

**UNIVERSITY OF BAHRAIN**  
**College of Engineering**  
**Department of Mechanical Engineering**



# **Improving Anode Quality Based on Process Parameters in Aluminum Industry**

A Case Study from the Kingdom of Bahrain

A Thesis Submitted in Partial Fulfilment of the Requirements for the  
Master's Degree in Engineering Management

**Submitted by:**

**Ahmed Mohamed Yusuf Mashkoor**

ID number: 20132683

**Supervised by:**

**Supervisor**

**Dr. Osama Al Jamal**

Assistant professor

**Co-Supervisor**

**Dr. Yaser Al Alawi**

Assistant professor

**Kingdom of Bahrain**

**February 2025**



## Thesis Discussion Committee Decision

Form Code: UDNGS013

|                 |                                                                          |
|-----------------|--------------------------------------------------------------------------|
| Student Name:   | Ahmed Mohamed Yusuf Mashkoor                                             |
| Student ID. No. | 20132683                                                                 |
| Department:     | Mechanical Engineering                                                   |
| College:        | Engineering                                                              |
| Program:        | Masters in Engineering Management                                        |
| Student GPA:    | 4.00                                                                     |
| Thesis Title:   | Improving Anode Quality Based on Process Parameter in Aluminium Industry |

The Discussion Committee met on 11.02.2025 in 14-162 and MS Teams upon discussing the student whose details are stated above in public. The discussion committee recommends the following:

The student is awarded the grade PASS in the thesis course IEN598.

☐ Without making any changes to the thesis.

☒ After making the modifications attached with this decision.

### Discussion Committee Members Approval:

|    |                    |                                          |        |  |
|----|--------------------|------------------------------------------|--------|--|
| 1. | Supervisor Name:   | Dr. Osama Al-Jamal<br>Dr. Yaser Al-Alawi | Sign.: |  |
| 2. | Internal Examiner: | Prof. Ghulam Hussain                     | Sign.: |  |
| 3. | External Examiner: | Prof. Mohamed Gadallah                   | Sign.: |  |

# Abstract

Carbon anode production is crucial in aluminum production and accounts for a high percentage of production costs. Therefore, anode quality improvement is essential for cost reduction and annual production increase. Raw material properties and process parameters highly influence carbon anode quality. The problem to be investigated in this research study is the unavailability of a well-known mathematical relationship or model to control the quality of anodes produced. The main objectives of the study are to develop a mathematical model of anode quality based on its process parameters, obtain the significant process parameters determining the anode quality, investigate the main causes of variation in anode quality, and define the relationship between anode quality and process parameters. The objectives of the study were achieved by using a multivariate statistical (latent variable modeling) method to build a mathematical model between anode quality properties and process parameters. The model built is used to predict a new dataset to validate prediction ability. Different interpretation methods are used to obtain significant process parameters and investigate the main causes of variation in anode quality. The data collected to achieve these objectives are actual anode quality properties, process parameters, and raw material properties data extracted from Carbon Plant 2 of Aluminium Bahrain (ALBA). Two projection to latent structure (PLS) models have been built using SIMCA software to improve anode physical properties, and chemical and reactivity properties. The research provides good anode quality models with 64.0% and 64.1% variances for the physical properties model and chemical and reactivity model, respectively. The model's prediction ability is 51.1% and 57.2% variances for the first and second models, respectively. Also, the primary source of variation in anode quality was mainly the variability of raw material properties, precisely calcined petroleum coke (CPC) properties. Moreover, the research illustrates that the most influential variable for the first model was green anode density (GAD), and for the second model, was CPC iron content and paste plant process parameters. Furthermore, the most important variables were green anode density and fire soaking time for the first and second models, respectively. The study estimated the net carbon consumption (NCC) reduction using the model built and found it to be 1.65 kg per ton of aluminum. Accordingly, the estimated annual cost reduction is \$825,000.

## **Dedication**

I would like to dedicate this work to my parents, family, friends, and everyone who supported me during all stages of the study period.

Thank you for your continuous love, support, and motivation throughout this journey.

## **Acknowledgment**

I would like to thank all those who contributed to this work, which otherwise would not have been achieved or completed. Especially my family, who supported me with everything in their hands.

Also, I would like to express my greatest thanks and appreciation to my supervisor, Dr. Osama Al Jamal, and co-supervisor, Dr. Yaser Al Alawi, for their unlimited support and guidance throughout the study.

Moreover, I would like to thank ALBA's top management as they were supportive in terms of the collection and extraction of needed study data. My appreciation also goes to IT specialist Mrs. Hayma Rahma for her support in data collection and Carbon Process Control specialist Mr. Bhoopesh Jain for his experience helping to synchronize and track the collected data during the research study.

# List of Content

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| <b>Abstract.....</b>                                                | <b>A</b>  |
| <b>Dedication .....</b>                                             | <b>B</b>  |
| <b>Acknowledgment.....</b>                                          | <b>C</b>  |
| <b>List of Content.....</b>                                         | <b>D</b>  |
| <b>List of Tables .....</b>                                         | <b>H</b>  |
| <b>List of Figures.....</b>                                         | <b>K</b>  |
| <b>List of Notations .....</b>                                      | <b>N</b>  |
| <b>List of Abbreviations .....</b>                                  | <b>Q</b>  |
| <b>Chapter 1: Introduction to the Study .....</b>                   | <b>1</b>  |
| <b>1.1 Introduction.....</b>                                        | <b>2</b>  |
| <b>1.2 Problem Statement .....</b>                                  | <b>5</b>  |
| <b>1.3 Research Questions.....</b>                                  | <b>6</b>  |
| <b>1.4 Research Objectives.....</b>                                 | <b>6</b>  |
| <b>1.5 Research Significance.....</b>                               | <b>6</b>  |
| <b>1.6 The Study’s Setting.....</b>                                 | <b>7</b>  |
| <b>1.7 Research Terminology.....</b>                                | <b>7</b>  |
| <b>Chapter 2: Theoretical Framework and Literature Review .....</b> | <b>9</b>  |
| <b>2.1 Introduction.....</b>                                        | <b>10</b> |
| <b>2.2 Theoretical Overview .....</b>                               | <b>10</b> |
| <b>2.2.1 Raw Materials Parameters .....</b>                         | <b>11</b> |
| <b>2.2.2 Paste Plant Process .....</b>                              | <b>13</b> |

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| 2.2.3 Baking Kiln Process .....                                            | 18 |
| 2.2.4 Anode Quality Measurement .....                                      | 23 |
| 2.3 Previous Studies on Carbon Plant Process Control .....                 | 30 |
| 2.3.1 Properties and Parameters Affecting Anode Quality .....              | 30 |
| 2.3.2 Anode Quality Models.....                                            | 34 |
| 2.4 Chapter Conclusion .....                                               | 37 |
| Chapter 3: Methods and Procedures .....                                    | 39 |
| 3.1 Introduction.....                                                      | 40 |
| 3.2 Evaluation of Prediction Model of Anode Quality Results .....          | 40 |
| 3.3 Multivariate Statistical Methods.....                                  | 42 |
| 3.3.1 Principal Component Analysis (PCA) .....                             | 43 |
| 3.3.2 Projection to Latent Structure (PLS) .....                           | 47 |
| 3.3.3 Cross-Validation to Estimate the Number of Latent Variables<br>..... | 53 |
| 3.3.4 Statistical Methods Analysis of the Model Built.....                 | 54 |
| 3.3.5 Prediction of New Dataset Using PLS Model Built.....                 | 55 |
| 3.3.6 Model Interpretation Tools .....                                     | 57 |
| 3.3.7 Net Carbon Consumption Reduction .....                               | 61 |
| 3.4 Research Tools .....                                                   | 62 |
| 3.5 Data Collection.....                                                   | 62 |
| 3.5.1 Anode Quality Results Data .....                                     | 64 |
| 3.5.2 Paste Plant Process Parameters Data .....                            | 65 |

|                                                                                       |     |
|---------------------------------------------------------------------------------------|-----|
| 3.5.3 Raw Material Properties Data .....                                              | 65  |
| 3.5.4 Baking Kiln Process Parameters Data .....                                       | 66  |
| 3.6 Chapter Conclusion .....                                                          | 68  |
| Chapter 4: Results and Discussion .....                                               | 70  |
| 4.1 Introduction.....                                                                 | 71  |
| 4.2 Discrete and Continuous Independent Variables.....                                | 71  |
| 4.3 Physical Properties Model .....                                                   | 73  |
| 4.3.1 Discrete Independent Variables of the Physical Properties Model .....           | 73  |
| 4.3.2 Exclusion of Non-contributory Independent Variables.....                        | 77  |
| 4.3.3 DmodX and Hotelling's ( $T^2$ ) plots.....                                      | 77  |
| 4.3.4 Physical Properties Model Main Findings .....                                   | 79  |
| 4.3.5 Test the Significance of the Independent Variables of Model 1 .....             | 85  |
| 4.3.6 Physical Properties Model Interpretation.....                                   | 89  |
| 4.3.7 Prediction of New Physical Properties Dataset .....                             | 98  |
| 4.4 Chemical and Reactivity Properties Model .....                                    | 101 |
| 4.4.1 Discrete Independent Variables of Chemical and Reactivity Properties Model..... | 102 |
| 4.4.2 Chemical and Reactivity Model Preparation .....                                 | 103 |
| 4.4.3 Chemical and Reactivity Model Main Findings .....                               | 106 |
| 4.4.4 Test the Significance of the Independent Variables of Model 2 .....             | 110 |

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| 4.4.5 Chemical and Reactivity Properties Model Interpretation...            | 113 |
| 4.4.6 Prediction of New Chemical and Reactivity Properties Dataset<br>..... | 120 |
| 4.5 Net Carbon Consumption of PLS Model Built.....                          | 122 |
| 4.6 Chapter Conclusion.....                                                 | 123 |
| Chapter 5: Conclusion and Recommendations.....                              | 124 |
| 5.1 Conclusion .....                                                        | 125 |
| 5.2 Recommendations.....                                                    | 127 |
| List of References .....                                                    | 129 |
| List of Appendices.....                                                     | 137 |

## List of Tables

|                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 1: CPC chemical and physical properties .....                                                                                                             | 11 |
| Table 2: Liquid tar pitch chemical and physical properties.....                                                                                                 | 12 |
| Table 3: Anode butts main properties.....                                                                                                                       | 13 |
| Table 4: Paste plant main process parameters .....                                                                                                              | 18 |
| Table 5: Baking kiln main process parameters.....                                                                                                               | 23 |
| Table 6: Physical and chemical anode properties.....                                                                                                            | 30 |
| Table 7: Regression model variances' titles and descriptions.....                                                                                               | 42 |
| Table 8: Physical and chemical models' dependent variables .....                                                                                                | 67 |
| Table 9: Physical and chemical models' independent variables .....                                                                                              | 67 |
| Table 10: Independent variables type: Discrete or Continuous variables ....                                                                                     | 72 |
| Table 11: Pitch level and green scrap content variables values and represented levels of the physical properties model .....                                    | 74 |
| Table 12: Butts, fine and medium particles percentage variables values and represented levels of the physical properties model.....                             | 75 |
| Table 13: Coarse particles percentage, anode forming speed and anode forming time variables values and represented levels of the physical properties model..... | 76 |
| Table 14: Fire cycle time and fire soaking time variables values and represented levels of the physical properties model.....                                   | 76 |
| Table 15: Cooling power and mixing temperature relationship strength with dependent variables.....                                                              | 77 |
| Table 16: Summary of physical properties model results .....                                                                                                    | 80 |
| Table 17: Model 1 - Physical properties model mathematical equation coefficients.....                                                                           | 83 |

|                                                                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 18: Geometric Density ( $y_1$ ) - test the significance of x-variables with near-zero coefficients.....                                                    | 86  |
| Table 19: Specific Electrical Resistance ( $y_2$ ) - test the significance of x-variables with near-zero coefficients .....                                      | 86  |
| Table 20: Flexural Strength ( $y_3$ ) - test the significance of x-variables with near-zero coefficients.....                                                    | 87  |
| Table 21: Raw material blend and supplier sources of physical properties model.....                                                                              | 91  |
| Table 22: VIP values for X-variables of the physical properties model .....                                                                                      | 98  |
| Table 23: Observed versus predicted dependent variables variance values of the physical properties model .....                                                   | 99  |
| Table 24: Fine, Coarse Particles Percentage, and Anode Forming Time variables values and represented levels of the chemical and reactivity properties model..... | 103 |
| Table 25: Cooling power relationship strength with dependent variables in the chemical and reactivity model .....                                                | 104 |
| Table 26: Summary of the chemical and reactivity properties model.....                                                                                           | 106 |
| Table 27: Model 2 - Chemical and reactivity properties model mathematical equation coefficients .....                                                            | 108 |
| Table 28: CO <sub>2</sub> Reactivity Residue ( $y_4$ ) - test the significance of x-variables with near-zero coefficients.....                                   | 111 |
| Table 29: Sulfur (S) Element Content ( $y_5$ ) - test the significance of x-variables with near-zero coefficients.....                                           | 112 |
| Table 30: Iron (Fe) Element Content ( $y_6$ ) - test the significance of x-variables with near-zero coefficients.....                                            | 112 |
| Table 31: VIP values for X-variables of the chemical and reactivity properties model.....                                                                        | 119 |

|                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 32: Observed versus predicted dependent variables variance values of the chemical and reactivity properties model ..... | 120 |
|-------------------------------------------------------------------------------------------------------------------------------|-----|

## List of Figures

|                                                                                                                              |    |
|------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1: Prebaked electrolytic cell .....                                                                                   | 3  |
| Figure 2: Carbon plant process flowchart.....                                                                                | 4  |
| Figure 3: Paste plant process flowchart .....                                                                                | 14 |
| Figure 4: Fire group in the baking kiln.....                                                                                 | 21 |
| Figure 5: Schematic representation of the PCA model .....                                                                    | 44 |
| Figure 6: NIPALS algorithm for PCA method: (a) Step 2, (b) Step 4 .....                                                      | 45 |
| Figure 7: Schematic representation of the PLS model .....                                                                    | 49 |
| Figure 8: NIPALS algorithm for PLS method.....                                                                               | 50 |
| Figure 9: Anode production number and sample ID number breakdown ....                                                        | 63 |
| Figure 10: DmodX plot of physical properties model.....                                                                      | 78 |
| Figure 11: Hotelling's ( $T^2$ ) plot of physical properties model .....                                                     | 79 |
| Figure 12: Independent variables variance ( $R_y^2$ ) explained by model versus<br>the number of latent variables used ..... | 81 |
| Figure 13: Cross-validation variance ( $Q^2$ ) explained by model versus the<br>number of latent variables used .....        | 82 |
| Figure 14: The 2D score plot for the first two latent variables of the physical<br>properties model.....                     | 90 |
| Figure 15: The score contribution plot for the change from Group 1 to Group<br>2.....                                        | 92 |
| Figure 16: The score contribution plot for the change from Group 4 to Group<br>5.....                                        | 93 |
| Figure 17: The 3D score plot for the first two latent variables of the physical<br>properties model.....                     | 94 |
| Figure 18: Loading plot of the first two latent variables of the physical<br>properties model.....                           | 95 |

|                                                                                                                                                                      |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 19: Loading plot of the physical properties model's latent variables 1 and 3.....                                                                             | 97  |
| Figure 20: Comparison between observed versus predicted Geometric Density ( $y_1$ ).....                                                                             | 100 |
| Figure 21: Comparison between observed versus predicted Specific Electrical Resistance ( $y_2$ ) .....                                                               | 100 |
| Figure 22: Comparison between observed versus predicted Flexural Strength ( $y_3$ ).....                                                                             | 101 |
| Figure 23: DmodX plot of the chemical and reactivity model .....                                                                                                     | 105 |
| Figure 24: Hotelling's ( $T^2$ ) plot of the chemical and reactivity model .....                                                                                     | 105 |
| Figure 25: Independent variables variance ( $R_y^2$ ) explained by model versus number of latent variables used of the chemical and reactivity properties model..... | 107 |
| Figure 26: Cross-validation variance ( $Q^2$ ) explained by model versus number of latent variables used of the chemical and reactivity properties model...          | 108 |
| Figure 27: The 2D score plot for the first two latent variables of the chemical and reactivity properties model.....                                                 | 114 |
| Figure 28: The 3D score plot for the first two latent variables of the chemical and reactivity properties model.....                                                 | 115 |
| Figure 29: Loading plot of the first two latent variables of the chemical and reactivity properties model .....                                                      | 117 |
| Figure 30: Loading plot of the latent variables 1 and 3 of the chemical and reactivity properties model .....                                                        | 118 |
| Figure 31: Comparison between observed versus predicted CO <sub>2</sub> reactivity residue ( $y_4$ ).....                                                            | 121 |
| Figure 32: Comparison between observed versus predicted sulfur (S) element content ( $y_5$ ).....                                                                    | 121 |

|                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------|-----|
| Figure 33: Comparison between observed versus predicted iron (Fe) element<br>content ( $y_6$ )..... | 122 |
|-----------------------------------------------------------------------------------------------------|-----|

## List of Notations

| Symbol          | Definition                                             |
|-----------------|--------------------------------------------------------|
| A               | Number of latent variables                             |
| $\alpha$        | Net carbon consumption empirical equation constant     |
| $\beta_0$       | Regression intercept of multiple linear regression     |
| $\beta_i$       | Regression coefficients of multiple linear regression  |
| $c_a$           | Loading vector of Y matrix of PLS model                |
| $c_a^T$         | Transpose loading vector of Y matrix of PLS model      |
| $C'$            | Loading matrix of Y of PLS model                       |
| df              | Degree of freedom                                      |
| $DmodX_i(a)$    | Residual of an observation in a latent variables space |
| $\varepsilon_i$ | Random error of multiple linear regression             |
| $E_a$           | Residual matrix of X                                   |
| $e_{ij}(a)$     | Residual of an observation of variable j               |
| $F_a$           | Residual matrix of Y                                   |
| $F_{(1)}$       | Residual matrix of group i of Y matrix                 |
| H0              | Null hypothesis                                        |
| H1              | Alternative hypothesis                                 |
| k               | Number of X variables                                  |
| M               | Number of Y variables                                  |
| N               | Number of observations                                 |
| P               | Loadings matrix of PCA model                           |
| $P_i'$          | Loadings vector of X matrix                            |
| $P_a^m$         | Magnitude of $P_a$ vector                              |
| $P_a$           | Loading vector of X matrix                             |

|                 |                                                                           |
|-----------------|---------------------------------------------------------------------------|
| $P_a^T$         | Transpose loading vector of X matrix                                      |
| $Q^2$           | Cross-validation dependent variables variance (cross-validation variance) |
| R               | Matrix of $r_a$ loadings vector                                           |
| $R_x^2$         | Independent variable variance                                             |
| $R_y^2$         | Dependent variables variance (model variance)                             |
| $\rho_a$        | Anode geometric density                                                   |
| $\rho_{ER}$     | Electrical resistivity                                                    |
| $SE_\beta$      | Standard error                                                            |
| $s_{ta}$        | Variance of $t_a$ vector                                                  |
| SSE             | Sum of Squares of Errors                                                  |
| $SSY(a)$        | Sum of squares of Y explained by the $a^{th}$ latent variable             |
| $SSY(tot)$      | Sum of squares of Y explained by the model                                |
| $t_a$           | Scores vector of X matrix of PCA and PLS models                           |
| $t_a^T$         | Transpose scores vector of X matrix                                       |
| $T_i^2$         | Hotelling's distance of an observation                                    |
| $t_{ia}$        | Score of an observation                                                   |
| T               | Scores matrix of PCA and PLS model                                        |
| $T_{new}$       | Matrix of new score vectors of new X variables                            |
| $u_a$           | Scores vector of Y matrix of PLS model                                    |
| $u_a^T$         | Transpose scores vector of X matrix                                       |
| U               | Scores matrix of Y of PLS model                                           |
| $Var(Y)$        | variance of preprocessed Y matrix                                         |
| $Var(F_{A,CV})$ | Variance of residuals matrix using cross-validation method                |
| $VIP_{j,A}$     | Importance index of variable j                                            |

|                  |                                                           |
|------------------|-----------------------------------------------------------|
| $w_a$            | Loading vector of X matrix of PLS model                   |
| $w_a^T$          | Transpose loading vector of X matrix of PLS model         |
| $W'$             | Loading matrix of X of PLS model                          |
| $w_{j,a}$        | Loading weight of variable j                              |
| $X_j^*$          | Preprocessed X                                            |
| $X_{(-i)}$       | Preprocessed X matrix of all groups except group i        |
| $X_{new}$        | Matrix of prediction dataset of X variables               |
| $X_j$            | Original X                                                |
| $X_{i,mean}$     | Mean of variable i                                        |
| $X_{i,STD}$      | Standard deviation of variable i                          |
| $X_a$            | Preprocessed X matrix                                     |
| $\hat{X}_a$      | Predicted variability of $X_a$ matrix                     |
| $X_{(i)}$        | Preprocessed group i of X matrix                          |
| $Y$              | Dependent (target) variable of multiple linear regression |
| $Y_{(-i)}$       | Preprocessed Y matrix of all groups except group i        |
| $\hat{Y}_{(1)}$  | Predicted variability of group i of Y matrix              |
| $\hat{Y}_{pred}$ | Matrix of the predicted Y variables                       |
| $Y_a$            | Preprocessed Y matrix                                     |
| $\hat{Y}_a$      | Predicted variability of $Y_a$ matrix                     |
| $Y_{(i)}$        | Preprocessed group i of Y matrix                          |

## List of Abbreviations

|        |                                            |
|--------|--------------------------------------------|
| ALBA   | Aluminium Bahrain                          |
| ANN    | Artificial Neural Network                  |
| AP     | Aluminium Pechiney                         |
| ARD    | Air Reactivity Dust                        |
| ARR    | Air Reactivity Residue                     |
| BAD    | Baked Anode Density                        |
| CPC    | Calcined Petroleum Coke                    |
| CTP    | Coal Tar Pitch                             |
| CRD    | Carboxy Reactivity Dust                    |
| CRR    | Carboxy Reactivity Residue                 |
| CTE    | Coefficient of Thermal Expansion           |
| CV     | Cross Validation                           |
| DC     | Direct Current                             |
| DUBAL  | Dubai Aluminium                            |
| FMR    | Fume Measurement Ramp                      |
| GAD    | Green Anode Density                        |
| GPC    | Green Petroleum Coke                       |
| GSR    | Grain to Sand Ratio                        |
| LMA    | Linear Multivariable                       |
| LV     | Latent Variable                            |
| MLR    | Multiple Linear Regression                 |
| NCC    | Net Carbon Consumption                     |
| NIPALS | Non-linear Iterative Partial Least Squares |
| PCA    | Principal Component Analysis               |

|       |                                             |
|-------|---------------------------------------------|
| PLS   | Projection to Latent Structure              |
| RDC   | Research and Development Carbon             |
| SIMCA | Soft Independent Modelling by Class Analogy |
| VBD   | Vibrated Bulk Density                       |
| VIP   | Variable Importance in Projection           |
| XRD   | X-Ray Diffraction                           |
| ZPR   | Zero Point Ramp                             |

# **Chapter 1: Introduction to the Study**

|                                        |          |
|----------------------------------------|----------|
| <b>1.1 Introduction.....</b>           | <b>2</b> |
| <b>1.2 Problem Statement.....</b>      | <b>5</b> |
| <b>1.3 Research Questions .....</b>    | <b>6</b> |
| <b>1.4 Research Objectives .....</b>   | <b>6</b> |
| <b>1.5 Research Significance .....</b> | <b>6</b> |
| <b>1.6 The Study's Setting .....</b>   | <b>7</b> |
| <b>1.7 Research Terminology .....</b>  | <b>7</b> |

## 1.1 Introduction

Aluminum is produced based on the Hall-Heroult process in electrolytic cells. In an electrolytic cell, a direct electrical current is passed to drive a chemical reaction. In aluminum production, the electrolyte solution of the cell where the chemical reaction occurs is the cryolite bath (mainly  $\text{Na}_3\text{AlF}_6$ ). The cell's electrodes are the carbon anodes and cathode blocks (carbon and graphite). Power stations provide the power source as direct current (DC). The main raw material for aluminum production, alumina ( $\text{Al}_2\text{O}_3$ ), is dissolved and dissociated in the cryolite bath to form positive and negative ions ( $\text{Al}^{3+}$  and  $\text{O}^{2-}$ ). Thus, the positive aluminum ions are attracted to the negative cathode charge at the bottom of the cell to form the aluminum metal. Meanwhile, carbon acts as a positive anode charge and attracts oxygen ions to form carbon dioxide ( $\text{CO}_2$ ) (Hulse, 2000). This process is based on the following reaction:

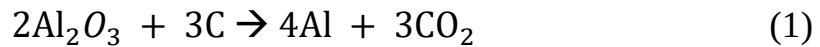


Figure 1 shows the arrangement of the electrolytic cell (pot) for aluminum production. The electrical current flows through the rod to the baked carbon anode and then through the bath to the molten aluminum. Finally, it flows through the cathode to the collector bar on its way to the next cell. Thus, carbon anodes play an essential part in electrical current conductivity in the cell (Meier & Carbon, 1996). The carbon anodes are consumed during the electrolysis process and are changed after nearly four weeks, as 75% of the anode is consumed. The remaining 25% is recycled and reused and known as anode butts. This percentage is left to avoid iron contamination from anode steel rod and cast iron (Hulse, 2000).

