	BB1 A T 2	134	BB1 flue T 2		
30	Butts Na/Ca	65	Inter Rt50+Rt100	100	BB1 A T R2	135	BB1 flue T R2		
31	Coarse (tph)	66	Fines Pt200	101	BB1 A T min	136	BB1 flue T min		
32	Fines (tph)	67	Agg Rt3/8	102	BB1 A T max	137	BB1 flue T max		
33	Inter. (tph)	68	Agg Rt4@Rt30	103	BB1 A T mean	138	BB1 flue T mean		
34	Butts (tph)	69	Agg Rt50+Rt100	104	BB2 A T 0	139	BB2 flue T 0		
35	Pitch (tph)	70	Agg Rt200+Pt200	105	BB2 A T 1	140	BB2 flue T 1		

Table 24 – **Y** variables for all the physical property models

Y variable ID
GAD Green anode dens
Green weight
Baked weight (mean)
Thermal cond
App dens
Real dens
Comp strengh
L c
Young's mod
Elect resis

Table 25 – **X** variables for the reactivity models

#	X variable ID	#	X variable ID	#	X variable ID
1	Coke real dens	32	Fines (tph)	63	Coarse Rt4
2	Coke Na	33	Inter. (tph)	64	Coarse Rt8
3	Coke Ca	34	Butts (tph)	65	Inter Rt50+Rt100
4	Coke S	35	Pitch (tph)	66	Fines Pt200
5	Coke V	36	Green recyc (tph)	67	Agg Rt3/8
6	Coke 28/48 app dens	37	Dry agg (tph)	68	Agg Rt4@Rt30
7	Coke Fe	38	Paste (tph)	69	Agg Rt50+Rt100
8	Coke Si	39	Coarse %	70	Agg Rt200+Pt200
9	Coke Ni	40	Fines %	71	Agg Pt200
10	Coke ash	41	Inter. %	72	Anode type (dim)(20)
11	Pitch SP	42	Butts %	73	Anode type (dim)(23)
12	Pitch TS	43	Pitch %	74	Anode type (dim)(26)
13	Pitch Beta	44	Green recyc %	75	Green anode height
14	Pitch QI	45	Fines rot valve speed	76	VC bellows P
15	Pitch B/QI	46	Agg pre-heater T	77	PP mean ext T
16	Pitch CV	47	Agg pre-heater_1 current	78	Oven
17	Pitch ash	48	Agg pre-heater_2 current	79	Fire
18	Pitch S	49	MX1 KW mean	80	Pit position
19	Pitch dist	50	MX1 KW max	81	Fire cycle T
20	Butts Al	51	MX1 KW min	82	BF pit starting T
21	Butts Ca	52	MX1 therm oil T	83	BB1 pit max T pos A
22	Butts %F	53	MX1 P	84	BB2 pit max T pos A
23	Butts ash	54	MX1 current	85	BB3 pit max T pos A
24	Butts Fe	55	Paste T between MX	86	BB3 pit max T pos B
25	Butts Ni	56	MX2 KW mean pan 1	87	BB1 flue 3 max T
26	Butts Si	57	MX2 KW mean pan 2	88	BB3 flue 3 max T
27	Butts Na	58	MX2 KW mean rotor	89	BB3 flue 3 T set point
28	Butts S	59	MX2 paste weight	90	BF mean ext T
29	Butts V	60	MX2 dump gate pos	91	BF min ext T
30	Butts Na/Ca	61	Paste T after MX2	92	BF max ext T
31	Coarse (tph)	62	Butts Rt3/8+Rt4	93	Bfext T var

Table 26 – **Y** variables for the reactivity models

Overall model Y variable ID	**Air residue model Y variable ID**	**CO_2 residue model Y variable ID**
CRD CO2 dust	ARR Air residue	CRR CO2 residue
CRL CO2 lost		
CRR CO2 residue		
ARD Air dust		
ARL Air lost		
ARR Air residue		