The carbon anodes produced for the electrolytic cells play a significant role in the quality and purity of aluminum and production cost. Thus, different important properties are required to achieve good-quality anodes. The main requirements of carbon anodes are good conductors of electricity, withstand physical impact during handling, withstand high temperatures, have high resistance to thermal shock, have low reactivity, do not produce dust or spikes, and have a low content of chemical impurities like iron and silicon. The quality of carbon anodes is crucial for the smooth operation of pots, the quality of aluminum metal produced, and the reduction of the overall cost of the aluminum smelting process (Aluminium Bahrain, 2020).

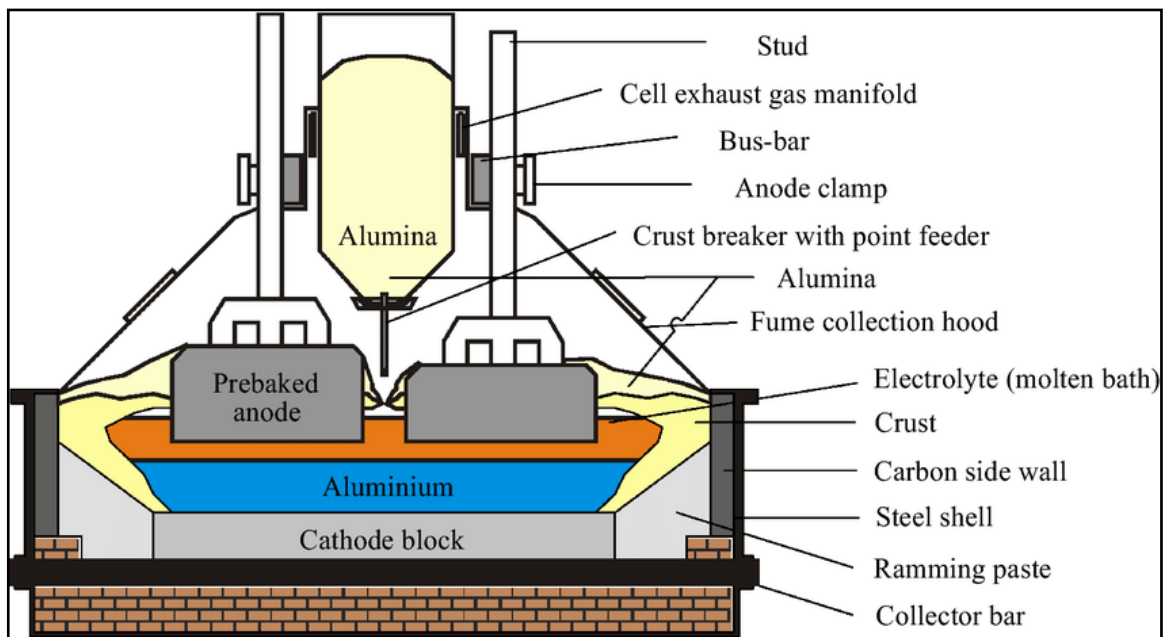


**Figure 1: Prebaked electrolytic cell**

The carbon plant consists of three main sections, during which the input raw materials end up with carbon anodes sent to potlines. Figure 2 shows the carbon process flow chart, and the three main sections are the paste plant, baking kilns, and anode rodding shop. Also, as shown, the main raw materials used for making green anodes are calcined petroleum coke (CPC), coal tar

pitch (CTP), anode butts, and green scrap. First, CPC is derived from the green coke by calcining it at temperatures up to 1300°C. Green coke is a by-product of petroleum refineries and contains mainly carbon and other impurities. The green coke was calcined to CPC because it is unsuitable for making the anodes due to its characteristics, such as shrinking at high temperatures, high electrical resistivity, and high reactivity.

Secondly, CTP is also derived from crude oil in petroleum refineries and acts as a binder for the anodes. Thirdly, anode butts are the remaining carbon of the spent anodes returning from potlines after the anode changing process, which takes place every 3-4 weeks. This anode butts is approximately 25% of the original anode. These anode butts are sent to the rodding section for separation of carbon from the aluminum rod and bath and then sent to the paste plant as recycled butts. Finally, the green scrap is mainly the rejected paste and rejected green anodes. During the start-up of the paste plant, the first production paste and green anodes are out of specification; thus, it is rejected and sent to the rodding section for crushing. However, the green scrap is not essential for producing green anodes. It is used for cost and waste reduction, and its percentage is less than 5%.

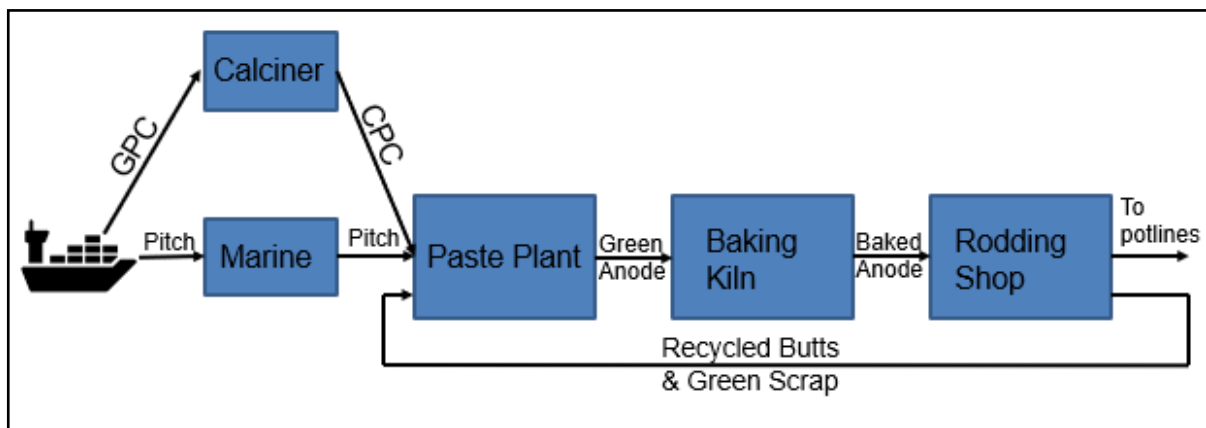


**Figure 2: Carbon plant process flowchart**

Carbon anode quality is a serious concern as the production of anodes represents a high percentage of aluminum production costs. Thus, all smelter industries seek to reduce carbon anodes consumption as much as possible. This could be achieved by improving the quality of carbon anodes. As the carbon anode quality improves, the number of anodes needed to produce the same aluminum metal will be minimized. Also, some industries tend to increase the amperage of potlines to increase aluminum production when carbon anode quality improves.

However, the anode production process is too complex, and improving the quality of anodes is not easy to control (Khaji & AlQassemi, 2016). Also, among the available publications, there were few mathematical models to control the quality of anodes based on their process parameters. Thus, high efforts are needed to reach the target quality of carbon anodes.

## **1.2 Problem Statement**

The carbon anode is a key component in the electrolysis process of aluminum production. Improving carbon anode quality is a high-importance concern in the aluminum smelting industry. As carbon anode quality is highly influenced by different raw material properties and process parameters, no well-known mathematical relationship or model is available to control the quality. As a result, aluminum industries are using core sample analysis to control the quality. However, there is a long delay between the start of anode production and the availability of the quality results, which could reach four weeks or even longer. Thus, mistakes in controlling parameters at the beginning of the production process would be observed after obtaining the quality results. Many bad-quality anodes are already produced or in production.

## **1.3 Research Questions**

1. What is the model that can help improve and predict anode quality?
2. What are the relationships between anode quality and process parameters?
3. What are the significant process parameters to determine the quality of anodes?
4. What are the main causes of variations in anode quality?

## **1.4 Research Objectives**

The objectives of the research are as follows:

1. To develop a mathematical model of anode quality based on its process parameters.
2. To obtain the significant process parameters determining the quality of the anode.
3. To investigate the main causes of variation in anode quality.
4. To define the relationship between the anode quality and process parameters.

## **1.5 Research Significance**

Currently, carbon plants in Aluminium Bahrain (ALBA) lack a mathematical model to improve and predict anode quality properties. Therefore, the research study is significant and will enrich the anode quality process controlling and improving. The research findings influence anode quality improvement, including obtaining the significant process parameters, the main causes of variation in anode quality, and defining the correlation between anode quality with process parameters and raw material properties.

## 1.6 The Study's Setting

This research will address the mathematical relationship of anode quality with process parameters in carbon plant number 2 in ALBA from November 2020 to October 2022. Therefore, the research results apply to carbon plants using Aluminium Pechiney (AP) technology.

## 1.7 Research Terminology

The common terminologies that will be used in the research study are:

### 1. **Carbon Anode:**

A carbon anode is an electrode made of carbon that is used in various electrochemical processes, including electrolysis, batteries, and aluminum production. *Electrochemical Engineering (Cambridge University Press, 2016)*.

### 2. **Mathematical Model:**

A mathematical model is an equation, or set of equations, that describes a physical problem or phenomenon originating from science, engineering, economics, or another area. *Applied Mathematics (Wiley, 2013)*

### 3. **Principal Component Analysis:**

Principal component analysis is a method of analysis that involves finding the linear combination of a set of variables that has maximum variance and removing its effect, repeating this successively. *Applied Multivariate Statistical Analysis (Pearson, 2014)*.

#### **4. Projection to Latent Structure:**

Projection to latent structure is a statistical method used for modeling relationships between observed variables and latent constructs. *Multivariate Data Analysis (Pearson, 2010).*

#### **5. Latent Variable:**

A latent variable is a variable that is not directly observed or measured but is inferred from other observed variables. *Applied Multivariate Statistical Analysis (Pearson, 2014).*

## **Chapter 2: Theoretical Framework and Literature Review**

|                                                                  |           |
|------------------------------------------------------------------|-----------|
| <b>2.1 Introduction.....</b>                                     | <b>10</b> |
| <b>2.2 Theoretical Overview.....</b>                             | <b>10</b> |
| <b>2.3 Previous Studies on Carbon Plant Process Control.....</b> | <b>30</b> |
| <b>2.4 Chapter Conclusion .....</b>                              | <b>37</b> |

## **2.1 Introduction**

This chapter aims to provide general information about controlling the carbon anode production process and predicting anode quality. Thus, it will review many published studies about the carbon process and anode quality prediction to gain insight into existing literature findings.

The chapter will be divided into two main parts. The first part (section 2.2) discusses the theoretical overview of carbon process control of raw materials, paste plant, and baking kiln, along with their most crucial controlling process parameters in each stage. This part will end with an explanation of the anode quality properties. The second part (section 2.3) reviews the literature on anode process parameters, anode quality, and their relationship. Moreover, anode quality prediction models based on process parameters literature are also reviewed.

## **2.2 Theoretical Overview**

Carbon anode raw materials are calcined petroleum coke (CPC), coal tar pitch (CTP), and anode butts. These raw materials in the carbon anode production process pass through three main sections, as shown in Figure 2: paste plant, baking kiln, and rodding shop. Each of these sections consists of different stages and processes. Many parameters could affect the quality of the anode produced during these processes. Thus, carbon plant process parameters are controlled to achieve the best quality anodes. Also, the quality of anodes produced is affected by raw material properties. This section discusses raw materials and carbon plant parameters, followed by anode quality properties, to get insight into carbon anode quality control.

### 2.2.1 Raw Materials Parameters

The first main raw material for carbon anode production is the CPC produced from green petroleum coke (GPC). The CPC represents 60-70% of the carbon anode. Thus, good-quality anodes are impossible if bad-quality CPC is used. Therefore, CPC chemical and physical properties are measured and controlled to maintain suitable quality anodes (Hulse, 2000). Table 1 shows the important CPC chemical and physical parameters that influence the quality of carbon anodes produced.

**Table 1: CPC chemical and physical properties**

| Property                           | Unit              | Target range |
|------------------------------------|-------------------|--------------|
| Moisture content                   | %                 | 0.0-0.2      |
| De-dusting oil content             | %                 | 0.0-0.4      |
| Coarse content (>4.75 mm)          | %                 | >25.0        |
| Fines content (<150 mm)            | %                 | 0.0-3.0      |
| Real density                       | g/cm <sup>3</sup> | 2.04-2.08    |
| Apparent density                   | g/cm <sup>3</sup> | >1.70        |
| Vibrated bulk density              | g/cm <sup>3</sup> | >0.84        |
| Hard grove grindability index      | -                 | 32-40        |
| Specific electrical resistance     | μΩm               | 460-540      |
| Crystallite size (L <sub>c</sub> ) | Å                 | 25-30        |
| Air reactivity                     | %/min             | 0.0-0.3      |
| CO <sub>2</sub> reactivity         | %                 | 0.0-15.0     |
| Ash content                        | %                 | 0.0-0.5      |
| Sulfur (S) element content         | %                 | 0.5-3.3      |
| Vanadium (V) element content       | ppm               | 0-350        |
| Iron (Fe) element content          | ppm               | 0-350        |
| Calcium (Ca) element content       | ppm               | 0-200        |
| Silicon (Si) element content       | ppm               | 0-400        |
| Nickel (Ni) element content        | ppm               | 0-300        |
| Sodium (Na) element content        | ppm               | 0-100        |

CTP is the second main raw material of carbon anodes, which is petroleum residue from petroleum refining industries. CTP acts as a binder in carbon anodes, and its quality is less variable than CPC quality. However, good-quality anodes could be achieved if the CTP properties are within the target range. The main pitch properties that influence the quality of anodes are given in Table 2.

**Table 2: Liquid tar pitch chemical and physical properties**

| <b>Property</b>              | <b>Unit</b> | <b>Target range</b> |
|------------------------------|-------------|---------------------|
| Softening point              | °C          | 111-115             |
| Viscosity at 150°C           | cP          | <4800               |
| Viscosity at 170°C           | cP          | <1200               |
| Viscosity at 190°C           | cP          | <400                |
| Coking value                 | %           | >56                 |
| Quinoline insoluble          | %           | 7-15                |
| Toluene insoluble            | %           | 26-34               |
| Ash content                  | %           | 0.0-0.2             |
| Sulfur (S) element content   | %           | 0.2-0.6             |
| Calcium (Ca) element content | ppm         | 0-80                |
| Sodium (Na) element content  | ppm         | 0-400               |
| Iron (Fe) element content    | ppm         | 0-300               |
| Zinc (Zn) element content    | ppm         | 0-500               |

The third main raw material in anode manufacture is the recycled butts which is the reminder of carbon anode. As the carbon anode cannot be wholly consumed due to iron contamination possibility, the reminder percentage of the anode is recycled as anode butts to manufacture the new anodes. Therefore, butts quality is essential to get high-quality anodes. The needed characteristics of anode butts are hard, low sodium content, and low reactivity.

Table 3 summarizes the measured properties of anode butts which are ash content and elements impurities content.

**Table 3: Anode butts main properties**

| <b>Property</b>              | <b>Unit</b> | <b>Target range</b> |
|------------------------------|-------------|---------------------|
| Ash content                  | %           | 0.0-1.0             |
| Iron (Fe) element content    | ppm         | 0-600               |
| Calcium (Ca) element content | ppm         | 0-200               |
| Silicon (Si) element content | ppm         | 0-400               |
| Sodium (Na) element content  | ppm         | 0-600               |

### **2.2.2 Paste Plant Process**

The first section of the carbon plant is the paste plant, where the raw materials from different sections are processed to form the green anodes. The major operations and processes in the paste plant are shown in Figure 3. So, it starts with the handling and storage of raw materials, each material in its silo (storage vessel). The silo size depends on the amount of raw material used for forming the green anode. For example, in ALBA, the CPC percentage in the green anode is the highest; thus, its silo is the largest compared to the CTP and anode butts.

The second stage in the paste plant is the dry aggregate mix preparation, which mainly prepares the anode recipe that will be mixed with the pitch. During this stage, different operations take place, such as crushing, screening, grinding, classifying, and proportioning. In this stage, the order of technology and operations differs between the old paste plant and the new one. In the old technology paste plant, each raw material is screened, crushed, and ground alone; after that, mixed with butts to be proportioned later. While in the new

technology, the CPC and butts are proportioned first. Then, the proportioned material is screened, crushed and grinded together (Aluminium Bahrain, 2020).

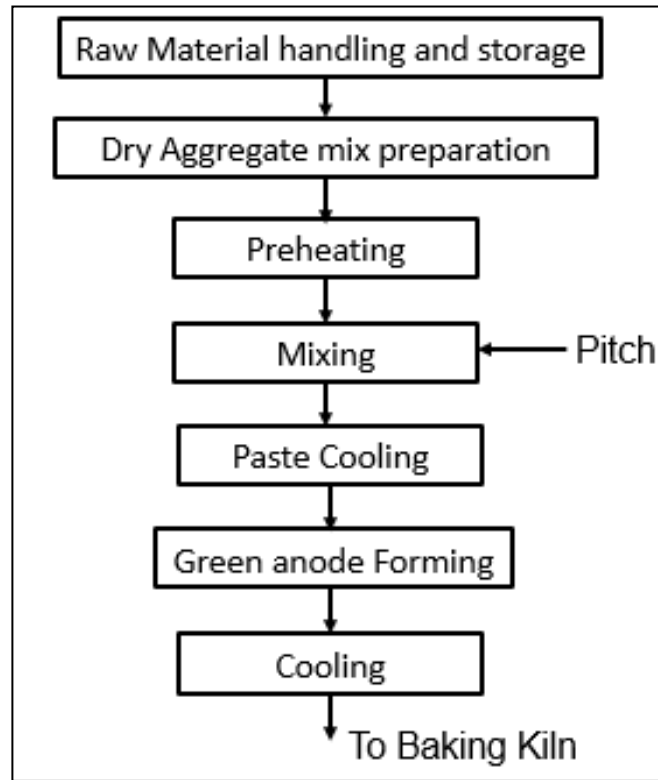


**Figure 3: Paste plant process flowchart**

The old technology dry aggregate mix is obtained and proportioned from four different particle sizes silos. These silos are very coarse (4.7-16 mm), coarse (1.7-4.7 mm), medium (0-1.7 mm), and fine ( $<75\ \mu\text{m}$ ). From each silo, a fraction of the dry aggregate mix is taken in order to obtain the optimum green anode proportioning recipe. The green anode proportioning recipe is a critical paste plant process parameter that should be controlled to get good anode quality. On the other hand, the new technology dry aggregate mix is proportioned from two silos which are the grains silo (coarse particles) and the fines silo (Aluminium Bahrain, 2020).

The green anode recipe is the other process parameter in the second stage of the paste plant that should be controlled to obtain a good-quality green anode. The green anode recipe is the fraction of each raw material in the dry aggregate mix, like the example of the recipe: 65% CPC, 20% butts, 2% green scrap, and 13% pitch. Moreover, each recipe is affecting different characteristics. For example, the increase in anode butts' content will increase the anode density. The size of particles is controlled using the gap size of the crushing and grinding devices (Hulse, 2000).

The third stage in the paste plant section, as shown in Figure 3, is the dry aggregate mix preheating. The main function of this stage is to increase the temperature of the dry aggregate mix from ambient temperature to nearly 190°C because, in the next stage, it will be mixed with the liquid pitch, which is at 190°C. This is done to get a good mixing quality and to avoid solidifying the liquid pitch when directly in contact with the low-temperature dry aggregate mix. Also, the higher the temperature of the dry aggregate mix, the better the green anode density. Thus, as a rule of thumb, the dry aggregate mix should be heated to a higher temperature than the pitch softening point. Therefore, the dry aggregate mix temperature is one of the important process parameters that should be controlled to have good-quality anodes (Aluminium Bahrain, 2020).

In the fourth stage, as shown in Figure 3, the heated dry aggregate is mixed with liquid pitch in the mixer to get the paste. In the mixing process, all the grains surface of the dry aggregate is wetted with the liquid pitch. Thus, the liquid pitch is called the binder of dry aggregate. Also, the liquid pitch fills the grain's pores and voids, and the grain sizes are distributed throughout the paste formed. The mixer parameters, such as mixing time, mixing energy, and

mixing temperature, are critical and directly affect the green anode density (Hulse, 2000).

The mixing time is increased enough to ensure that the pores and voids of grains are filled and coated well. However, it cannot be increased for so long because this may lead to a delay in green anode production and very high green anode density that may lead to thermal shock in baking kilns. Also, the mixing energy is a key parameter for the homogenizing of paste and allowing the liquid pitch to penetrate the dry aggregate. Moreover, the mixing temperature controls the effectiveness of coating grains by liquid pitch, which depends on the pitch viscosity (Aluminium Bahrain, 2020).

In the mixing stage, one crucial parameter controlled for a good quality green anode is the pitch percentage in the paste, called the pitch level. It is one of the essential process parameters that should be optimized because a lack or excess of pitch percentage will lead to poor anode quality. The green anodes produced are soft at a low pitch percentage due to inappropriate grain coating and insufficient pores filling. This leads to air burn of anodes in the potlines due to more gaps available in the anodes produced. On the other hand, green anodes with excess pitch level will propagate cracks and reduced dry density, which means gaps will be formed during the baking of anodes, resulting in air burn in potlines (Aluminium Bahrain, 2020).

The next stage in the paste plant is paste cooling, as the paste is cooled between the mixing stage and the anode forming stage. The paste exits the mixer at a temperature up to 190°C and is cooled to a temperature suitable for green anode forming. The paste is cooled to a temperature that suits the anode-forming unit. The cooling is performed by spraying water into the paste, resulting in the evaporation of water and cooling of paste temperature. The

temperature of the cooled paste is controlled by the water flow rate. During this stage, the cooler power is an important parameter to control the anode quality because as the cooler power increases, the green anode density increases (Aluminium Bahrain, 2020).

The most critical stage in a paste plant is the anode-forming stage. The anode-forming stage is done using either of two methods: hydraulic pressing or vibro-compacting. The paste needed for each anode is collected and weighed in the hopper and then fed to the mold, where the pressing or vibro-compacting takes place. The process parameters that influence the anode quality in the vibro-compaction method are forming temperature, vibrating speed, and vibrating time. Higher forming temperatures and vibrating speeds are more favorable for producing dense anodes, but over-dense anodes may result in cracks and segregation problems. After the green anode forms, the green anode density is measured, which is a critical process parameter in carbon plants (Hulse, 2000).

The last stage of the paste plant, as shown in Figure 3, is the green anode cooling. The green anode formed is cooled down in the cooling tunnel unit by spraying water for nearly one hour before discharge to the storage area. The cooling is performed to reduce the green anode temperature below the pitch softening point to avoid physical deformation and cohesion of the anodes during handling. (Hulse, 2000). Table 4 summarizes the important process parameters in paste plants that affect the quality of anodes produced.

**Table 4: Paste plant main process parameters**

| <b>Parameter</b>               | <b>Unit</b>       |
|--------------------------------|-------------------|
| Green anode density (GAD)      | g/cm <sup>3</sup> |
| Pitch level                    | %                 |
| Green scrap content            | %                 |
| Butts (very coarse) percentage | %                 |
| Coarse particles percentage    | %                 |
| Medium particles percentage    | %                 |
| Fine particles percentage      | %                 |
| Dry aggregate temperature      | °C                |
| Preheater speed                | rpm               |
| Mixing power                   | kW                |
| Mixing temperature             | °C                |
| Mixing speed                   | rpm               |
| Cooling power                  | kW                |
| Cooling temperature            | °C                |
| Anode forming temperature      | °C                |
| Anode forming speed            | rpm               |
| Anode forming time             | Seconds           |

### **2.2.3 Baking Kiln Process**

The second section of the carbon plant is the baking kilns, as shown in Figure 2. In the baking kilns, the green anodes received from paste plants are baked in furnaces to produce baked anodes ready to be rodded in the rodding shop. The green anodes with binder pitch are poor electrical conductors, soften if heated to a temperature higher than the softening point of pitch, and are highly chemically reactive. Thus, the green anodes are sensitive to mechanical shocks, especially during handling, and cannot be heated without

collapsing under their weight. Therefore, the green anodes are baked to achieve three critical purposes (Aluminium Bahrain, 2020):

1. Make the anodes electrically conductive.
2. Make the anodes resistant to thermal shocks and cracks, and withstand various handling processes (improve the anode's mechanical characteristics).
3. To achieve as low as possible chemical reactivity of anodes.

The baking of anodes is a process of transforming pitch into pitch coke. This is done by removing volatile matters as the green anodes are heated and avoiding air contact with green anodes. The baking furnace's heating energy is obtained from fuel combustion and pitch volatile matters. After loading the green anodes into the baking kilns, the green anodes are covered with packing coke. The packing coke is used to prevent the airburn of green anodes and to act as a heat transfer medium from the flue gas and green anodes (Aluminium Bahrain, 2020).

The baking kiln process is carried out using several identical groups of equipment. Each group of equipment is called a fire group. The fire group is moved in a circular manner from section to section. Each section in the kiln consists of pits, where each pit can take several anodes. For example, some pits can take eighteen anodes in three layers; each layer can hold six anodes. The number of fire groups in each kiln depends on the number of sections in that kiln. A kiln with 50 sections can hold three fires only, while 68 68-section kiln can hold four fires operated simultaneously. Each fire group is performed in twelve sections plus two sections for loading and discharge (Aluminium Bahrain, 2020).

Figure 4 shows the fire group with its equipment. The first equipment in the fire group is the exhaust manifold, which is the connection from the source of draught suction to the flue walls. The main role of the exhaust manifold is to supply the needed negative pressure (vacuum) for the fire group. Secondly, the fume measurement ramp (FMR) is a measuring device that communicates and controls the exhaust manifold by feeding the temperature and vacuum pressure using the thermocouple and sensors. Third, the burner ramp, a three-burner bridge, heats the three sections to the target temperature.

The next ramp in the fire group is the zero-point ramp (ZPR), and at this point is the transition from the negative pressure produced by the exhaust manifold to the positive pressure achieved from the blower ramp. So, the ZPR communicates with the blower ramp to maintain zero pressure in the zero-point ramp section. Thus, the blower ramp is a fan to each flue wall that supplies positive pressure to maintain the flow of oxygen (or air) to the burning area. Moreover, the anodes are cooled by blowing air in that section. The last ramp of the fire group is the cooling ramp, also called the secondary cooling manifold. In this section of the fire group, the blowing fans are cooling the adjacent anodes directly with all the covers closed (Meier & Carbon, 1996).

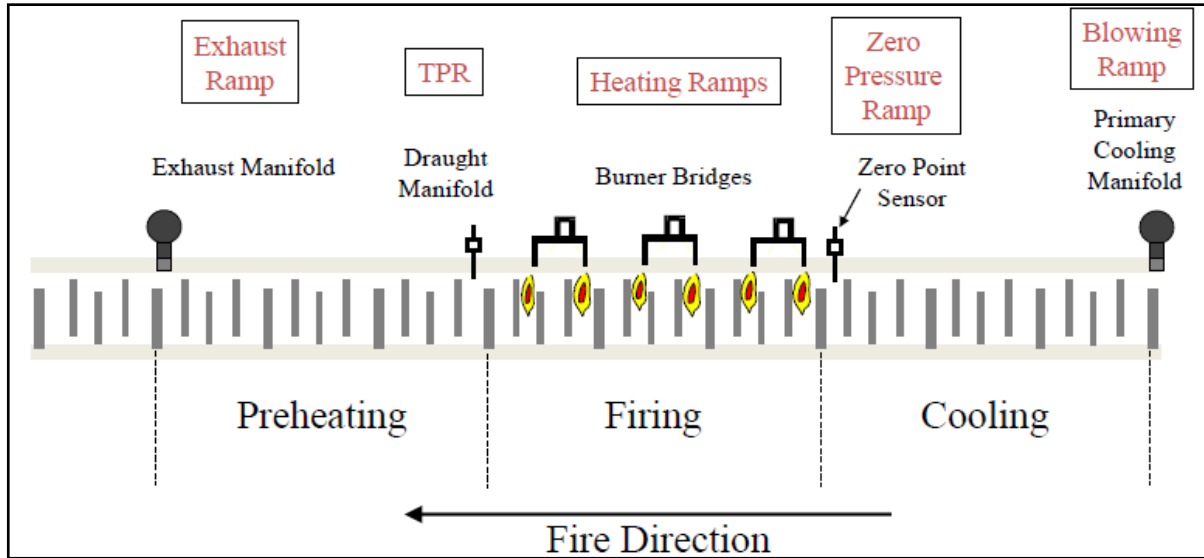


**Figure 4: Fire group in the baking kiln**

After some time, a fire change process is performed in the baking kiln. This is the main process in the kiln operation, where the fire equipment from the exhaust manifold to the cooling ramp is displaced in one section. The period in which the fire change takes place is called the fire cycle. Thus, the fire cycle is the time elapsed before the fire change process is needed. For example, the cycle of 26 hours means that every 26 hours, the fire equipment will be moved to one section. The fire cycle is one of the most critical baking kiln parameters that should be controlled to achieve a good anode quality. Also, the fire cycle depends on the production rate needed and the furnace's safety. Thus, due to furnace safety, most baking kilns' minimum fire cycle time is 24 hours, even though reducing it means more production (Aluminium Bahrain, 2020).

The green anodes are transformed and exposed to different zones and stages of baking during their period in the baking kilns. The heating zones during the baking process can be divided into three main zones: preheating zone, degassing zone, and baking (force heating) zone. The baking process

parameters during these zones and anode quality are highly related and manipulated accordingly to achieve the goals of the baking process effectively.

However, some goals of the baking process could be achieved by controlling one of the process parameters while affecting other goals negatively. For example, increasing the final anode temperature will improve anodes' resistance to air reactivity. It will increase the thermal conductivity dramatically, resulting in high air reactivity at the top of anodes in pots. Thus, the final anode temperature highly correlates to anode properties, especially CO<sub>2</sub> reactivity, air reactivity, and thermal conductivity. Also, the increase of final anode temperature to these high values will have a negative impact on refractory bricks and fuel efficiency (Keller, 2004).

The third important baking process parameter (after fire cycle time and final anode temperature) is the heating up rate of anodes. The heating-up rate is highly relevant to electrical resistivity and flexural strength, which are crucial to achieving good-quality anodes. Increasing the heating-up rate to certain levels increases the possibility of anode cracking and decreases flexural strength sharply. Also, increasing the heating-up rate causes an increase in electrical resistivity and increases the risk of thermal shock behavior.

The last two crucial baking process parameters are the soaking time and temperature. The soaking temperature is the maximum flue wall temperature, and it is kept at this temperature for a period called the soaking time. The soaking temperature is controlled based on the needed final anode temperature, which will affect the anode reactivity. The soaking time is manipulated to ensure that anodes in different positions in the pit are baked to

the final baking temperature. Thus, these two parameters are controlled to decrease the dusting problem in potlines (Keller, 2004). Table 5 shows the main five process parameters in baking kilns.

**Table 5: Baking kiln main process parameters**

| <b>Parameter</b>        | <b>Unit</b> |
|-------------------------|-------------|
| Cycle time              | hours       |
| Final anode temperature | °C          |
| Heating rate            | °C/hours    |
| Soaking time            | hours       |
| Soaking temperature     | °C          |

#### **2.2.4 Anode Quality Measurement**

The quality of the anode is very important for pot operation, as a bad anode will result in bad-quality aluminum being produced and sold to customers. Thus, the anode properties required to produce good-quality and cost-effective aluminum should be achieved. These anode quality requirements are good electricity conductors, withstand physical impact during handling, withstand high temperature and thermal shock, low reactivity, no dusting, low voltage drops, and low chemical impurities such as iron and silicon (Aluminium Bahrain, 2020).

To measure the quality of the anodes produced, sample cores are drilled from the anodes using a diamond wheel saw. The drilled sample cores are in a cylindrical shape with a diameter of approximately 50 mm. It is necessary to measure enough samples to characterize the anode population. The percentage of drilled anodes to total anodes is nearly 2%, which differs from one baking kiln to another. The sample cores should represent the sample size,

position, and frequency, representing the complete batch. However, not all the drilled sample cores are analyzed. Some sample cores are stored for abnormal situations, and others are rejected due to the bad situation of sample cores. In the end, nearly 1% of the total anode population is measured and analyzed by conducting different tests in the laboratory. The anode quality is determined based on the obtained results (Meier & Carbon, 1996).

The anode cores' physical and chemical properties are tested to measure the quality of anodes produced in the carbon department. Different properties are tested for anode cores, but the most important properties are geometric (apparent) density, real density, electrical resistivity, crystallite size (Lc), flexural strength, Young's modulus, coefficient of thermal expansion (CTE), thermal shock resistance, thermal conductivity, air permeability, CO<sub>2</sub> reactivity, air reactivity and impurities elements content. The quality of anodes produced is decided based on these measured properties, and decisions are made according to the results of these properties (Aluminium Bahrain, 2020).

The first important property is geometric density, the geometric space occupied by solid material, including interior voids, porosity, and cracks. The geometric density of anode cores is measured by measuring the mass of the sample, length, and diameter. The anode performance in potlines is characterized based on the geometric density of the core sample, which represents the anode's apparent density. Geometric density indicates the amount of mass available for reaction in electrolytic cells. The higher the density, the lower the permeability and gaps. Also, higher densities provide less dusting and better mechanical strength.

The geometric density above the limit is unfavorable to be exceeded because this will result in a negative effect related to thermal shock damage. Hence, the target geometric density for anode core is between 1.58 g/cm<sup>3</sup> and 1.62 g/cm<sup>3</sup>. The process parameters that should be controlled to obtain the target geometric density are the green anode proportioning recipe, grain-to-sand ratio (GSR), green anode recipe (raw materials recipe), mixing parameters (temperature, time, and energy), forming parameters (temperature, time & speed), heating up rate and final baking temperature (baking level). Thus, the geometric density is very important and difficult to control because it is affected by different parameters (Meier & Carbon, 1996).

The real density is the second density measured and is highly important for the anode production process. Real density is the volume excluding the closed and open pores occupied by material. The real density directly indicates the baking level of the anode. The most important process parameters directly affecting the real density are anode final temperature and soaking time. The target real density of the anode is greater than 2.10 g/cm<sup>3</sup>. Because lower real density indicates a lower baking level; as a result, anodes will create dusting and increased reactivity (Aluminium Bahrain, 2020).

The crystallite size (Lc) is another analysis performed in the laboratory for anode using the X-ray diffraction (XRD) method. The crystallite size is the size of a single crystal inside a particle (consisting of a group of crystals). Anode baking aims to improve the crystallite size related to the arrangement of atoms to form crystals. The crystallite size indicates the baking level of anodes, where the baking level shall be at the optimum level. As the baking temperature increases, the crystallite size will increase. However, above certain limits, the pores and voids in the anode will increase, resulting in

reactivity problems. The process parameters are controlled based on the crystallite size, which is the final anode temperature and soaking time (Aluminium Bahrain, 2020).

The fourth important property of anode is the electrical resistivity ( $\rho_{ER}$ ), which is a measure of the resistance to conducting electrical currents of a material of known size. Electrical resistivity indicates the hairline cracks, mechanical strength, and electrical conductivity of anodes in potlines. Thus, the electrical resistivity should be as minimal as possible because this directly affects the current efficiency in the electrolytic cell (minimum voltage drop). The process parameters affecting the electrical resistivity are final anode temperature, heating up rate, and pitch percentage. A high heating rate increases the electrical resistivity, while a higher pitch percentage decreases the electrical resistivity. Also, the sulfur content is directly proportional to electrical resistivity (Meier & Carbon, 1996).

Another significant anode property to characterize the mechanical strength of anodes is the flexural strength (bending strength). The flexural strength is the maximum applied stress on a material just before it fractures. This test gives an idea about the strength of an anode and whether the anodes produced will withstand the mechanical stress during handling. As the strength of anodes increases, the pot abnormality decreases due to less anode problems. However, very high strength is not preferable as this will result in thermal shock behavior (Meier & Carbon, 1996).

The process parameters influencing the flexural strength are green anode proportioning recipe, green anode recipe, pitch percentage, final anode temperature, heating up rate, soaking time, and soaking temperature. The butts percentage of the anode recipe positively influences the flexural strength,

while the heating-up rate negatively influences the strength of anodes. Also, increasing the pitch percentage increases the anode's flexural strength (Aluminium Bahrain, 2020).

One of the other physical properties that indicates the mechanical strength of anode core is Young's modulus, also known as the dynamic modulus of elasticity. The process parameters affecting Young's modulus are similar to those affecting flexural strength. Also, the CTE is one of the important parameters for understanding the mechanical strength and thermal shock behavior of anodes in electrolytic cells. This parameter is mainly associated with the quality of CPC used. The CTE of the anode core should be as low as possible for good anode quality (Meier & Carbon, 1996).

Thermal conductivity is another parameter used to understand the thermal shock behavior of anodes. Thermal conductivity is used to indicate the baking level of anodes directly. The anodes should have optimal thermal conductivity, as low values will produce less current efficiency in electrolytic cells. On the other hand, high thermal conductivity values will lead to air burns and thermal shock behavior. Process parameters that influence the thermal conductivity of anodes are the final temperature of the anode, soaking time, and soaking temperature (Aluminium Bahrain, 2020).

As discussed, 40% of the pitch is volatilized during baking. This leads to open pores development. That is why air permeability is an important property to measure of anodes. The air permeability of an anode is a measure of airflow passed through the anode. As the air permeability of anodes depends on porosity, thus it influences the density inversely. The air permeability of anodes is desired to be of low values because high values will result in air and CO<sub>2</sub> passing through the anodes, increasing the reactivity of

the anodes. The air permeability is affected by different process parameters such as the green anode sizing recipe, mixing parameters, forming parameters, pitch percentage, and anode final temperature (Meier & Carbon, 1996).

The chemical properties of anodes are measured along with the physical properties to characterize the quality of anodes produced. The main chemical properties are the reactivity and impurities elements of anodes. First, the carbon anode reacts with carbon dioxide to form carbon monoxide. This reaction is also called the carboxy reaction. In the electrolytic cell, carbon dioxide generated by the electrochemical oxidation in the anode surface attacks the carbon anode to form carbon monoxide as a side reaction. Equation (3) shows that reaction depends on anode raw materials quality and air permeability of anodes. After the reaction, the amount of carbon left out is measured and called carboxy reactivity residue (CRR) (Meier & Carbon, 1996).



High values of CRR are required for good performance and ensuring that carbon anodes will be spent in the preferable reaction to produce aluminum metal. The main parameters that affect the CRR value are Sodium content which acts as a catalyst of carboxy reaction, sulfur content which acts as an inhibitor of carboxy reaction, and butts sodium level. The sodium level in butts is a function of cleaning quality and removal of bath particles. Also, the final anode temperature affects the CRR as higher temperature increases CRR values (Aluminium Bahrain, 2020).

The second reactivity property of the anode is the air reactivity, as shown in Equation (2). This reactivity occurs on the surface of anodes at a certain temperature. Research and Development Carbon (RDC) apparatus is used to measure the air reactivity residue (ARR), which represents the amount of carbon left after the reaction with air. ARR is affected by the quality of raw materials used to form green anodes and the final anode temperature. The increase of final anode temperature affects the ARR positively until certain limits. The air reactivity dust (ARD) is reduced with the increase of final anode temperature. Also, impurities such as vanadium act as catalysts for air reactivity and negatively affect ARR (Meier & Carbon, 1996).

**Table 6: Physical and chemical anode properties**

| <b>Property</b>                    | <b>Unit</b>         | <b>Range</b> |
|------------------------------------|---------------------|--------------|
| Geometric density                  | g/cm <sup>3</sup>   | 1.58-1.62    |
| Specific electrical resistance     | μΩcm                | <5500        |
| Flexural strength                  | MPa                 | 8.0-12.0     |
| Young's modulus                    | GPa                 | 6.0-9.0      |
| Coefficient of thermal expansion   | 10 <sup>-6</sup> /K | 4.0-5.5      |
| Thermal conductivity               | W/mK                | 3.0-4.5      |
| Air permeability                   | nPm                 | 0.5-1.5      |
| Real density                       | g/cm <sup>3</sup>   | >2.06        |
| Crystallite size                   | Å                   | 31-35        |
| Ash content                        | %                   | <0.5         |
| CO <sub>2</sub> reactivity residue | %                   | >80.0        |
| CO <sub>2</sub> reactivity dust    | %                   | <5.0         |
| CO <sub>2</sub> reactivity loss    | %                   | <10.0        |
| Air reactivity residue             | %                   | >65.0        |
| Air reactivity dust                | %                   | <10.0        |
| Air reactivity loss                | %                   | <30.0        |
| Sulfur (S) element content         | %                   | 0.6-2.8      |
| Vanadium (V) element content       | ppm                 | 80-220       |
| Iron (Fe) element content          | ppm                 | 200-500      |
| Calcium (Ca) element content       | ppm                 | 100-300      |
| Silicon (Si) element content       | ppm                 | 50-250       |
| Nickel (Ni) element content        | ppm                 | 50-200       |
| Sodium (Na) element content        | ppm                 | 100-300      |

## **2.3 Previous Studies on Carbon Plant Process Control**

### **2.3.1 Properties and Parameters Affecting Anode Quality**

Reviewing existing literature, it has been observed that anode properties are analyzed to find correlations between them with raw material

properties and process parameters. Thus, many researchers have investigated the relationship between some anode properties and raw material properties and process parameters in paste plants and baking kilns to decide the most important factors that influence the anode produced quality.

#### ***2.3.1.1 Raw material properties and anode quality***

Khaji and AlQassemi (2015) analyzed the baked anode properties to correlate them with CPC properties and process parameters in paste plants and baking kilns. The study's main results have shown that CPC's vibrated bulk density (VBD) affects the baked anode density directly proportional. Also, CPC with high sulfur content increases the electrical resistivity of anodes produced. Impurities like sodium, calcium, and iron adversely impact CRR as these elements act as catalysts for carboxy reactions. Moreover, higher vanadium content reduces ARR. On the other hand, sulfur content increases the CRR because it acts as an inhibitor for secondary (unwanted) reactions.

Ndjom et al. (2016) have implemented a roadmap of 3 years in Dubai Aluminium (DUBAL) intending to improve baked anode density and air permeability in order to increase the anode lifetime in electrolytic cells or increase the current of pots, which will result in increasing the metal production. Different raw material properties were studied to improve these two anode quality properties. The VBD of CPC is correlated with GAD, and it was found that these two parameters are directly proportional. Another characteristic of CPC that influences the CRR of anode is the CO<sub>2</sub> reactivity. Thus, a regression line correlating the anode's CO<sub>2</sub> reactivity of CPC and CRR is developed. As expected, this characteristic of CPC highly affects the CRR of anodes. Moreover, the Vanadium content in the anode, which acts as a catalyst for secondary anode reactions, is correlated with ARR.

The impurities in CPC directly impact anode quality, as Khaji and AlQassemi (2016) found that the high level of impurities in coke negatively affects anodes' reactivities (ARR and CRR). Therefore, the net carbon consumption (NCC) is negatively affected as anode reactivities increase. Moreover, the study stated that an annual savings of nearly half a million US Dollars could be achieved for every kg reduced in NCC in one million tons of aluminum metal produced yearly.

Sijabat et al. (2017) have achieved an increase in the ARR and CRR of anodes by reducing the impurity element Sodium in butts. This is achieved by adding additional manual cleaning of butts used in forming anodes. Gauthier et al. (2011) investigated the raw material properties and process parameters that are the main causes of variation in the anode quality. The study results show that process variability in anode quality is due to changes in raw material properties.

### ***2.3.1.2 process parameters and anode quality***

In addition to raw material properties, paste plant and baking kiln process parameters directly impact anode quality. Paste plant parameters such as green anode recipe, mixing parameters, and forming parameters have been studied by many researchers to understand their impact on anode quality. Khaji and AlQassemi (2015) have optimized the pitch percentage to obtain the best GAD. The study has shown that increasing the pitch percentage increases the flexural strength of baked anodes, decreases CRD and ARD, and decreases the electrical resistivity of anodes.

Also, Khaji and AlQassemi (2016) indicate that increasing pitch content will decrease CRR, reducing the NCC. Moreover, Armani et al. (2021) have suggested an optimal percentage of pitch (15%) to minimize

cracks and defects in anodes, resulting in a baked anode density (BAD) increase. Furthermore, increasing the level of butts content in the green anode increased the anode defects.