Appendix B Models VIP list

Table 27 – VIP for the physical transport delay model

Variable ID	VIP	Variable ID	VIP	Variable ID	VIP	Variable ID	VIP
Pit position	2,850	Inter. %	0,982	Butts Al	0,835	MX1 current	0,404
Green anode height	2,291	MX2 KW mean pan 2	0,978	Butts Na	0,819	Oven	0,397
Butts V	1,544	Green recyc %	0,968	BF pit starting T	0,818	Fire	0,233
VC bellows P	1,470	Butts Ni	0,967	Butts Ca	0,807		
Pitch ash	1,456	Coke Si	0,965	Inter Rt50+Rt100	0,796		
Coke V	1,425	Butts (tph)	0,962	Fire cycle T	0,759		
Pitch QI	1,355	Core state(2)	0,962	Coke Ni	0,759		
Anode type (dim)(20)	1,313	Coke ash	0,953	BF ext T var	0,754		
Pitch B/QI	1,312	MX2 KW mean pan 1	0,944	MX1 KW max	0,739		
Coke 28/48 app dens	1,273	Coarse Rt4	0,944	Butts Rt3/8+Rt4	0,727		
BF max ext T	1,239	Core state(1)	0,944	BB1 pit max T pos A	0,699		
Butts S	1,239	Coke Ca	0,942	Paste T after MX2	0,690		
Pitch Beta	1,215	Butts %	0,940	Agg Rt3/8	0,658		
BF mean ext T	1,214	Green recyc (tph)	0,938	BB2 pit max T pos A	0,655		
MX1 P	1,197	Coarse (tph)	0,937	Pitch dist	0,645		
MX2 dump gate pos	1,173	Agg Rt4@Rt30	0,934	MX2 KW mean rotor	0,638		
Pitch %	1,173	Agg Rt50+Rt100	0,920	Butts ash	0,637		
BF min ext T	1,168	Paste (tph)	0,902	Coarse Rt8	0,623		
BB3 pit max T pos B	1,166	Coarse %	0,900	Agg Rt200+Pt200	0,621		
Pitch (tph)	1,153	Fines rot valve speed	0,898	Fines Pt200	0,617		
Pitch CV	1,137	Coke Na	0,895	BB3 pit max T pos A	0,593		
Pitch SP	1,129	MX2 paste weight	0,895	Agg pre-heater_1 current	0,568		
Coke S	1,126	Butts %F	0,894	Butts Si	0,558		
Coke Fe	1,113	Fines %	0,888	Agg Pt200	0,533		
MX1 KW min	1,102	Inter. (tph)	0,881	MX1 therm oil T	0,511		
Anode type (dim)(26)	1,081	Fines (tph)	0,879	BB3 flue 3 max T	0,498		
Anode type (dim)(23)	1,063	Butts Fe	0,879	Core state(3)	0,495		
PP mean ext T	1,057	Butts Na/Ca	0,873	BB1 flue 3 max T	0,488		
Dry agg (tph)	1,055	Agg pre-heater_2 current	0,861	MX1 KW mean	0,460		
Pitch TS	1,026	Paste T between MX	0,851	Pitch S	0,454		
Coke real dens	1,012	Agg pre-heater T	0,848	BB3 flue 3 T set point	0,452		

Table 28 – VIP for the overall reactivity model

Variable ID	VIP	Variable ID	VIP	Variable ID	VIP
Pit position	3,060	PP mean ext T	1,000	Coke Na	0,623
Butts V	1,797	Paste T between MX	0,982	BB3 flue 3 max T	0,621
Coke 28/48 app dens	1,605	BF pit starting T	0,977	Coke ash	0,618
Pitch B/QI	1,578	Inter Rt50+Rt100	0,974	Butts Fe	0,601
Coke V	1,544	Inter. (tph)	0,957	BB2 pit max T pos A	0,601
Pitch QI	1,523	Butts Si	0,953	Paste T after MX2	0,596
Pitch Beta	1,511	Fines rot valve speed	0,939	Dry agg (tph)	0,593
Pitch SP	1,509	Anode type (dim)(23)	0,921	Fines (tph)	0,592
VC bellows P	1,499	Butts Al	0,916	Butts ash	0,591
MX2 KW mean pan 1	1,491	MX2 paste weight	0,912	BB3 flue 3 T set point	0,587
Pitch ash	1,449	Butts %F	0,902	Green recyc (tph)	0,580
MX2 dump gate pos	1,415	Coke real dens	0,883	Fire cycle T	0,560
Butts S	1,400	Inter. %	0,881	Agg pre-heater_1 current	0,548
Pitch %	1,378	Agg pre-heater_2 current	0,879	Green recyc %	0,530
Butts Na/Ca	1,347	BF min ext T	0,872	Agg Rt3/8	0,524
Pitch (tph)	1,344	BB3 pit max T pos B	0,861	Oven	0,485
Coke S	1,330	Coke Si	0,843	Agg Pt200	0,428
MX1 P	1,298	Paste (tph)	0,842	Fines Pt200	0,426
Coke Fe	1,270	MX1 current	0,818	Coarse %	0,422
Anode type (dim)(26)	1,219	Pitch S	0,794	BB1 flue 3 max T	0,401
Anode type (dim)(20)	1,217	MX1 KW min	0,774	Butts (tph)	0,385
MX1 KW max	1,171	Butts Rt3/8+Rt4	0,752	Butts %	0,384
Coke Ca	1,156	Butts Ni	0,746	Agg Rt50+Rt100	0,321
Bfext T var	1,131	Butts Ca	0,717	Coarse Rt4	0,287
BF max ext T	1,116	BB3 pit max T pos A	0,702	Coarse (tph)	0,271
BF mean ext T	1,100	Pitch TS	0,698	Pitch dist	0,263
Pitch CV	1,079	BB1 pit max T pos A	0,697	Fire	0,192
Butts Na	1,045	MX2 KW mean pan 2	0,670	Green anode height	0,173
Coke Ni	1,040	Coarse Rt8	0,666	MX1 KW mean	0,160
MX1 therm oil T	1,037	MX2 KW mean rotor	0,660		
Agg pre-heater T	1,013	Agg Rt4@Rt30	0,643		
Fines %	1,011	Agg Rt200+Pt200	0,632		