Ndjom et al. (2016) found that increasing coarse particles (above 4.75 mm) content could improve the BAD and reduce air permeability. Another contribution regarding the coarse particle content by Sijabat et al. (2017) suggested that ARR and CRR increased by increasing coarse particles. Also, Khaji and AlQassemi (2015) studied the mixing parameters' effect on anode quality, suggesting that an increase in mixing energy will increase BAD by enabling the filling of micropores of coke particles.

Moreover, Sommerseth (2016) investigated the mixing parameters and suggested that a high mixing temperature is required to obtain the target physical properties such as BAD, electrical resistivity, and air permeability. In addition, it has been found that mixing temperatures are more critical than baking temperatures in obtaining the needed anode properties. Gauthier et al. (2012) model interpretation showed that the parameters that contribute the most to process variability changes are coke impurities content and mixer temperature.

The forming (compaction) parameters are highly correlated to anode quality properties. Thus, Ndjom et al. (2016) correlated the compaction speed with the GAD and found that these two parameters are directly proportional. However, increasing the compaction parameters over the limit would result in cracks and defects in green anodes, as Armani et al. (2021) suggested. Therefore, optimal compaction parameters are needed to get the best quality green anode, as over- and under-compaction have increased defects and crack formation.

The baking kiln process parameters strongly influence anode quality; thus, many studies have been carried out to study the impact of these parameters on the properties of the anode produced. Fischer et al. (2016) studied the effect of baking kiln parameters and found that a high heating rate dramatically increases flexural strength and electrical resistivity deviation. On the other hand, a low heating rate leads to low production and a shortage of anodes in potlines. Thus, the maximum allowable heating rate achieved through the study was between 10°C/hr to 14°C/hr.

Low final baking temperature or short soaking time leads to high anode reactivities (low ARR and CRR). Too high a final baking temperature leads to Fairburn behavior due to high desulphurization and high thermal conductivity of anodes. Also, low soaking time could lead to a high-temperature difference between the coldest and hottest anodes.

Khaji and AlQassemi (2015) have investigated the effect of final anode temperature on quality. The study has shown that increasing baking temperature increases the real density, crystallite size, and thermal conductivity. On the other hand, as the baking temperature increases, the desulphurization increases, resulting in a decrease in CRR and increasing in air permeability. Therefore, optimal final anode temperature should be obtained for the best quality. Sommerseth (2016) also found that the final anode temperature directly impacts the reactivities of anodes. Furthermore, Armani et al. (2021) suggested that the percentage of defects was decreased by reducing the heating rate during the baking of anodes.

### **2.3.2 Anode Quality Models**

There is a lack of models relating anode quality properties with raw material properties and process parameters. However, some researchers have

tried to build a model for anode quality properties prediction based on raw material and process parameters. The anode quality models that had been used to predict anode properties: multivariate latent variable partial least square (PLS) (Gauthier et al., 2011; Gauthier et al., 2012) and artificial neural network (ANN) (Bhattacharyay et al., 2015; Bhattacharyay et al., 2016; Kocaefer et al., 2017).

Gauthier et al. (2011) used a multivariate latent variable called the projection to latent structure (PLS) method to build a model for predicting anode quality. The model developed has been used to investigate the process variability and main variables causing variation in anode quality. In addition to the model developed, the contribution plot method was used to identify the most contributed variables that lead to a change in anode quality results. Also, the anode quality properties prediction using the model is compared with observed properties, and the results show a good correlation model as most variables have good predictions using the PLS method. Thus, the prediction model developed could be used as a tool for faster process controlling and variation detection.

Gauthier et al. (2012) developed a soft-sensor model using a multivariate latent variable PLS regression method to predict baked anode properties. The historical database of the Alcoa carbon plant was used to build the model and validate its prediction ability. The PLS model shows good prediction ability for some anode properties and low prediction ability for others. The best fit was obtained for baked weight, and the worst fit was for thermal conductivity.

Bhattacharyay et al. (2015) studied the ANN model for predicting the CO<sub>2</sub> reactivity of carbon anodes used in primary aluminum production.

The study focused on one anode quality parameter, the CO<sub>2</sub> reactivity of anodes, and the input parameters were coke characteristics and pitch percentage. However, the data was not industrial data; laboratory-made anodes were used. Thus, as in industrial data, the noise and missing data were not challenging. The results show that the ANN model is a good method for CO<sub>2</sub> reactivity prediction based on input parameters. Also, the average error was 6%, a highly excellent prediction error compared with the small number of samples. Moreover, the obtained result trends agreed with the results of the literature.

The ANN method was compared with linear multivariable (LMA) and partial least square (PLS) regression to predict anode properties by Bhattacharyay et al. (2015). BAD, the most important anode property, was studied and predicted based on four process parameters using the three methods mentioned. The developed ANN model could provide additional information regarding the effect of process parameters on BAD compared to the other two methods. Also, the ANN model could give the trend of each parameter with anode-quality results. However, the results show that all three methods identified the expected trend of the anode quality based on different parameters studied. The study's main finding was that the ANN model was the best tool for predicting anode quality.

In their third study of ANN, Bhattacharya et al. (2016) studied the application of the ANN in predicting three anode properties along with other methods, such as the LMA method. The anode properties studied were BAD, electrical resistivity, and Young's modulus. The input parameters were nineteen variables representing raw material properties. Results of the study show that the feedforward ANN method appeared to be the best tool for

predicting anode properties based on raw materials properties and process parameters, as this method's coefficient of determination values were the highest. Linear multivariable analysis was the worst method, which indicates the nonlinearity of anode properties prediction.

Furthermore, Kocaefe et al. (2017) analyzed anode core properties to improve the quality of anodes produced. The research's main results show that the ANN model is a useful method to determine the effect of a particular parameter on anode quality without the effect of other parameters.

## **2.4 Chapter Conclusion**

Prediction of anode quality is an important aspect of the carbon department in the aluminum industry. Anode quality has a direct influence on aluminum production and NCC values. Thus, any smelter company that plans to increase its production should investigate how to improve the quality of its anodes. Many researchers have studied the relationship between anode quality and process parameters.

These past works have shown the direct effect of anode raw material characteristics and process parameters on the anode-produced quality. The researchers have used historical data from previous years to study the effect of raw materials and process parameters on anode quality. Also, some researchers have used laboratory anode fabrication to correlate anode quality with different parameters. These researches focus on correlating anode quality results with some process parameters to manipulate these parameters in order to get favorable anode quality results.

The anode production process is too complex; changing one parameter could affect other parameters. Thus, the correlation between the quality

variable and process parameter is inefficient and could take a long time of trial and error to achieve the needed quality of anodes due to the interrelation between variables. Moreover, there is a time delay between the production of anodes and the availability of anode results. This situation makes controlling the production process difficult, and actions are reactive after damage occurs. Therefore, an efficient way to predict the anode quality results based on process parameters and raw material properties is needed.

However, literature has shown the absence of a general mathematical model that can relate the anode quality results to raw material characteristics and process parameters. Also, there is a lack of research that provides mathematical relationships to control anode quality. Most researchers in the literature correlate parameter by parameter to the anode quality characteristics. This way of tackling the problem could end up with a solution, but it is not as efficient and effective as a mathematical model.

Hence, this study addresses the need for a mathematical model between anode quality results, raw material characteristics, and process parameters. Some modeling methods could be used to handle missing data, massive amounts of data, and noisy processes like carbon anode production. The research aims to develop a mathematical model to predict anode quality results. Thus, anode quality results would be predicted based on that model, along with the main parameters that cause variation in anode quality. This study will improve the carbon production process control at the ALBA company.

## **Chapter 3: Methods and Procedures**

|                                                                              |           |
|------------------------------------------------------------------------------|-----------|
| <b>3.1 Introduction.....</b>                                                 | <b>40</b> |
| <b>3.2 Evaluation of Prediction Model of Anode Quality Results<br/>.....</b> | <b>40</b> |
| <b>3.3 Multivariate Statistical Methods .....</b>                            | <b>42</b> |
| <b>3.4 Research Tools .....</b>                                              | <b>62</b> |
| <b>3.5 Data Collection .....</b>                                             | <b>62</b> |
| <b>3.6 Chapter Conclusion .....</b>                                          | <b>68</b> |

### 3.1 Introduction

This chapter illustrates methods used to build the anode quality models, the evaluation of these methods, and the tools used to build models. Also, the chapter contains the procedure followed to collect the carbon plant data needed to build and validate the models.

### 3.2 Evaluation of Prediction Model of Anode Quality Results

Previous studies used simple linear regression to evaluate one dependent anode quality variable based on one independent process or raw material parameter. These studies help improve the anode quality but neglect the interrelationship between different anode quality variables and process parameters variables. However, Gauthier et al. (2011) used multivariate statistical methods to improve anode quality results without neglecting the interrelationship between variables.

Regression analysis is a predictive modeling technique used to analyze the relationship between different variables. It is built based on the relationship between a dependent variable (target) and independent variables (predictors). There are different types of regression analysis techniques used to analyze prediction models between variables, and the most popular types are linear regression, logistic regression, nonlinear regression, and polynomial regression (Alexopoulos, 2010). However, the most popular regression analysis for different sets of variables is multiple linear regression. The general expression for multiple linear regression is as follows (Montgomery & Runger, 2010):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + e_i \quad i = 1, 2, \dots, n \quad (4)$$

Where  $Y$  is the dependent (target) variable,  $x_i$  ( $i=1, 2, \dots, n$ ) are the independent (predictors) variables,  $\beta_i$  ( $i=1, 2, \dots, n$ ) are the regression coefficients, and  $e_i$  is the random error.

Multiple linear regression explains only one dependent variable. So, a dataset of more than one dependent variable would have more than one expression model. In the case of anode quality results, the dependent variables are interrelated to each other. So, the set of independent variables results would be different from one expression model to another. Thus, one model that explains the relationship between a set of dependent variables and a set of independent variables is needed for anode quality results prediction.

Multivariate regression analysis techniques are commonly used for this purpose. The most used methods are Principal Component Analysis (PCA) and Projection to Latent Structures (PLS). These two methods are suitable for analyzing and predicting anode quality results as the datasets are very large and noisy. Also, the variables are highly interrelated and collinear (Abdi, 2010).

In this research study, two multivariate regression models will be used where more than one dependent variable will be predicted simultaneously. The first model dependent variables will be physical variables, and the second will be chemical and reactivity ones. The model prediction strength is calculated based on variance values ( $R_x^2$ ,  $R_y^2$ , and  $Q^2$ ). Table 7 summarizes the three residuals' titles and descriptions. A high value of  $R^2$  is an indication of a good fit model, and a high value of  $Q^2$  indicates a good predictivity model (Wold et al., 2010).

**Table 7: Regression model variances' titles and descriptions**

| <b>Symbols</b> | <b>Title</b>                                  | <b>Description</b>                                                                                                                                   |
|----------------|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| $R_x^2$        | Independent variables variance                | Percent of variation of the training set of X variables explained by the model.                                                                      |
| $R_y^2$        | Dependent variables variance                  | Percent of variation of the training set of Y variables explained by the model (how well the model fits the data).                                   |
| $Q^2$          | Cross-validation dependent variables variance | Percent of variation of the training set of Y variables predicted by the model according to cross-validation (how well the model predicts new data). |

### **3.3 Multivariate Statistical Methods**

As mentioned before, two of the most popular methods for multivariate analysis are PCA and PLS. These two methods are based on the latent variables principle. The main idea of latent variables is to capture and investigate the hidden phenomenon in the data to be studied and modeled by using a small number of variables. So, a few latent variables are used to understand and correlate between variables instead of using massive raw data. The latent variable modeling methods are suitable for noisy, non-casual, and missing data in the system investigated. Moreover, these methods are mostly used for data with a lack of independence between variables (Dunn, 2019).

### 3.3.1 Principal Component Analysis (PCA)

The basic multivariate analysis method is the PCA, which is used for modeling and analyzing massive datasets. PCA is basically a reduction method that is used for large datasets by simplifying them into smaller sets of variables and keeping the large dataset information and pattern. So, by using the PCA method, complex, massive industrial data is driven by a small number of independent variables called latent variables (principal components) (Holland, 2008).

The main idea behind the PCA method is to use principal components as the new variables of the model to be constructed. These variables are constructed based on linear combinations of the original independent variables. The initial principal component explains the highest percentage of variance. Then, some of the remaining variability is explained by the second principal component, and so on. So, geometrically, the line direction of principal components is calculated to capture the highest variance of the raw data (Smith, 2002).

Figure 5 shows the schematic representation of the PCA method. The  $X$  matrix represents independent variables with  $k$  number of variables, and  $N$  represents the number of observations. The  $Y$  matrix consists of one column as PCA is used for a single dependent variable. The direction (loading) vectors are represented by a  $P$  matrix in which the  $k \times 1$  vector represents each latent variable direction vector. The second matrix that explains the latent variable is the scores matrix ( $T$ ). The  $T$  matrix consists of  $N \times A$  scores vectors as the model consists of  $A$  latent variables. So, each score vectors  $t_1, t_2, \dots, t_A$  has  $N \times 1$  vector values. Thus, latent variables are explained using two matrices:

P matrix and T matrix which are representing directions and scores respectively (Kettaneh et al., 2019).

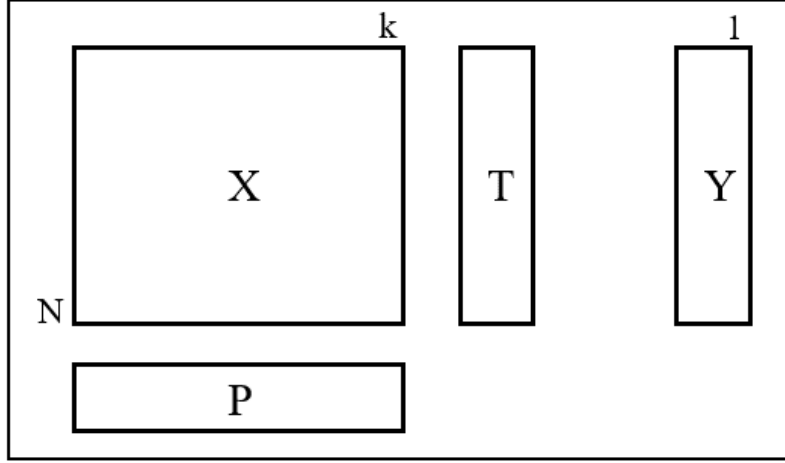


**Figure 5: Schematic representation of the PCA model**

#### **3.3.1.1 NIPALS algorithm for PCA method**

Different methods and algorithms are used to build the PCA model, and each method could be used to provide different insights and interpretations for the model. The most common three algorithms are Eigenvalue decomposition, singular value decomposition, and Non-linear Iterative Partial Least Squares (NIPALS) (Dunn, 2019). Most application packages and software use the NIPALS algorithm, the most convenient and helpful algorithm for interpreting the model built. Also, the missing data could be handled only using the NIPALS algorithm. NIPALS is an algorithm that calculates the latent variables sequentially. The NIPALS algorithm starts by preprocessing the raw data. The preprocessing of data is applied for each variable (i) from  $X_1$  to  $X_k$ , where each column is centered and scaled using the following equation:

$$X_j^* = \frac{X_j - X_{i,mean}}{X_{i,STD}} \quad (5)$$

Where  $X_j^*$  ( $j=1, 2, \dots, N$ ) are the preprocessed X values,  $X_j$  ( $j=1, 2, \dots, N$ ) are the original X values,  $X_{i,mean}$  ( $i=1, 2, \dots, k$ ) are mean values of each variable  $i$ , and  $X_{i,STD}$  ( $i=1, 2, \dots, k$ ) are standard deviation values of each variable  $i$ .

The primary purpose of preprocessing the data before building the model is to standardize the variables, ensuring variations around their respective means can be interpreted effectively. This step is crucial because, in most cases, variables are measured on different scales and in varying engineering units. Standardizing the data gives equal importance to all variables, preventing those with larger scales from dominating the model. Once the raw data has been preprocessed, the steps of the NIPALS algorithm, as illustrated in Figure 6, are applied.

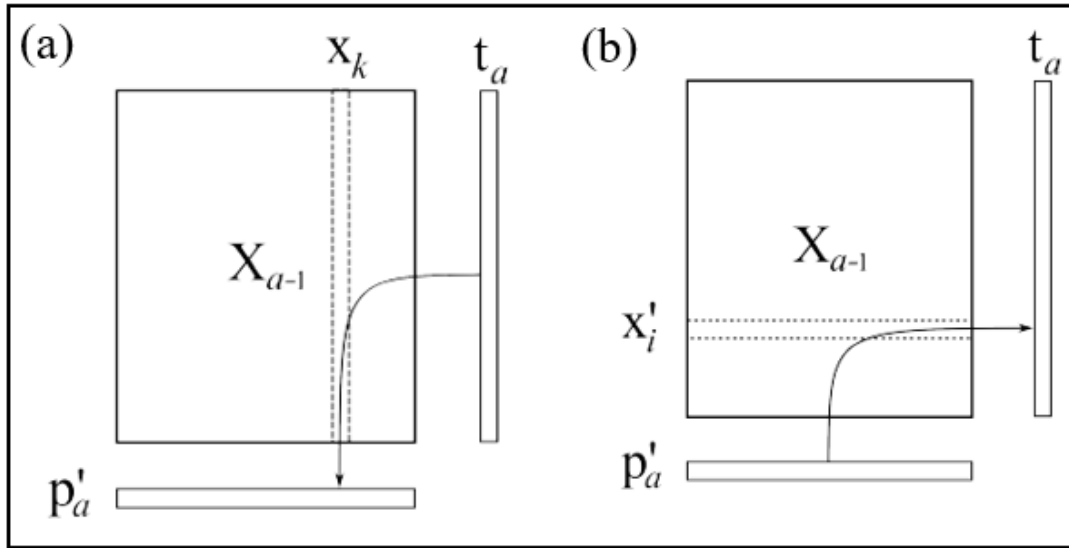


**Figure 6: NIPALS algorithm for PCA method: (a) Step 2, (b) Step 4**

1. First, the initial arbitrary column for  $t_a$  is selected by using random numbers or by taking any  $X_i$  variable column from the X matrix.

2. Then, take every column in the X matrix and call it  $X_k$  to regress it onto the initial column  $t_a$ . After that, the regression coefficient is stored as an entry to  $P_a'$ . Figure 6(a) explains this step.

This procedure is repeated for every column in the X matrix until the  $P_a'$  is filled with all  $1 \times k$  values. Normally,  $P_a'$  matrix is calculated in one step by using matrix multiplication as in Equation (6):

$$P_a' = \frac{1}{t_a^T t_a} t_a^T X_a \quad (6)$$

Where  $P_a'$  is the directions (loadings) vector ( $1 \times k$  row vector),  $t_a$  is the scores vector ( $N \times 1$  column vector),  $t_a^T$  is the transpose of  $t_a$  vector, and  $X_a$  is the preprocessed X matrix ( $N \times k$  matrix).

3. Rescale directions (loadings) vector  $P_a'$  to unit vector having a magnitude of 1. This is done by calculating vector magnitude as in Equation (7) and then dividing the vector by its magnitude to get the unit vector.

$$P_a^m = \sqrt{P_1^2 + P_2^2 + \dots + P_k^2} \quad (7)$$

Where  $P_a^m$  is the magnitude of  $P_a$  vector, and  $P_i$  ( $i=1, 2, \dots, k$ ) are the values of  $P_a$  vector.

4. Regress every row in the X matrix onto the normalized loading vector ( $P_a'$ ) and regression coefficients are stored in the scores vector ( $t_a$ ). Same as step 2, as shown in Figure 6(b),  $t_a$  vector is calculated in one step using matrix multiplication as in the following equation:

$$t_a = \frac{1}{P_a'^T P_a'} X_a P_a' \quad (8)$$

5. Keep iterating steps 2 to 4 until  $t_a$  vector changing  $< 10^{-6} - 10^{-9}$ .
6. Once  $t_a$  coverage and does not change, the  $t_a$  and  $P_a'$  values are stored in matrices T and P, respectively, as the first latent variable. Then, the deflation step is carried out by removing the variability explained by the first latent variable using the following equation:

$$E_a = X_a - t_a P_a' \quad (9)$$

Where  $E_a$  is the residual after using latent variable a (uncaptured variability) as in matrix  $N \times k$ . Then, the next latent variable calculation is carried out using this residual as in Equation (10).

$$X_{a+1} = E_a \quad (10)$$

Where  $X_{a+1}$  is the next X matrix with uncaptured variability as in matrix  $N \times k$ .

So,  $X_{a+1}$  matrix is used to calculate the next latent variable starting from step 1. This procedure is repeated until the number of latent variables is enough to explain the raw data. The number of latent variables needed for the model is determined by the cross-validation method. This method will be explained later in Section (3.3.3).

### 3.3.2 Projection to Latent Structure (PLS)

Projection to latent structure (PLS) is one of the other latent variable methods used for multivariate datasets. The PLS method is used for more than one dependent variable (Y) in the raw data, which PCA cannot handle. The PLS method can be used for different purposes, such as process monitoring,

performance optimization, and prediction of variables. This method is suitable for predicting anode quality variables using process parameters. The key advantage of PLS over other multivariate methods, such as multiple linear regression (MLR) and PCA, is that a single model could be built for multiple Y variables instead of building M number of models for all dependent variables. So, as Dunn (2019) suggested, PLS is one of the most convenient and useful methods to build a model for inter-correlated Y variables. In this research study, the number of Y variables is greater than one and are collinear. Therefore, the PLS method was chosen over the PCA method to build the model.

The PLS method is visualized using different matrices, as shown in Figure 7. As with the PCA method, independent variables are represented using an X matrix with a K number of variables. However, the Y matrix has M columns representing the number of dependent variables. In PLS, the loading and score vectors of the X block are  $W'$  and T matrices, respectively. Also, the Y block has loading and score vectors, which are  $C'$  and U matrices, unlike PCA, where Y has only one column. So, the A latent variables in the PLS method are represented by  $W'$ , T,  $C'$ , and U matrices (Kettaneh et al., 2019).

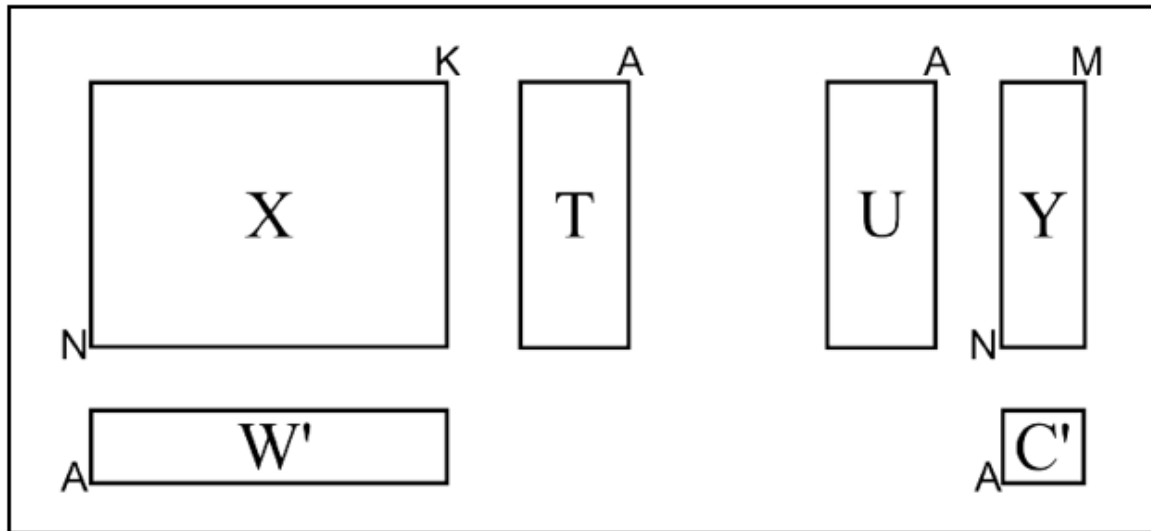


**Figure 7: Schematic representation of the PLS model**

### ***3.3.2.1 NIPALS algorithm for PLS method***

NIPALS algorithm is also used to calculate the PLS model, and the procedure is the same as PCA, but with some additional steps for block Y. Figure 8 shows the first five steps for calculating the PLS model. As with PCA, both matrices X and Y are preprocessed, as explained in subsection (3.3.1.1). Below this figure is the explanation of these five iteration steps, followed by the three next steps after completing the iteration steps to prepare data for the next latent variable calculation.

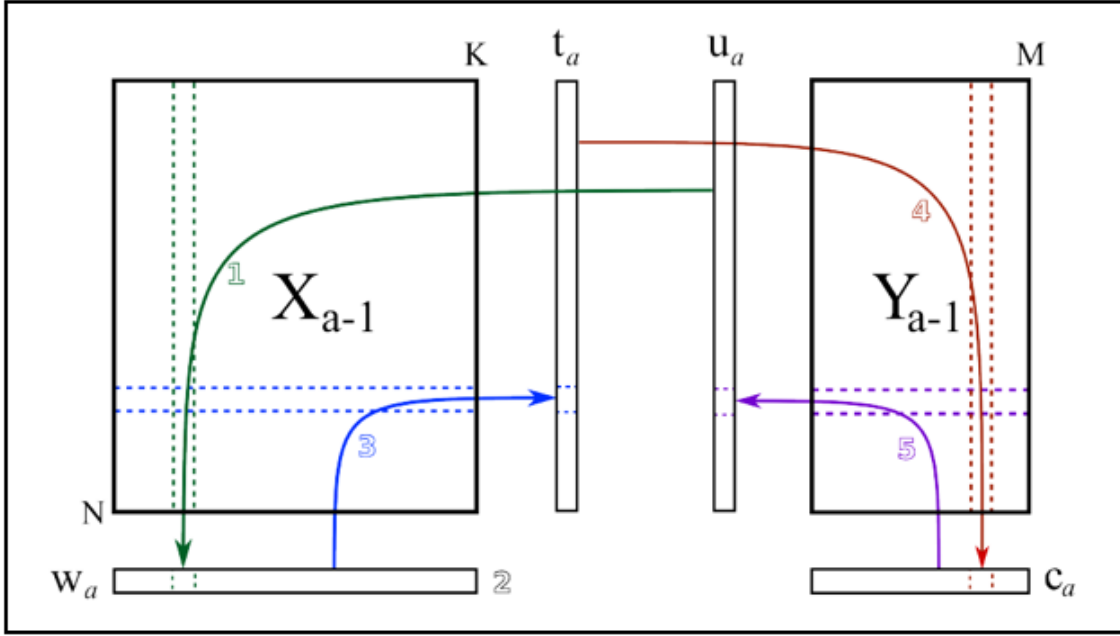


**Figure 8: NIPALS algorithm for PLS method**

1. Initial arbitrary column for  $u_a$  is selected by random values or by taking one column from the Y matrix. Then, regress every column in the X matrix onto the initial column  $u_a$  and store regression coefficients in  $w_a$ . Thus,  $w_a$  is calculated in one step using matrix multiplication as in the following equation:

$$w_a = \frac{1}{u_a^T u_a} X_a^T u_a \quad (11)$$

Where  $w_a$  is the loading vector of X block ( $1 \times k$  row vector),  $u_a$  is the scores vector of the Y block ( $N \times 1$  column vector),  $u_a^T$  is the transpose of  $u_a$  vector, and  $X_a^T$  the transpose of  $X_a$  matrix.

2. Rescale loading vector  $w_a$  to unit vector as in Equation (5). However,  $w_a$  notation is used for the PLS method instead of  $P_a'$  which is used for the PCA method.

3. Regress every row in the X matrix onto the normalized loading vector  $w_a$  and regression coefficients are stored in  $t_a$  vector. So,  $t_a$  vector is calculated in one step using the following equation:

$$t_a = \frac{1}{w_a^T w_a} X_a w_a \quad (12)$$

Where  $w_a^T$  is the transpose of  $w_a$  vector.

4. Then, regress every column in the Y matrix onto the calculated  $t_a$  vector and regression coefficients are stored in  $c_a$ . The following equation is used to calculate  $c_a$ :

$$c_a = \frac{1}{t_a^T t_a} Y_a^T t_a \quad (13)$$

Where  $c_a$  is the loading vector of the Y block ( $1 \times M$  row vector),  $Y_a$  is the preprocessed Y matrix ( $N \times M$  matrix), and  $Y_a^T$  is the transpose of  $Y_a$  matrix.

5. Last, each row in the Y matrix is regressed to  $c_a$  calculated to compute  $u_a$  by using the regression coefficients as entries:

$$u_a = \frac{1}{c_a^T c_a} Y_a c_a \quad (14)$$

Where  $c_a^T$  is the transpose of  $c_a$  vector.

6. Repeat the five steps until  $u_a$  vector does not change much. This is a calculation of one latent variable only. Values of  $w_a$ ,  $t_a$ ,  $c_a$ , and  $u_a$  are stored in latent variable (a).

After calculating the first latent variable, deflation steps (6 and 7) are performed to calculate the next latent variable.

7. A loading vector  $P_a$  for X space is calculated using the following equation:

$$P_a = \frac{1}{t_a^T t_a} X_a^T t_a \quad (15)$$

8. First, the predicted variability of  $X_a$  is calculated. Secondly, this predicted variability is removed from  $X_a$  to get the residual after latent variable (a). Lastly, the calculated residual is used as the next X matrix values. Equations (16), (17), and (18):

$$\hat{X}_a = t_a P_a' \quad (16)$$

Where  $\hat{X}_a$  is the predicted variability of  $X_a$  ( $N \times K$  matrix).

$$E_a = X_a - \hat{X}_a \quad (17)$$

$$X_{a+1} = E_a \quad (18)$$

The Y block is treated in same way exactly using equation (19), (20), and (21).

$$\hat{Y}_a = t_a c_a' \quad (19)$$

Where  $\hat{Y}_a$  is the predicted variability of  $Y_a$  ( $N \times M$  matrix).

$$F_a = Y_a - \hat{Y}_a \quad (20)$$

$$Y_{a+1} = F_a \quad (21)$$

Where  $F_a$  is the residual after using latent variable a of Y block as in matrix  $N \times M$ , and  $Y_{a+1}$  is the next Y matrix with uncaptured variability as in matrix  $N \times M$ .

### 3.3.3 Cross-Validation to Estimate the Number of Latent Variables

In PLS and PCA, the explained variability by the model would be increased by increasing the latent variables of the model. However, increasing the number of latent variables more than needed could lead to over-fitting, explaining noise and unstructured variance. Wold (1978) has introduced a method to select the correct number of latent variables necessary for the model. This method is called the Cross-Validation (CV) method. CV method adds latent variables to the model until adding more variables does not improve the predictivity of the model.

In the CV method, matrices  $X$  and  $Y$  are divided into  $G$  groups of rows. Each group of  $G$  has several rows, and rows that belong to group 1 are called  $X_{(1)}$  and  $Y_{(1)}$ , and all other remaining rows are called  $X_{(-1)}$  and  $Y_{(-1)}$ . In same way, groups 1, 2, ...,  $G$  are called  $X_{(1)}$ ,  $X_{(2)}$ , ...,  $X_{(G)}$ . In the random  $g^{\text{th}}$  group,  $X_{(g)}$  and  $Y_{(g)}$  represent rows of the  $g$  group, and  $X_{(-g)}$  and  $Y_{(-g)}$  represent all rows except the  $g$  group rows. The procedure of CV method is to build model on  $X_{(-1)}$  and  $Y_{(-1)}$  and use  $X_{(1)}$  and  $Y_{(1)}$  data to test the model as in following equation:

$$\hat{Y}_{(1)} = T_{(1)}C' \quad (22)$$

$$F_{(1)} = Y_{(1)} - \hat{Y}_{(1)} \quad (23)$$

Similarly, the remaining groups residual  $F_{(g)}$  ( $g=1, 2, \dots, G$ ) are calculated. Then, the calculated residuals ( $F_1, F_2, \dots, F_G$ ) are gathered in one matrix called  $F_{A,CV}$  where  $A$  represents the number of latent variables used and  $CV$  represents the cross-validation method used for calculation of the residuals matrix  $F$ . Finally, Eq. (24) is used to calculate CV variance:

$$Q_A^2 = 1 - \frac{Var(F_{A,CV})}{Var(Y)} \quad (24)$$

Where  $Q_A^2$  is the CV variance using the number of latent variables  $A$ ,  $Var(F_{A,CV})$  is the variance of the residuals matrix calculated using the CV method, and  $Var(Y)$  is the variance of the preprocessed  $Y$  matrix.

However, model (dependent variables) variance ( $Q_Y^2$ ) are calculated using all the rows of data ( $N$ ) using the following equation:

$$R_Y^2 = 1 - \frac{Var(F_A)}{Var(Y)} \quad (25)$$

Where  $R_Y^2$  is the model variance using the number of latent variables  $A$ , and  $Var(F_A)$  is the variance of the residuals' matrix using all  $N$  rows. Also, the independent variables variance ( $R_X^2$ ) are calculated in the same way as  $R_Y^2$ , but with using the  $X$  block matrices:  $P'$ ,  $\hat{Y}$  and  $E$ .

The model variance ( $R_Y^2$ ) will keep increasing as more LVs are added to the model, while CV variance ( $Q_A^2$ ) will decrease after a certain number of LVs. This is because the newly added LVs are not systematic and explain the noise of data. However, the best number of latent variables needed for datasets is not easy to judge because dataset types are different. So, a few extra LVs should be added to the model, and based on the calculated  $R_Y^2$  and  $Q_A^2$  values, judgment should be made (Dunn, 2019).

### 3.3.4 Statistical Methods Analysis of the Model Built

Previous studies used different quantitative measures and statistical methods to evaluate the significance of independent variables in predicting the dependent variable in the model built. Lazraq and Cl  roux (2001) have

stated that meeting statistical tests and testing the significance of the results is mandatory to accept or reject the hypothesis related to the significance of an independent variable in the model built. Therefore, the t-test is used to test the significance of an independent variable based on its coefficient. The t-test is used to determine whether there is a significant difference between the means of two groups or between a sample mean and a known population mean. It involves using the t-distribution to calculate a p-value.

Following Eriksson and Wold (2008) methodology, statistical t-test analysis (analysis of variance) will be performed on the independent variables to test the significance of these variables. Statistical significance was set at a 5% level, which results in a percentage of error related to accepting or rejecting the null hypothesis. The t-test will test the null hypothesis that an independent variable does not affect a dependent variable against the alternative hypothesis that an independent variable significantly affects a dependent variable. The p-value of the t-test is calculated during data fitting and indicates the probability of observing the given coefficient under the null hypothesis. If the value is less than the significance level, the null hypothesis will be rejected, implying the significance of the independent variable. Otherwise, the null hypothesis will be accepted, implying that there is no effect on the independent variable.

### **3.3.5 Prediction of New Dataset Using PLS Model Built**

This section explains and illustrates the procedure to predict new dataset-dependent variables using the existing built PLS model. In order to verify and validate the accuracy and ability of the model built, a new dataset that is not used for building the model is used. The previous sections explained building a model using a set of data called the training set of data. The data

used to predict the dependent variables or validate the model built is called a prediction or validation set. As the dataset available is massive and the model is complex, James et al. (2013) have suggested having 70% to 80% of the dataset be used for the training of data and 20% to 30% for prediction. Therefore, 75% of the data collected will be used for training purposes and 25% for prediction and validation of the model.

The first step in predicting a new dataset is to select the 25% of the dataset collected that will be used for prediction. The most crucial aspect in selecting the prediction set is to ensure randomness to avoid biases and to ensure that any inferences made from the observations are representative of the entire dataset. Juster (2013) proposed a procedure to be followed using Excel to ensure randomness of the selected portion of the dataset using the Excel function “*Rand*”. This procedure has been followed to select the prediction portion of the dataset collected. After that, the selected portion of the dataset is ready to be used for prediction purposes without worries of biases.

After building the PLS model, different steps and calculations should be performed to predict a new dataset. First, a new loadings matrix is introduced, and an  $R$  matrix is calculated using the two training loading matrices calculated during building the model, which are  $W$  and  $P'$ . Equation (26) presents the formula used to calculate the  $R$  matrix:

$$R = W(P'W)^{-1} \quad (26)$$

Where  $R$  is the matrix of the  $r_a$  loading vectors,  $W$  is the matrix of the  $w_a$  loading vectors and  $P'$  is the matrix of the  $P_a$  loading vectors. Secondly,

the new scores matrix is calculated for the new dataset of X variables using Equation (27):

$$T_{new} = X_{new}R \quad (27)$$

Where  $T_{new}$  is the matrix of the new score vectors based on the new dataset,  $X_{new}$  is the matrix of the new (prediction) dataset of X variables. Lastly, the predicted dependent variables are calculated using the new scores matrix and the training data loading matrix as in Equation (28):

$$\hat{Y}_{pred} = T_{new}C' \quad (28)$$

Where  $\hat{Y}_{pred}$  is the matrix of the predicted dependent variables. After that, the predicted and observed dependent variables could be compared to validate the model's ability to predict new data. The variance between predicted and observed dependent variables data is calculated for all dependent variables to validate prediction ability.

### 3.3.6 Model Interpretation Tools

After building the PCA or PLS model, different methods and tools are used to interpret the model built and its findings. One of these tools is the score plot, which is used to capture any clustering of data, outlier observations, and time-based trends in the raw data. The score value represents the distance from the origin along the direction vector of the latent variable to the point where the observation projects onto the direction vector. So, the weight of each observation is represented by a point in the score plot. The first two latent variables are used in the score plot as the highest variance of the model captured by them. The score plot is used to see similar patterns in portions of data, as groups of observations cluster together because of a similar combination of process parameters or raw material properties. Thus,

the relationship between observations is investigated using the score plot (Worley et al., 2013).

The second plot used to interpret the model is the loading plot. First, the loading (direction) value is used to describe how much a specific variable contributes to a specific latent variable. So, any model is going to have a number of directions equal to the number of latent variables. Loading plot is plotted using two pairs of latent variables (usually the first two LVs). Then, the plot is used to interpret the correlation between independent variables and dependent variables. The loading plot is interpreted by capturing the sign of variables' loadings and the location of variables in the plot. Variables located near each other and had loadings with the same signs are positively correlated. On the other hand, variables located far from each other and had loadings with different signs are negatively correlated (Ergon, 2004).

The third method used to interpret the PLS model is Hotelling's ( $T^2$ ) plot. The Hotelling's value is the distance from the origin of the model space to the observation projection in the model plane. So, Hotelling's plot shows these distances for N observations in the data. The main use of Hotelling's plot is to detect the outlier observations. Equation (29) is used to calculate the Hotelling's value for each observation in the model. Hotelling's plot is useful for determining outliers with large  $T^2$  values. An observation with a high  $T^2$  value indicates that the given observation is far from other observations and could be considered an outlying observation (Mehmood, 2016).

$$T_i^2 = \sum_{a=1}^A \frac{t_{ia}^2}{s_{ta}^2} \quad (29)$$

Where  $T_i^2$  is the Hotelling's distance of an observation,  $t_{ia}$  is the score of an observation for the latent variables  $a$ , and  $s_{ta}$  is the variance of  $t_a$ .

The other tool used to detect outlier observations in the dataset is the DmodX plot. DmodX value is the distance of an observation to its residual on the latent variable space. Observations with high DmodX values indicate differences from other observations in terms of the correlation structure. Equation (30) is used to calculate the DmodX value of an observation:

$$DmodX_i(a) = \sqrt{\frac{\sum_{j=1}^J e_{ij}(a)^2}{J - A}} \quad (30)$$

Where  $DmodX_i(a)$  is the residual of an observation for a latent variables space, and  $e_{ij}(a)$  is the residual of an observation of variable  $j$  obtained from a model built using a latent variables.

Silva et al. (2019) have set a rule of thumb for determining outlier observations in PCA and PLS models. An observation with DmodX or Hotelling's ( $T^2$ ) value higher than the critical value of 95% of the F-distribution of data by twice is considered an outlier observation. This indicates that such an observation is different from normal observations with respect to the correlation structure of the variables.

Farrés et al. (2015) have explained another tool that could be used to discover the important variables in the PLS model is the Variable Importance in the Projection (VIP) method. This tool indicates the importance of X variables in predicting Y variables. The higher the VIP value of the X variable, the higher the importance of that variable in predicting Y variables. Equation (31) is used to calculate the VIP value for the X variable. Dunn (2019) has

suggested that X variable with a VIP value greater than one is important to predict the model.

$$VIP_{j,A} = \sqrt{\sum_{a=1}^A w_{j,a}^2 \left( \frac{SSY(a)}{SSY(tot)} \right)} \quad (31)$$

Where  $VIP_{j,A}$  is the importance index of variable j,  $w_{j,a}$  is the loading weight of variable j in the  $a^{th}$  latent variable,  $SSY(a)$  is the sum of squares of Y explained by the  $a^{th}$  latent variable, and  $SSY(tot)$  is the sum of squares of Y explained by the model.

The VIP scores for each X variable could be used to rank their importance. The variables with VIP values higher than one are considered important for predicting the Y variables. Moreover, the top ten variables are used most to find the best set of Y variables.

The VIP plot gives the most important variables in predicting Y-variables, while the loading plot indicates the most significant variables and their contribution to specific latent variables. Therefore, the loading plot helps identify significant variables by showing how strongly and in what direction each variable contributes to the latent variables, indicating their influence on data variation. However, the VIP plot determines important variables by quantifying their overall contribution to the model, with higher VIP scores indicating greater importance in predicting the dependent variables. So, the terms significant variables and important variables each have distinct interpretations in the PLS method (Farrés et al., 2015).

### 3.3.7 Net Carbon Consumption Reduction

Net carbon consumption refers to the amount of carbon consumed during the aluminum smelting process, including all the sources of consumption. In other words, the number of kilograms of carbon consumed to produce one ton of aluminum.

The main consumption component is the theoretical carbon consumption source, which is related to the chemical stoichiometry of the electrochemical equation shown in Equation (1), which is going to be 334 kg of carbon consumed for every ton of aluminum produced. The other components are related to the non-theoretical carbon consumption, which is related to the anode quality, process conditions, and potline efficiency parameters. Therefore, achieving 334 kg of NCC is not possible, and modern aluminum smelting companies are achieving 400-500 kg of carbon per 1 ton of aluminum. Thus, most companies spend a lot of money and effort to reduce the NCC as it reflects an indicator of performance.

Khaji and Al-Qassemi (2016) have suggested Equation (32) to estimate the NCC value based on anode geometric density:

$$NCC = 334 + \alpha \left( \frac{1}{\rho_a} \right) \quad (32)$$

Where  $\alpha$  is the empirical equation constant, and  $\rho_a$  is the anode geometric density. The equation constant ( $\alpha$ ) can be calculated for any aluminum smelter by knowing the NCC values and the anode geometric density. Then, this  $\alpha$  could be used to estimate NCC based on the built model using the predicted density of the best set of variables.

### **3.4 Research Tools**

The research study was performed using two software programs: Excel and SIMCA. All data extracted from Aluminium Bahrain (ALBA) were transferred to Excel spreadsheets. Then, Excel software was used to filter and prepare data for the model. SIMCA software was used to build and interpret the model.

SIMCA stands for Soft Independent Modelling by Class Analogy. This software is developed by Umetrics Company. Using SIMCA, appropriate models of PLS and PCA are built, along with interpretation tools.

### **3.5 Data Collection**

In this section, the procedure followed during data collection is summarized for all the stages. Before the collection of data needed for the study, ethical clearance of the Carbon department and Laboratory analysis data was sought from ALBA. The data collected include raw material properties, paste plant parameters, baking kiln parameters, and anode core quality results. The dataset collected was from November 2020 to October 2022. The carbon plant data used was for carbon plant 2 because this plant, during the period of collection, had stable production and available data access. The data was extracted from ALBA domains during November 2022 and February 2023.

Raw material laboratory analyses data were collected from laboratory analyses Excel sheets, which are stored in the database daily for calcined petroleum coke and weekly basis for butts and pitch. Paste plant process parameters data were collected from the ALBA database and updated every minute. Baking kiln process parameters are available on a weekly average

basis only. Lastly, anode core quality results were collected from laboratory analyses Excel sheets in ALBA domains, which are updated daily. Data collection procedures and treatment of these data before model building are described and explained in this section.

Before data collection, strategy and procedure should be set to trace and synchronize the data correctly, as there are many complications related to the identification of the anode cores analyzed. Each anode core has two identification numbers: anode production number and sample ID number. As mentioned in Section (2.2) and shown in Figure 2, raw materials are mixed to produce the green anode. After green anode production, the anode is stamped with an anode production number. This anode production number represents that anode produced along with its process parameters. So, the anode paste plant process parameters are traced through the anode production number. Then, green anodes with their production numbers are stored in the storage area for at least one day.

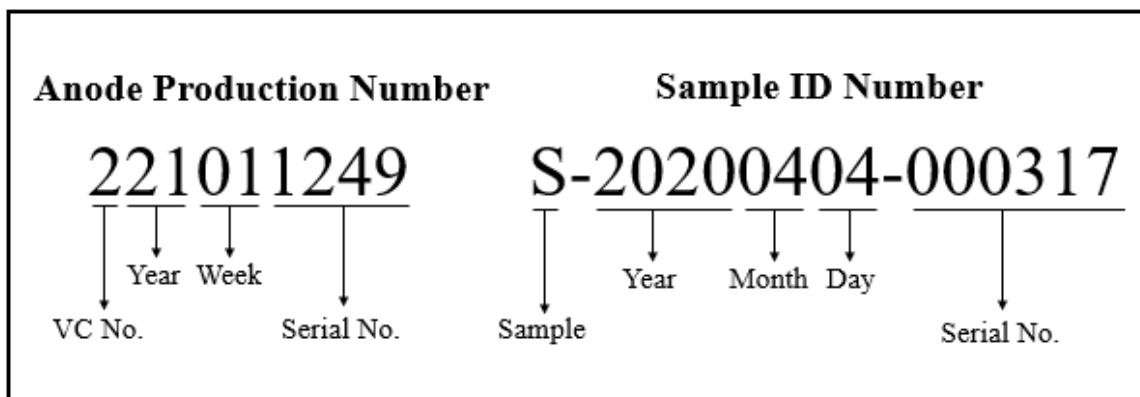


**Figure 9: Anode production number and sample ID number breakdown**

After baking the green anodes, they are cored and sent to a laboratory for analysis. After coring, sample ID numbers are set for these anode cores. Every day in Carbon 2, twelve anodes are cored and sent to the laboratory

randomly. Only five out of these twelve anode cores are analyzed to get their quality results. So, synchronization and data collection start with anode quality results using sample ID numbers.

### **3.5.1 Anode Quality Results Data**

The quality of the daily five anode cores is tested with different chemical and physical properties. However, not every anode core is tested with all the chemical and physical properties listed in Table 6 because the core samples are too small to perform all tests, and manpower and machine capabilities cannot perform all tests for all anode cores. So, the ALBA laboratory follows a standardized testing plan to perform different tests on different anode cores received.

Three physical properties tests, geometric density, specific electrical resistance, and flexural strength, are performed for all five daily anode cores. On the other hand, several chemical and reactivity analyses are performed only for a maximum of five anode cores per week.

Hence, two PLS models can be built for the dataset of anode quality results. The first model is the physical properties model consisting of geometric density, specific electrical resistance, and flexural strength. The second model is the chemical and reactivity properties model consisting of CO<sub>2</sub> reactivity residue, sulfur (S) content, and iron (Fe) content, as this will help in using all the massive data of the two years for the first model instead of using less than 20% of the data for building one model for all properties. Moreover, the second model will study the chemical and reactivity properties using the anode cores tested for all properties.

The anode quality results in data collection start by removing all untested anodes. Then, two sets of data are extracted for the two models. For the two-year study, the dataset available for the first and second models is 3344 and 587 observations, respectively. Appendix (1A) shows the results of the anode properties before filtration for the two models. Appendix (2) shows a portion of data for the two models after filtration and preparation; these prepared data are ready for the next stages of collection.

### **3.5.2 Paste Plant Process Parameters Data**

The next stage of the collection after anode quality results in data is the paste plant parameters data. The maximum capacity of the carbon 2 paste plant is 864 anodes daily. Mostly, a lower number of green anodes is produced due to failures and maintenance in the plant. An average of five anodes are analyzed out of daily green anodes produced. Thus, the anode production number of analyzed anodes is used to filter needed green anode process parameters. During this stage of data collection, all the analyzed anodes paste plant process parameters are extracted from the massive dataset of the two-year research study period.

The paste plant process parameters collected are listed in Table 4, which are the same for both models to be built. The data collected in this stage are synchronized with anode quality results to be in one table, as shown in Appendix (3). The data collected and filtered in this appendix are ready for raw material properties data collection and filtration.

### **3.5.3 Raw Material Properties Data**

The raw material properties include butts, fines, coal tar pitch (CTP), and calcined petroleum coke (CPC) properties. Butts, fines, and CTP

properties are available weekly, and CPC properties daily. The anode production number is used to synchronize raw material data with data collected in previous stages. Butts, fines, and CTP data are extracted by using a week of production from the anode production number, as shown in Figure 9.

Based on discussions with plant process controllers and engineers, the raw materials data are synchronized with the data collected. Each week, analysis data is used for the anodes produced that week. For example, butts, fines, and CTP data from week 42 are used for anodes produced on week 42.

CPC data are synchronized with anode data daily. According to process engineers, the CPC data should be used two weeks before the green anode is produced for 2020 and 2021 production. The CPC data will be used one week before the 2022 anode production. At the end of this stage, the data collected and synchronized are ready for baking kiln process data collection.

#### **3.5.4 Baking Kiln Process Parameters Data**

The last stage of carbon anode collection data is the baking kiln parameters data collection. For baking kiln data, the sample ID number is used to synchronize the data with other anode-collected data. The baking kiln process parameters used for the model are cycle time, soaking time, final anode temperature, and heating rate. The data collected for the baking kiln is available every week. The data from one week before the sample ID number are synchronized for the first three process parameters (cycle time, soaking time, and final anode temperature).

For heating rate data, two weeks before the sample ID number date is used. This is because the heating rate is conducted at the beginning of the

anode baking cycle, and other parameters are conducted at the end of the cycle. At the end of this stage, the data collected is ready to be used to build the model. The model variables are listed in Tables 8 and 9, and a portion of the final data collected is shown in Appendix (4).

**Table 8: Physical and chemical models' dependent variables**

| <b>Y Variables</b> | <b>Variable ID</b> | <b>Variable</b>                    | <b>Unit</b>          |
|--------------------|--------------------|------------------------------------|----------------------|
| <b>Model 1</b>     | $y_1$              | Geometric Density                  | $\text{g/cm}^3$      |
|                    | $y_2$              | Specific Electrical Resistance     | $\mu\Omega\text{cm}$ |
|                    | $y_3$              | Flexural Strength                  | MPa                  |
| <b>Model 2</b>     | $y_4$              | CO <sub>2</sub> Reactivity Residue | %                    |
|                    | $y_5$              | Sulfur (S) Element Content         | %                    |
|                    | $y_6$              | Iron (Fe) Element Content          | ppm                  |

**Table 9: Physical and chemical models' independent variables**

| <b>X Variables</b> | <b>Variable</b>                | <b>Unit</b>     |
|--------------------|--------------------------------|-----------------|
| $x_1$              | Green Anode Density (GAD)      | $\text{g/cm}^3$ |
| $x_2$              | Pitch level                    | %               |
| $x_3$              | Mixing power                   | kW              |
| $x_4$              | Cooling power                  | kW              |
| $x_5$              | Mixing temperature             | °C              |
| $x_6$              | Dry aggregate temperature      | °C              |
| $x_7$              | Anode forming temperature      | °C              |
| $x_8$              | Green scrap content            | %               |
| $x_9$              | Butts (very coarse) percentage | %               |
| $x_{10}$           | Fine particles percentage      | %               |
| $x_{11}$           | Medium particles percentage    | %               |
| $x_{12}$           | Coarse particles percentage    | %               |
| $x_{13}$           | Anode forming speed            | rpm             |
| $x_{14}$           | Anode forming time             | Seconds         |

|          |                                    |          |
|----------|------------------------------------|----------|
| $x_{15}$ | Butts iron (Fe) element content    | ppm      |
| $x_{16}$ | Butts sodium (Na) element content  | ppm      |
| $x_{17}$ | Fines iron (Fe) element content    | ppm      |
| $x_{18}$ | Fines silicon (Si) element content | ppm      |
| $x_{19}$ | Fines sodium (Na) element content  | ppm      |
| $x_{20}$ | Pitch softening point              | °C       |
| $x_{21}$ | Pitch coking value                 | %        |
| $x_{22}$ | Pitch sulfur (S) element content   | %        |
| $x_{23}$ | CPC sulfur (S) element content     | %        |
| $x_{24}$ | CPC vanadium (V) element content   | ppm      |
| $x_{25}$ | CPC iron (Fe) element content      | ppm      |
| $x_{26}$ | CPC CO <sub>2</sub> reactivity     | %        |
| $x_{27}$ | CPC crystallite size ( $L_c$ )     | Å        |
| $x_{28}$ | CPC coarse content (>4.75 mm)      | %        |
| $x_{29}$ | Fire cycle time                    | Hours    |
| $x_{30}$ | Fire soaking time                  | Hours    |
| $x_{31}$ | Final anode temperature            | °C       |
| $x_{32}$ | Heating rate                       | °C/hours |

### 3.6 Chapter Conclusion

In this chapter, the methodologies used to build a multivariate analysis model were discussed. Two popular multivariate analysis methods, PCA and PLS, were investigated and explained to use them for building the model. These methods will be used to develop a model that examines the relationship between anode quality properties, raw material properties, and process parameters. Different model tools will also be used to interpret the model findings.

In the second part of the chapter, the procedures followed in collecting data to build the physical and chemical models are illustrated. The four stages

of data collection are discussed. These stages contain the procedure and strategy for collecting data on anode quality results, paste plant process parameters, raw material properties, and baking kiln process parameters.

## **Chapter 4: Results and Discussion**

|                                                               |            |
|---------------------------------------------------------------|------------|
| <b>4.1 Introduction.....</b>                                  | <b>71</b>  |
| <b>4.2 Discrete and Continuous Independent Variables.....</b> | <b>71</b>  |
| <b>4.3 Physical Properties Model.....</b>                     | <b>73</b>  |
| <b>4.4 Chemical and Reactivity Properties Model .....</b>     | <b>101</b> |
| <b>4.5 Net Carbon Consumption of PLS Model Built.....</b>     | <b>122</b> |
| <b>4.6 Chapter Conclusion .....</b>                           | <b>123</b> |

## **4.1 Introduction**

This chapter analyses and discusses the research study's results, including the models built, the most important process parameters, and the significant relationship between anode quality variables and process parameters variables. The analysis and discussion are performed for both models: the physical model and the chemical and reactivity model.

The first part of the chapter analyses the physical properties model, and the second part analyses the chemical and reactivity properties model.

## **4.2 Discrete and Continuous Independent Variables**

The thirty-two independent variables are collected in different ways depending on the database. Some variables are measured directly in the plant using measuring devices, while others are set by the plant controller based on specialist requests. Additionally, some variables are analyzed and measured in a laboratory. All the independent variables are of a continuous scale and nature.

However, despite being continuous, some of these variables are treated as discrete variables because the plant controller sets them to operate at specific fixed levels or intervals. This treatment reflects the system's practical constraints, where variables are adjusted in stepwise increments rather than continuously, influencing how the model interprets and incorporates these variables. These collection methods result in the availability of two types of independent variables: discrete variables and continuous variables. Table 10 shows the list of discrete and continuous variables.

**Table 10: Independent variables type: Discrete or Continuous variables**

| <b>Variable ID</b> | <b>Variable</b>                    | <b>Type</b> |
|--------------------|------------------------------------|-------------|
| $x_1$              | Green anode density (GAD)          | Continuous  |
| $x_2$              | Pitch level                        | Discrete    |
| $x_3$              | Mixing power                       | Continuous  |
| $x_4$              | Cooling power                      | Continuous  |
| $x_5$              | Mixing temperature                 | Continuous  |
| $x_6$              | Dry aggregate temperature          | Continuous  |
| $x_7$              | Anode forming temperature          | Continuous  |
| $x_8$              | Green scrap content                | Discrete    |
| $x_9$              | Butts (very coarse) percentage     | Discrete    |
| $x_{10}$           | Fine particles percentage          | Discrete    |
| $x_{11}$           | Medium particles percentage        | Discrete    |
| $x_{12}$           | Coarse particles percentage        | Discrete    |
| $x_{13}$           | Anode forming speed                | Discrete    |
| $x_{14}$           | Anode forming time                 | Discrete    |
| $x_{15}$           | Butts iron (Fe) element content    | Continuous  |
| $x_{16}$           | Butts sodium (Na) element content  | Continuous  |
| $x_{17}$           | Fines iron (Fe) element content    | Continuous  |
| $x_{18}$           | Fines silicon (Si) element content | Continuous  |
| $x_{19}$           | Fines sodium (Na) element content  | Continuous  |
| $x_{20}$           | Pitch softening point              | Continuous  |
| $x_{21}$           | Pitch coking value                 | Continuous  |
| $x_{22}$           | Pitch sulfur (S) element content   | Continuous  |
| $x_{23}$           | CPC sulfur (S) element content     | Continuous  |
| $x_{24}$           | CPC vanadium (V) element content   | Continuous  |
| $x_{25}$           | CPC iron (Fe) element content      | Continuous  |
| $x_{26}$           | CPC CO <sub>2</sub> reactivity     | Continuous  |
| $x_{27}$           | CPC crystallite size ( $L_c$ )     | Continuous  |
| $x_{28}$           | CPC coarse content (>4.75 mm)      | Continuous  |
| $x_{29}$           | Fire cycle time                    | Discrete    |
| $x_{30}$           | Fire soaking time                  | Discrete    |
| $x_{31}$           | Final anode temperature            | Continuous  |
| $x_{32}$           | Heating rate                       | Continuous  |

## **4.3 Physical Properties Model**

For the physical model, the final data collected in Appendix (4A) are tested and filtered to remove all invalid observations due to system or database errors. Each independent variable is also tested to check its relationship with dependent variables. Variables that do not contribute to the model were removed from the data.

### **4.3.1 Discrete Independent Variables of the Physical Properties Model**

Before building the physical properties model, discrete variables shown in Table 10 are ordered numerically based on their values. Therefore, the discrete variables are treated ordinally to model the levels as a single continuous variable for each discrete variable. The discrete variables, their values, and represented ordinal levels are shown in Tables 11 to 14. The first three tables represent the paste plant process parameters, and the last represents the baking kiln process parameters. The paste plant process parameters have been split into three tables because they have many variable levels.

A variable with a missing value is represented, and the ordinal level is left blank. For example, the fine particle percentage of 19% is represented with the ordinal level of 2, even though there is no such observation in this variable that has a value of 19%. This allows for the representation of this value in predicting new datasets that may have this value. These discrete variables are not preprocessed like other continuous variables to ensure that ordinal data are treated correctly.

**Table 11: Pitch level and green scrap content variables values and represented levels of the physical properties model**

| Variable                      | Variable value | Represented level |
|-------------------------------|----------------|-------------------|
| Pitch level ( $x_2$ )         | 12.9           | 1                 |
|                               | 13.0           | 2                 |
|                               | 13.1           | 3                 |
|                               | 13.2           | 4                 |
|                               | 13.3           | 5                 |
|                               | 13.4           | 6                 |
|                               | 13.5           | 7                 |
|                               | 13.6           | 8                 |
|                               | 13.7           | 9                 |
|                               | 13.8           | 10                |
|                               | 13.9           | 11                |
| Green scrap content ( $x_8$ ) | 0              | 1                 |
|                               | 0.5            | 2                 |
|                               | 1.0            | 3                 |
|                               | 1.5            | 4                 |
|                               | 2.0            | 5                 |
|                               | 2.5            | 6                 |
|                               | 3.0            | 7                 |
|                               | 3.5            | 8                 |

**Table 12: Butts, fine and medium particles percentage variables values and represented levels of the physical properties model**

| Variable                                    | Variable value | Represented level |
|---------------------------------------------|----------------|-------------------|
| Butts (very coarse)<br>percentage ( $x_9$ ) | 18             | 1                 |
|                                             | 20             | 3                 |
|                                             | 21             | 4                 |
|                                             | 22             | 5                 |
|                                             | 23             | 6                 |
|                                             | 24             | 7                 |
|                                             | 25             | 8                 |
|                                             | 26             | 9                 |
|                                             | 27             | 10                |
|                                             | 28             | 11                |
|                                             | 30             | 13                |
|                                             | 32             | 15                |
| Fine particles<br>percentage ( $x_{10}$ )   | 18             | 1                 |
|                                             | 20             | 3                 |
| Medium particles<br>percentage ( $x_{11}$ ) | 21             | 1                 |
|                                             | 22             | 2                 |
|                                             | 23             | 3                 |
|                                             | 24             | 4                 |
|                                             | 25             | 5                 |
|                                             | 26             | 6                 |
|                                             | 27             | 7                 |
|                                             | 28             | 8                 |
|                                             | 29             | 9                 |
|                                             | 30             | 10                |

**Table 13: Coarse particles percentage, anode forming speed and anode forming time variables values and represented levels of the physical properties model**

| Variable                                 | Variable value | Represented level |
|------------------------------------------|----------------|-------------------|
| Coarse particles percentage ( $x_{12}$ ) | 25             | 1                 |
|                                          | 26             | 2                 |
|                                          | 27             | 3                 |
|                                          | 28             | 4                 |
|                                          | 29             | 5                 |
|                                          | 30             | 6                 |
|                                          | 31             | 7                 |
|                                          | 33             | 9                 |
|                                          | 35             | 11                |
| Anode forming speed ( $x_{13}$ )         | 1101           | 1                 |
|                                          | 1120           | 20                |
|                                          | 1122           | 22                |
| Anode forming time ( $x_{14}$ )          | 65             | 1                 |
|                                          | 68             | 4                 |

**Table 14: Fire cycle time and fire soaking time variables values and represented levels of the physical properties model**

| Variable                       | Variable value | Represented level |
|--------------------------------|----------------|-------------------|
| Fire cycle time ( $x_{29}$ )   | 24             | 1                 |
|                                | 25             | 2                 |
|                                | 26             | 3                 |
|                                | 28             | 5                 |
| Fire soaking time ( $x_{30}$ ) | 41             | 1                 |
|                                | 42             | 2                 |
|                                | 45             | 5                 |
|                                | 46             | 6                 |
|                                | 50             | 10                |

### 4.3.2 Exclusion of Non-contributory Independent Variables

Before building the model, the independent variables were tested to evaluate their relationships with the dependent variables. Two independent variables, cooling power ( $x_4$ ) and mixing temperature ( $x_5$ ) showed fragile relationships with all dependent variables, as shown in Table 15. Therefore, these two independent variables were excluded from the model. In the following sections, the significance and effect of the other independent variables on the dependent variables will be investigated and tested after building the model.

**Table 15: Cooling power and mixing temperature relationship strength with dependent variables**

| <b>X-variable</b>  | <b>Y-variable</b>              | <b>Variance (<math>R^2</math>)</b> |
|--------------------|--------------------------------|------------------------------------|
| Cooling power      | Geometric density              | 3.8%                               |
|                    | Specific electrical resistance | 2.3%                               |
|                    | Flexural strength              | 2.9%                               |
| Mixing temperature | Geometric density              | 1.7%                               |
|                    | Specific electrical resistance | 0.1%                               |
|                    | Flexural strength              | 4.4%                               |

### 4.3.3 DmodX and Hotelling's ( $T^2$ ) plots

The number of observations of the final data collected before removing invalid observations was 3344 observations. The number of observations was reduced to 3133 after all invalid observations had been removed. The model was built on 75% of those reduced observations (2343 observations), and outlier observation detection methods were used for that model. These methods are DmodX and Hotelling's ( $T^2$ ) plots. DmodX and Hotelling's ( $T^2$ ) plots are shown in Figures 10 and 11, respectively. Outlier observations are

detected using these two methods. As these outliers affect the model variance negatively, the outlier observations are excluded from the model.

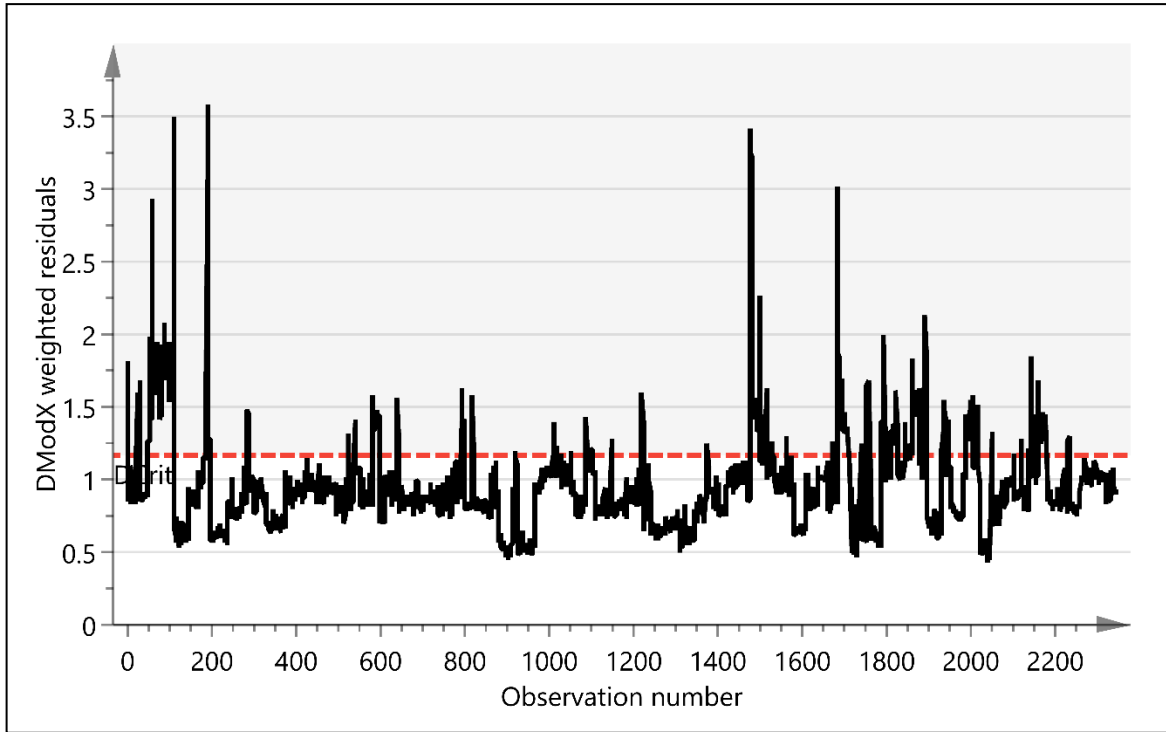


**Figure 10: DmodX plot of physical properties model**

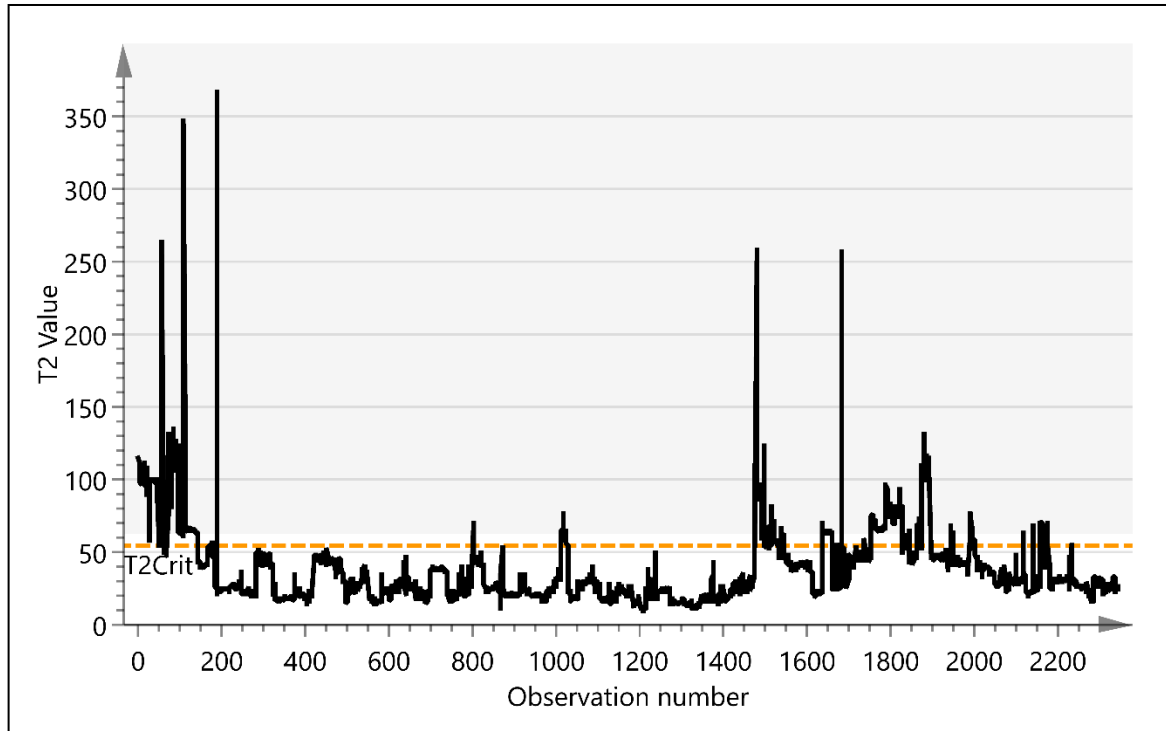


**Figure 11: Hotelling's ( $T^2$ ) plot of physical properties model**

After removing outlier observations, the model number of observations was reduced to 2284 observations. Therefore, the model variance and capability to explain the dataset have increased. The outliers are determined as the DmodX or Hotelling's ( $T^2$ ) value of an observation that is more than twice the critical value of DmodX ( $D_{crit}$ ) or the critical value of Hotelling's ( $T^2_{crit}$ ), respectively. Figures 10 and 11 show that  $D_{crit}$  and  $T^2_{crit}$  have been found to be 1.17 and 54.4, respectively. Thus, outlier observations have been detected based on these values.

#### **4.3.4 Physical Properties Model Main Findings**

The non-contributory variables and outlier observations were removed from the data analysis, ensuring the dataset was ready for model building. A summary of the results of the physical properties model is presented in Table 16. The results include key model performance metrics such as the number of

observations, the number of X variables, the number of Y variables, the number of latent variables (A), the variance of independent variables ( $R_x^2$ ), the variance of dependent variables ( $R_y^2$ ), and the cross-validation variance of dependent variables ( $Q^2$ ).

**Table 16: Summary of physical properties model results**

| <b>Model Performance Metrics</b>                  | <b>Symbol</b> | <b>Result</b> |
|---------------------------------------------------|---------------|---------------|
| Number of observations                            | N             | 2284          |
| Total number of X variables                       | -             | 30            |
| Number of discrete X variables                    | -             | 10            |
| Number of continuous X variables                  | -             | 20            |
| Number of Y variables                             | -             | 3             |
| Number of latent variables                        | A             | 18            |
| Independent variables variance (%)                | $R_x^2$       | 85.2          |
| Dependent variables variance (%)                  | $R_y^2$       | 64.0          |
| Cross-validation dependent variables variance (%) | $Q^2$         | 62.7          |

The model demonstrates strong explanatory power, as the variance of the dependent variable exceeds 50% and approaches 70%. Additionally, the model utilizes a substantial portion of the independent variables' data, with 85.2% of the X variables contributing to the prediction of the dependent variables. This is a sizable proportion, especially considering the substantial number of observations and X variables.

The number of latent variables used in the model is high, which is reasonable given that the dataset includes more than two thousand observations. Furthermore, the selection of latent variables is based on the cross-validation method, with a cross-validated variance of 62.7%.

The behavior of the dependent variables' variance ( $R_y^2$ ) as a new latent variable is added is illustrated in Figure 12. It is evident that while  $R_y^2$  increases with the addition of latent variables, the rate of increase diminishes as more latent variables are included. Figure 13 presents the cross-validated variance ( $Q^2$ ) in relation to the number of latent variables. The  $Q^2$  value initially increases with the addition of new latent variables, reaching an optimal point before it starts to decline. However, the magnitude of the decline in  $Q^2$  is insignificant compared to the earlier increase in  $R_y^2$ . Therefore, adding new latent variables is beneficial in explaining the data until the increase in  $R_y^2$  becomes too small relative to the decrease in  $Q^2$ . Once this point is reached, the optimal number of latent variables has been identified, and adding more would contribute only to explaining noise without improving the model.

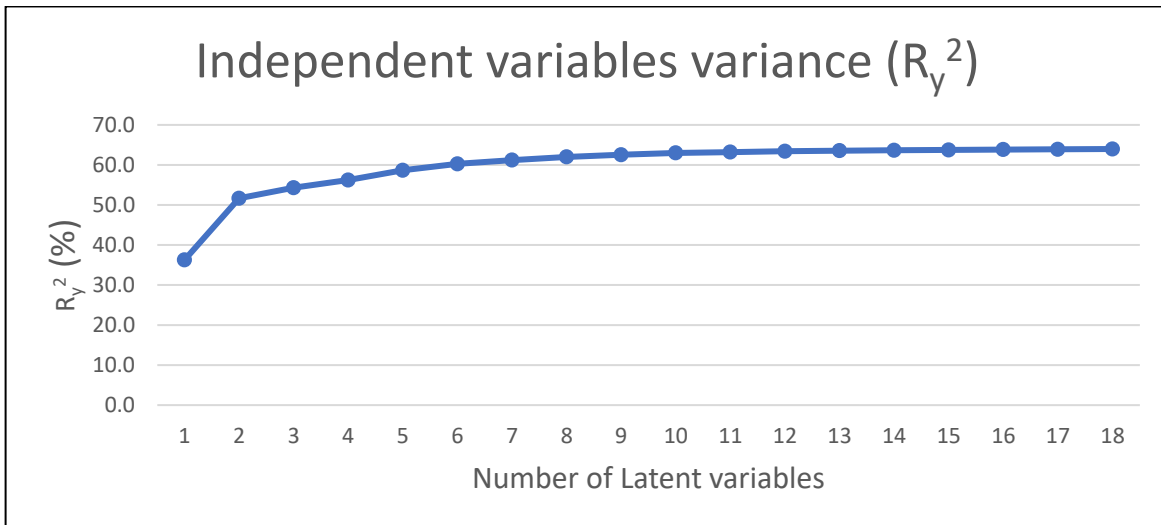


**Figure 12: Independent variables variance ( $R_y^2$ ) explained by model versus the number of latent variables used**

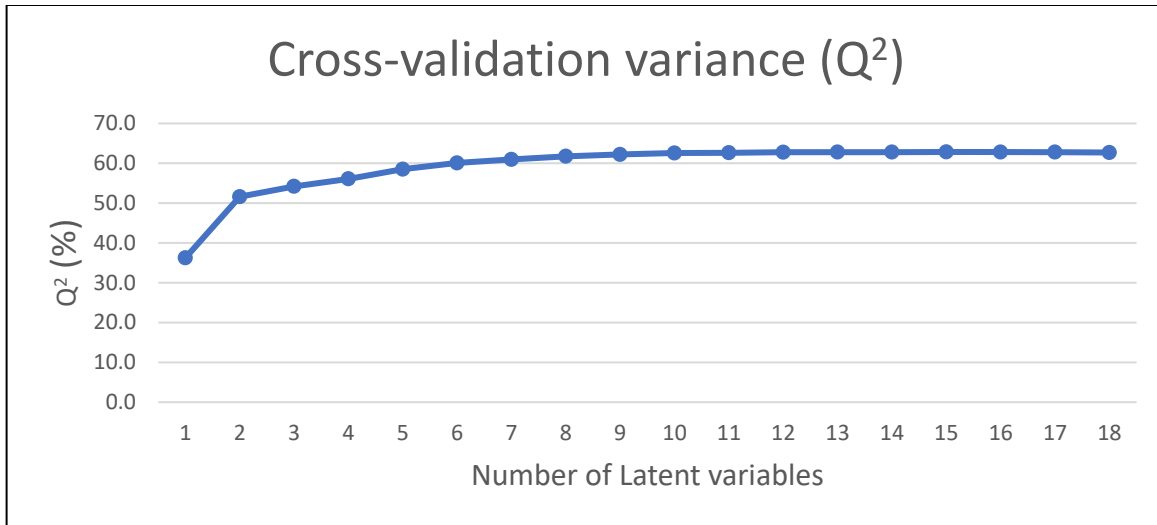


**Figure 13: Cross-validation variance ( $Q^2$ ) explained by model versus the number of latent variables used**

The cross-validation variance ( $Q^2$ ) began to decline after reaching its optimal value of 62.8% with 15 latent variables. Despite this, additional latent variables were incorporated, which reduced  $Q^2$  by only 0.1%. This 0.1% reduction in  $Q^2$  resulted in a 0.4% increase in the variance explained by the model, bringing the independent variables' variance to a maximum of 64.0%. Therefore, adding latent variables contributed to describing more of the data rather than introducing noise.

The mathematical model equation for the physical properties was derived using the PLS model. Table 17 presents the coefficients of the independent variables for each of the three Y-variables. These coefficients can be used to formulate the mathematical expression shown in Equation (4). These coefficients correspond to the scaled and mean-centered X-variables for the continuous variables. In contrast, the discrete variables are unstandardized, allowing for a direct comparison of their effects on the Y-variables. The coefficients' magnitude indicates each independent variable's

impact: a higher coefficient value implies a greater influence on the corresponding dependent variable.

Moreover, as the model built is based on mean-centered and scaled data, the data removes the effect of any constant or intercept. Therefore, the coefficients table does not include a constant or intercept. As in this modeling, the coefficients represent the relationship between the standardized variables and the response, so the need for a constant term is bypassed. Furthermore, some near-zero values of coefficients indicate a low relationship between dependent and independent variables. This is expected as these relationships are not found in the literature. So, these non-zero values are consistent with the literature.

**Table 17: Model 1 - Physical properties model mathematical equation coefficients**

| <b>Coefficient</b> | <b>Geometric Density (<math>y_1</math>)</b> | <b>Specific Electrical Resistance (<math>y_2</math>)</b> | <b>Flexural Strength (<math>y_3</math>)</b> |
|--------------------|---------------------------------------------|----------------------------------------------------------|---------------------------------------------|
| Intercept          | 0                                           | 0                                                        | 0                                           |
| $x_1$              | 0.303                                       | -0.020                                                   | 0.002                                       |
| $x_2$              | 0.132                                       | -0.067                                                   | 0.098                                       |
| $x_3$              | -0.034                                      | 0.009                                                    | -0.009                                      |
| $x_6$              | -0.003                                      | 0.082                                                    | -0.071                                      |
| $x_7$              | 0.033                                       | 0.005                                                    | 0.009                                       |
| $x_8$              | -0.040                                      | -0.058                                                   | 0.036                                       |
| $x_9$              | 0.073                                       | 0.038                                                    | -0.048                                      |
| $x_{10}$           | -0.248                                      | -0.002                                                   | 0.034                                       |
| $x_{11}$           | -0.113                                      | -0.124                                                   | 0.119                                       |
| $x_{12}$           | 0.110                                       | 0.076                                                    | -0.069                                      |
| $x_{13}$           | 0.195                                       | 0.017                                                    | 0.019                                       |
| $x_{14}$           | -0.044                                      | -0.005                                                   | 0.001                                       |

|          |        |        |        |
|----------|--------|--------|--------|
| $x_{15}$ | -0.031 | 0.064  | -0.026 |
| $x_{16}$ | -0.102 | 0.017  | -0.016 |
| $x_{17}$ | 0.069  | -0.054 | 0.038  |
| $x_{18}$ | -0.005 | 0.002  | -0.007 |
| $x_{19}$ | -0.059 | -0.003 | 0.005  |
| $x_{20}$ | 0.011  | 0.028  | -0.004 |
| $x_{21}$ | 0.019  | 0.054  | -0.038 |
| $x_{22}$ | 0.058  | 0.011  | -0.014 |
| $x_{23}$ | -0.126 | -0.013 | 0.027  |
| $x_{24}$ | 0.088  | -0.027 | 0.038  |
| $x_{25}$ | -0.031 | 0.016  | -0.004 |
| $x_{26}$ | -0.007 | -0.045 | 0.044  |
| $x_{27}$ | -0.053 | -0.001 | -0.003 |
| $x_{28}$ | -0.066 | 0.014  | 0.011  |
| $x_{29}$ | 0.121  | -0.014 | -0.019 |
| $x_{30}$ | 0.122  | -0.071 | 0.033  |
| $x_{31}$ | -0.153 | -0.051 | 0.013  |
| $x_{32}$ | -0.138 | -0.098 | 0.124  |

The results in Table 17 indicate that the obtained mathematical expression is consistent with the literature. First, GAD is directly proportional to geometric density and flexural strength and inversely proportional to specific electrical resistance, which aligns with previous findings. Second, the anode forming speed, another key parameter in the paste plant process, directly correlates with geometric density. The coefficient for anode forming speed increases with higher geometric density at greater speeds. Lastly, using more coarse particles in the anode recipe increases the geometric density. In comparison, a lower percentage of coarse particles reduces it, which is consistent with the expected process behavior.

For the baking kiln process parameters, the behavior of the variables aligns with the literature, as the geometric density is directly proportioned with fire cycle time. On the other hand, the fire cycle time inversely affects the specific electrical resistance. Additionally, the final anode temperature and heating rate are inversely proportional to geometric density, consistent with expected process behavior. Finally, as noted in the literature, an increase in final anode temperature has a negative effect on flexural strength due to crack formation caused by high temperatures.

#### **4.3.5 Test the Significance of the Independent Variables of Model 1**

Two-tailed t-tests were conducted to evaluate the statistical significance of the independent variables with near-zero coefficients. The hypothesis tested is that these variables do not affect the dependent variable ( $y_j$ ). This test aims to determine whether there is a significant relationship between a specific independent variable and a specific dependent variable. The null hypothesis ( $H_0$ ) and alternative hypothesis ( $H_1$ ) are defined as follows:

**H0:** The variable  $x_i$  has no significant effect on  $y_j$  ( $\beta_i=0$ ).

**H1:** The variable  $x_i$  has a significant effect on  $y_j$  ( $\beta_i\neq 0$ ).

If the null hypothesis is rejected,  $x_i$  is considered significant in predicting  $y_j$ .

The Sum of Squares of Errors (SSE) for the dependent variables  $y_1$ ,  $y_2$ , and  $y_3$  were found to be 1381.09, 2232.54, and 2148.77, respectively. These values were used to calculate the corresponding t-values. Tables 18, 19, and 20 present the near-zero coefficients, t-values, p-values, significance levels, and hypothesis decisions for  $y_1$ ,  $y_2$ , and  $y_3$ .

The degree of freedom (df) for all tests were calculated as 2253, consistent with the total number of observations and independent variables. For additional details on the statistical testing parameters, including the sum of squares deviations for the independent variables and their standard errors ( $SE_{\beta}$ ), refer to Appendix (6).

**Table 18: Geometric Density ( $y_1$ ) - test the significance of x-variables with near-zero coefficients**

| Variable | Coefficient ( $\beta_i$ ) | t-value | p-value | Significance   | Hypothesis decision ( $H_0$ ) |
|----------|---------------------------|---------|---------|----------------|-------------------------------|
| $x_6$    | -0.003                    | -0.183  | 0.855   | $0.855 > 0.05$ | Accepted                      |
| $x_{18}$ | -0.005                    | -0.305  | 0.760   | $0.760 > 0.05$ | Accepted                      |
| $x_{20}$ | 0.011                     | 4.143   | 0.000   | $0.000 < 0.05$ | Rejected                      |
| $x_{21}$ | 0.019                     | 1.160   | 0.246   | $0.246 > 0.05$ | Accepted                      |
| $x_{26}$ | -0.007                    | -0.427  | 0.669   | $0.669 > 0.05$ | Accepted                      |

**Table 19: Specific Electrical Resistance ( $y_2$ ) - test the significance of x-variables with near-zero coefficients**

| Variable | Coefficient ( $\beta_i$ ) | t-value | p-value | Significance   | Hypothesis decision ( $H_0$ ) |
|----------|---------------------------|---------|---------|----------------|-------------------------------|
| $x_3$    | 0.009                     | 0.432   | 0.666   | $0.666 > 0.05$ | Accepted                      |
| $x_7$    | 0.005                     | 0.240   | 0.810   | $0.810 > 0.05$ | Accepted                      |
| $x_{10}$ | -0.002                    | -0.071  | 0.944   | $0.944 > 0.05$ | Accepted                      |
| $x_{13}$ | 0.017                     | 2.811   | 0.005   | $0.005 < 0.05$ | Rejected                      |
| $x_{14}$ | -0.005                    | -0.341  | 0.733   | $0.733 > 0.05$ | Accepted                      |
| $x_{16}$ | 0.017                     | 0.816   | 0.415   | $0.415 > 0.05$ | Accepted                      |
| $x_{18}$ | 0.002                     | 0.096   | 0.924   | $0.924 > 0.05$ | Accepted                      |
| $x_{19}$ | -0.003                    | -0.144  | 0.886   | $0.886 > 0.05$ | Accepted                      |
| $x_{22}$ | 0.011                     | 0.528   | 0.598   | $0.598 > 0.05$ | Accepted                      |

|          |        |        |       |                |          |
|----------|--------|--------|-------|----------------|----------|
| $x_{23}$ | -0.013 | -0.624 | 0.533 | $0.533 > 0.05$ | Accepted |
| $x_{25}$ | 0.016  | 0.768  | 0.443 | $0.443 > 0.05$ | Accepted |
| $x_{27}$ | -0.001 | -0.048 | 0.962 | $0.962 > 0.05$ | Accepted |
| $x_{28}$ | 0.014  | 0.672  | 0.502 | $0.502 > 0.05$ | Accepted |
| $x_{29}$ | -0.014 | -0.832 | 0.406 | $0.406 > 0.05$ | Accepted |

**Table 20: Flexural Strength ( $y_3$ ) - test the significance of x-variables with near-zero coefficients**

| Variable | Coefficient ( $\beta_i$ ) | t-value | p-value | Significance   | Hypothesis decision ( $H_0$ ) |
|----------|---------------------------|---------|---------|----------------|-------------------------------|
| $x_1$    | 0.002                     | 0.098   | 0.922   | $0.922 > 0.05$ | Accepted                      |
| $x_3$    | -0.009                    | -0.440  | 0.660   | $0.660 > 0.05$ | Accepted                      |
| $x_7$    | 0.009                     | 0.440   | 0.660   | $0.660 > 0.05$ | Accepted                      |
| $x_{13}$ | 0.019                     | 3.203   | 0.001   | $0.001 < 0.05$ | Rejected                      |
| $x_{14}$ | 0.001                     | 0.069   | 0.945   | $0.945 > 0.05$ | Accepted                      |
| $x_{16}$ | -0.016                    | -0.783  | 0.434   | $0.434 > 0.05$ | Accepted                      |
| $x_{18}$ | -0.007                    | -0.342  | 0.732   | $0.732 > 0.05$ | Accepted                      |
| $x_{19}$ | 0.005                     | 0.245   | 0.807   | $0.807 > 0.05$ | Accepted                      |
| $x_{20}$ | -0.004                    | -1.208  | 0.227   | $0.227 > 0.05$ | Accepted                      |
| $x_{22}$ | -0.014                    | -0.685  | 0.493   | $0.493 > 0.05$ | Accepted                      |
| $x_{25}$ | -0.004                    | -0.196  | 0.845   | $0.845 > 0.05$ | Accepted                      |
| $x_{27}$ | -0.003                    | -0.147  | 0.883   | $0.883 > 0.05$ | Accepted                      |
| $x_{28}$ | 0.011                     | 0.538   | 0.591   | $0.591 > 0.05$ | Accepted                      |
| $x_{31}$ | 0.013                     | 0.636   | 0.525   | $0.525 > 0.05$ | Accepted                      |

Independent variables with no significant effect were excluded from the mathematical expression model. However, accepting or rejecting the null hypothesis is subject to an error percentage ( $\alpha = 5\%$ ). Note that  $X_{18}$ , the fines Si content, has no significant effect on all three dependent variables; thus, it

is excluded from all the three expressions model. This is expected as  $X_{18}$  is not significant in terms of physical properties, which is consistent with the literature. As a result, Equations (33) to (35) present the expressions for the three dependent variables as the coefficients for the independent variables are collected from Table 17. It is important to note that the intercept is not included in the expression, as it is equal to zero due to data preprocessing.

$$\begin{aligned}\hat{y}_1 = & 0.303x_1 + 0.132x_2 - 0.034x_3 + 0.033x_7 - 0.040x_8 + 0.073x_9 - 0.248x_{10} \\ & - 0.113x_{11} + 0.110x_{12} + 0.195x_{13} - 0.044x_{14} - 0.031x_{15} \\ & - 0.102x_{16} + 0.069x_{17} - 0.005x_{18} - 0.059x_{19} + 0.011x_{20} \\ & + 0.058x_{22} - 0.126x_{23} + 0.088x_{24} - 0.031x_{25} - 0.053x_{27} \\ & - 0.066x_{28} + 0.121x_{29} + 0.122x_{30} - 0.153x_{31} - 0.138x_{32}\end{aligned}\quad (33)$$

$$\begin{aligned}\hat{y}_2 = & -0.020x_1 - 0.067x_2 + 0.082x_6 - 0.058x_8 + 0.038x_9 - 0.124x_{11} + 0.076x_{12} \\ & + 0.017x_{13} + 0.064x_{15} - 0.054x_{17} + 0.028x_{20} + 0.054x_{21} \\ & - 0.027x_{24} - 0.045x_{26} - 0.071x_{30} - 0.051x_{31} - 0.098x_{32}\end{aligned}\quad (34)$$

$$\begin{aligned}\hat{y}_3 = & 0.098x_2 - 0.071x_6 + 0.036x_8 - 0.048x_9 + 0.034x_{10} + 0.119x_{11} - 0.069x_{12} \\ & + 0.019x_{13} - 0.026x_{15} + 0.038x_{17} - 0.038x_{21} + 0.027x_{23} \\ & + 0.038x_{24} + 0.044x_{26} - 0.019x_{29} + 0.033x_{30} + 0.124x_{32}\end{aligned}\quad (35)$$

Some of the insignificant independent variables are of the 10 discrete variables, which means that the dependent variables are not affected by these independent variables for the ordinal levels covered by the model built as listed in Section (4.3.1). However, these variables may exhibit significance at levels beyond the range of the datasets used in this research study. Therefore, datasets encompassing a broader range of levels should be analyzed further to investigate the potential significance of these independent variables. In

contrast, the continuous variables are not subject to this limitation, as the model does not impose range constraints on them as it does for the discrete variables.

#### **4.3.6 Physical Properties Model Interpretation**

After building the model, various interpretation methods were employed to explain the model and discuss the results obtained. The first tool used for interpreting the model is the score plot. The 2D score plot of the model is presented in Figure 14. This plot represents the first two latent variables, which account for 28% of the X-variables' variance explained by the model. This constitutes nearly a quarter of the X-data variance explained by the 18 latent variables used in the model.

The plot reveals five distinct data groups, where different observations cluster together to form each group. The physical properties model data is naturally organized into these five groups. Typically, data within each group exhibit similar patterns, suggesting that each group might share common characteristics such as raw material properties, paste plant process conditions, or baking kiln process parameters.

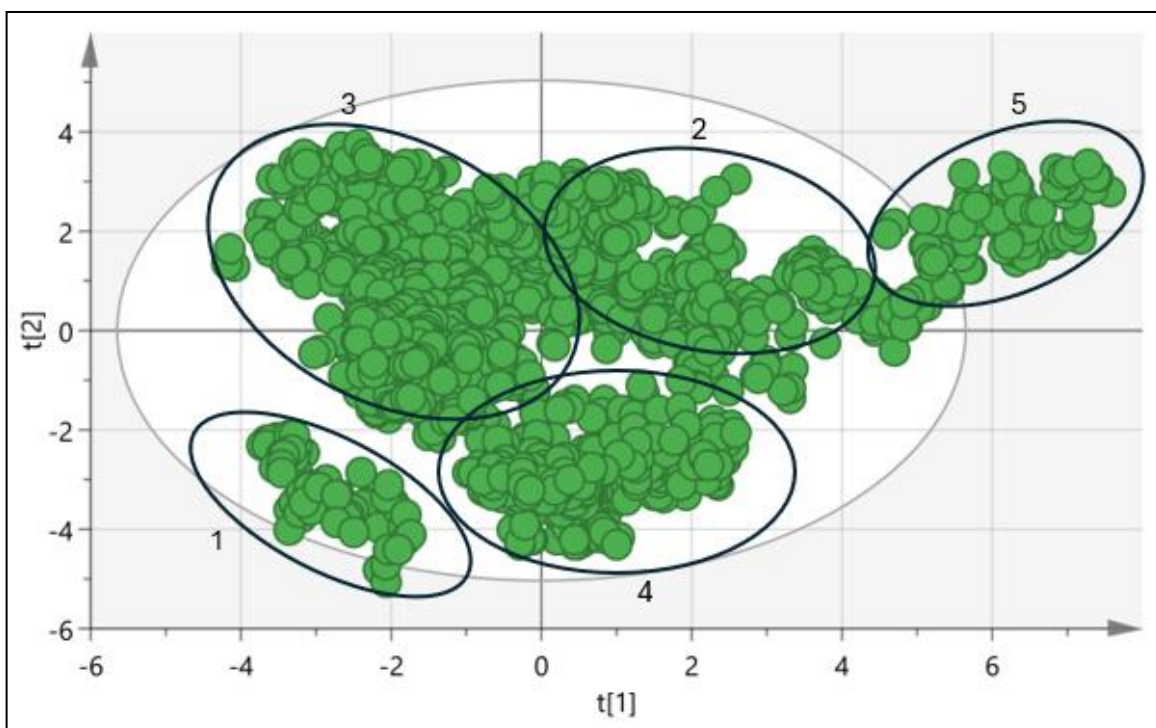


**Figure 14: The 2D score plot for the first two latent variables of the physical properties model**

For the raw material data, CPC properties vary across different groups due to using different blends to produce green anodes. Each blend represents a unique combination of CPC percentages from various supplier sources, as detailed in Table 21. For instance, blend 30088 consists of 55% CPC from Source 1, 35% from Source 2, and 10% from Source 3. These blends are adjusted based on the availability of supplier materials and variations in CPC quality. Similarly, the properties of liquid tar pitch depend on the supplier sources used. However, fewer sources are utilized for pitch, and source changes occur less frequently than CPC properties. Table 21 indicates that only two sources were used during the data collection period.

**Table 21: Raw material blend and supplier sources of physical properties model**

| Group | CPC blend | CPC supplier source (%) |    |    |    | Pitch supplier |   |
|-------|-----------|-------------------------|----|----|----|----------------|---|
|       |           | 1                       | 2  | 3  | 4  | 1              | 2 |
| 1     | 30087     | 50                      | 50 | -  | -  | ✓              |   |
| 2     | 30088     | 55                      | 35 | 10 | -  | ✓              |   |
|       | 20040     | 70                      | -  | 30 | -  |                |   |
| 3     | 40151     | 65                      | 25 | -  | 10 | ✓              | ✓ |
|       | 30094     | 50                      | 40 | -  | 10 |                |   |
|       | 40160     | 50                      | 33 | 10 | 7  |                |   |
| 4     | 30098     | 50                      | 45 | -  | 5  |                | ✓ |
|       | 30099     | 35                      | 60 | -  | 5  |                |   |
| 5     | 40161     | 50                      | 30 | 15 | 5  | ✓              |   |

In combination with the score plot, the contribution plot method is used to investigate the variables responsible for changes from one group to another. Figures 15 and 16 show each independent variable contribution of change from group 1 to group 2 and from group 4 to group 5. These two changes are investigated because both include the main supplier source change of raw material properties. As a change from group 1 to group 2, include introducing supplier source 3 with 10% and reducing supplier source 2 by 15%. Moreover, a change from group 4 to group 5 results in adding 15% of CPC from supplier source 3 and a change of pitch supplier source from source 2 into source 1, as shown in Table 21.

First, Figure 15 shows that green anode density ( $X_1$ ) is the main negative contributor to changing latent variable space from group 1 to group 2. This is investigated as changing the raw material supplier sources to produce green anode strongly affects green anode density. Furthermore, butts

percentage is the other main contributor as it is the part of the anode that remains and is reused in producing green anode. Therefore, it has been influenced by green anode parameters. Secondly, Figure 16 shows that CPC sulfur content significantly affects the change of LV space from group 4 to group 5, as the addition of 15% supplier source 3 of CPC raw material increases sulfur percentage. Moreover, the change in pitch supplier source has contributed to the change in LV space.

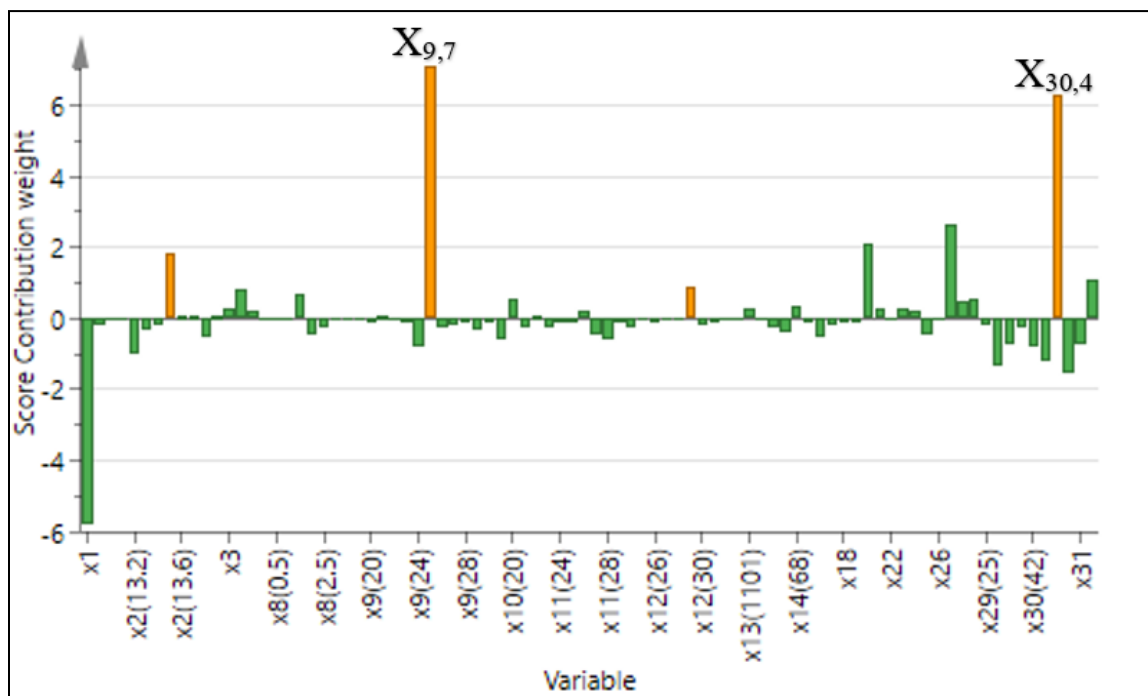


**Figure 15: The score contribution plot for the change from Group 1 to Group 2**

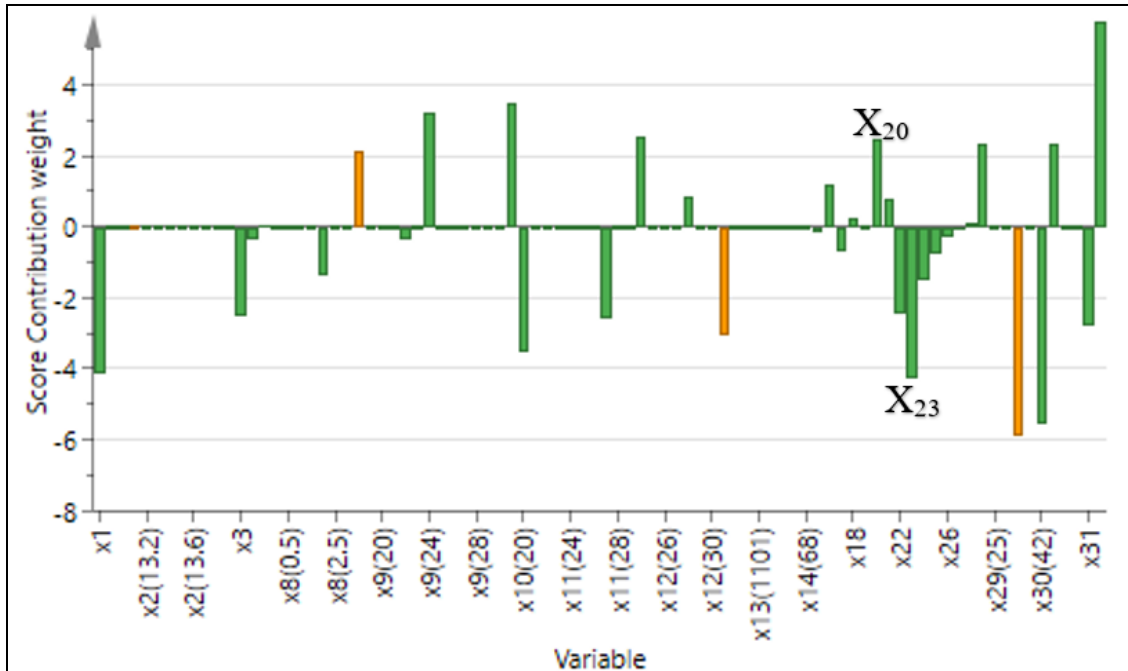


**Figure 16: The score contribution plot for the change from Group 4 to Group 5**

Figure 17 shows the 3D score plot with the third latent variable added. Group 1 and Group 5 data are still outside the space of explained data as they have a combination of X-data that the model does not capture with the first three latent variables. This indicates the variability of these data and the need for more latent variables to capture and explain more of the X-variables database by the model.

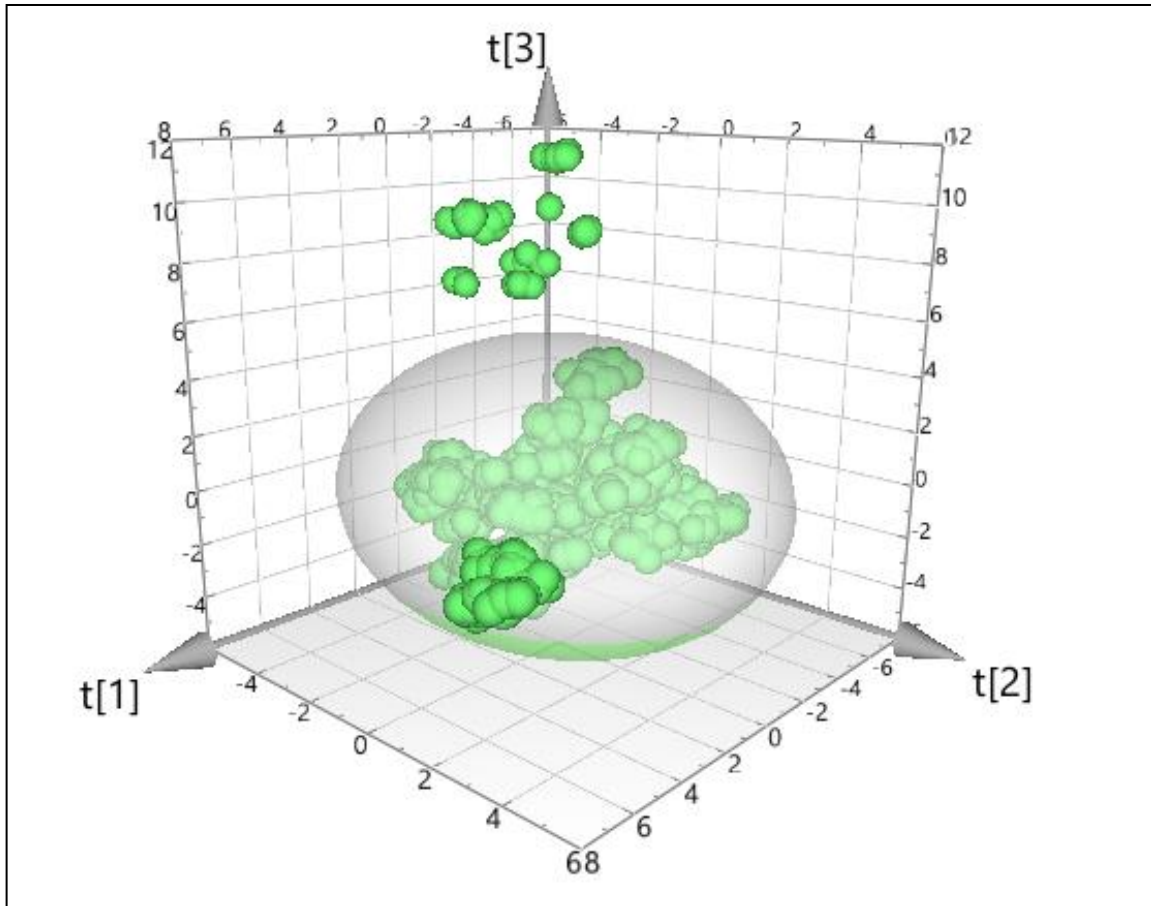


**Figure 17: The 3D score plot for the first two latent variables of the physical properties model**

The loading plot is another method to investigate the relationships between different variables within the PLS model. Figure 18 presents the loading plot for the first two latent variables, displaying the loading values for each variable (X or Y). While the high number of variables makes it difficult

to visualize all correlations, the key relationships are represented and can be effectively analyzed.

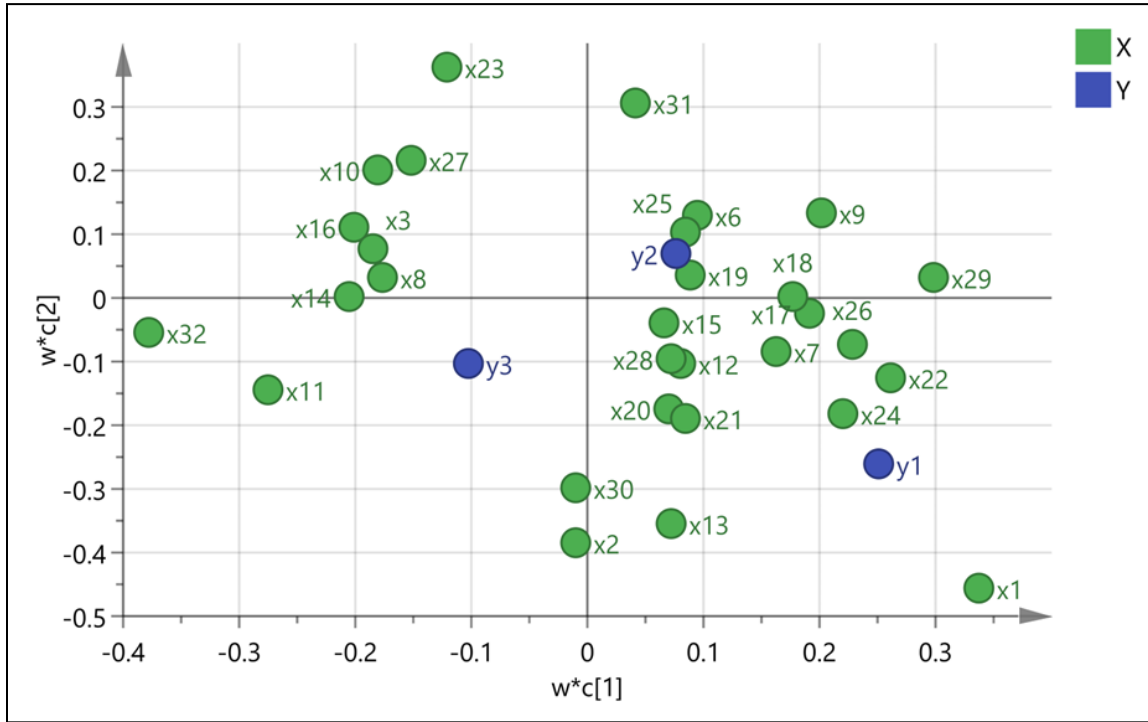


**Figure 18: Loading plot of the first two latent variables of the physical properties model**

The correlation structure between X-variables and Y-variables is illustrated in this plot. It can be observed that green anode density, heating rate, fire cycle time, and medium particles percentage are the most influential variables on the model for the first latent variable, as indicated by their high absolute loading values. Pitch level, CPC sulfur content, anode forming speed, and final anode temperature strongly influence the model based on the second latent variable's high absolute loading values. Therefore, CPC raw material properties and baking parameters significantly impact the PLS model.

The second aspect investigated using the loading plot is the positively correlated variables, which appear close to each other with the same sign of

loading value. First, geometric density is highly correlated with green anode density. Specific electrical resistance is positively correlated with final anode temperature. Additionally, flexural strength and anode forming time are positively correlated, indicating that higher flexural strength is associated with longer anode forming times. These findings are consistent with the literature.

The loading plot for latent variables 1 and 3 is presented in Figure 19, with the vertical axis representing the loading values for the third LV. From the figure, it can be observed that medium particles percentage, fine iron element content, GAD, and butts percentage have the highest absolute loading values for the third LV. This suggests that paste plant process parameters primarily influence the third latent variable. Additionally, specific electrical resistance is positively correlated with fire soaking time, while geometric density is positively affected by butts percentage. Furthermore, the third LV loading value aligns with the first LV loading value and the literature, indicating that geometric density negatively correlates with the heating rate.

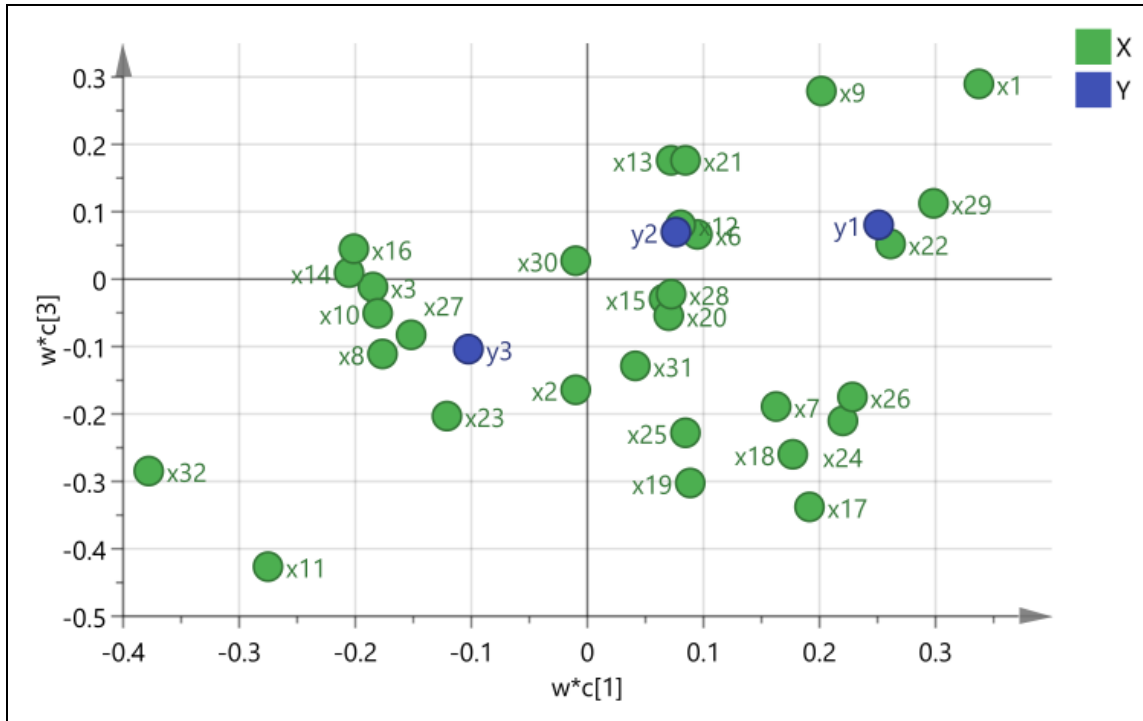


**Figure 19: Loading plot of the physical properties model's latent variables 1 and 3**

The final method used to interpret the PLS model is the Variable Importance in Projection (VIP) method. Table 22 presents the VIP values for the 10 most important X-variables. The table includes the highly important X-variables in the PLS model, their ranking for predicting Y-variables, and the process they represent (raw material property, paste plant parameter, or baking kiln parameter). The full list of VIP values for all X-variables is provided in Appendix (7A).

The most important variable is green anode density, as it represents the quality of the paste plant process. The top 10 VIP variables are distributed across raw material properties, paste plant process parameters, and baking kiln process parameters, with 2, 5, and 3 variables, respectively. This is evidence of the importance of raw material properties and process parameters in predicting

anode quality and the need to consider them to ensure a good model. As expected, raw material properties significantly impact process variation and quality. The highest-ranking VIP variables –green anode density, heating rate, and medium particles percentage– are at a similar level of importance compared to the other VIP variables. Additionally, the VIP method results align with the loading plot method, as both identify green anode density and heating rate as the most important and influential variables in the PLS model.

**Table 22: VIP values for X-variables of the physical properties model**

| Rank | VIP Value | X variable ID | X variable                       | Process      |
|------|-----------|---------------|----------------------------------|--------------|
| 1    | 1.699     | $x_1$         | Green anode density              | Paste plant  |
| 2    | 1.612     | $x_{32}$      | Heating rate                     | Baking kiln  |
| 3    | 1.421     | $x_{11}$      | Medium particles percentage      | Paste plant  |
| 4    | 1.289     | $x_{29}$      | Fire cycle time                  | Baking kiln  |
| 5    | 1.234     | $x_2$         | Pitch level                      | Paste plant  |
| 6    | 1.139     | $x_{23}$      | CPC sulfur (S) element content   | Raw material |
| 7    | 1.123     | $x_{31}$      | Final anode temperature          | Baking kiln  |
| 8    | 1.098     | $x_9$         | Butts percentage                 | Raw material |
| 9    | 1.088     | $x_{13}$      | Anode forming speed              | Paste plant  |
| 10   | 1.043     | $x_{22}$      | Pitch sulfur (S) element content | Raw material |

#### 4.3.7 Prediction of New Physical Properties Dataset

Once the model has been developed, a separate prediction dataset is used to estimate the dependent variables. This step is essential for validating the PLS model's predictive accuracy and assessing its generalization capability to new, unseen data. The number of observations to predict the dependent variables is 790, 25% of the total dataset. Table 23 presents the

variance of predicted versus observed dependent variables for the three Y variables and the average prediction variance. The geometric density has the highest variance of the three variables.

**Table 23: Observed versus predicted dependent variables variance values of the physical properties model**

| <b>Y-Variable ID</b> | <b>Variable</b>                | <b>Variance (<math>R^2</math>)</b> |
|----------------------|--------------------------------|------------------------------------|
| $y_1$                | Geometric Density              | 56.27%                             |
| $y_2$                | Specific Electrical Resistance | 52.24%                             |
| $y_3$                | Flexural Strength              | 44.89%                             |

The average prediction variance is 51.13%, less than the model built (training variance) by 12.9%. This suggests that the prediction dataset is more complex and variable than the training dataset. Moreover, the prediction dataset is from the period after the training dataset. This indicates that the raw material blends and properties used in the training and prediction datasets differ. However, the prediction variance remains satisfactory, as it exceeds 50%, making it reliable for predicting the quality of anodes. Figures 20, 21, and 22 present the predicted versus observed results of geometric density, Specific electrical resistance, and flexural strength, respectively. The black line represents the predicted values, and the blue dotted line represents the observed values.

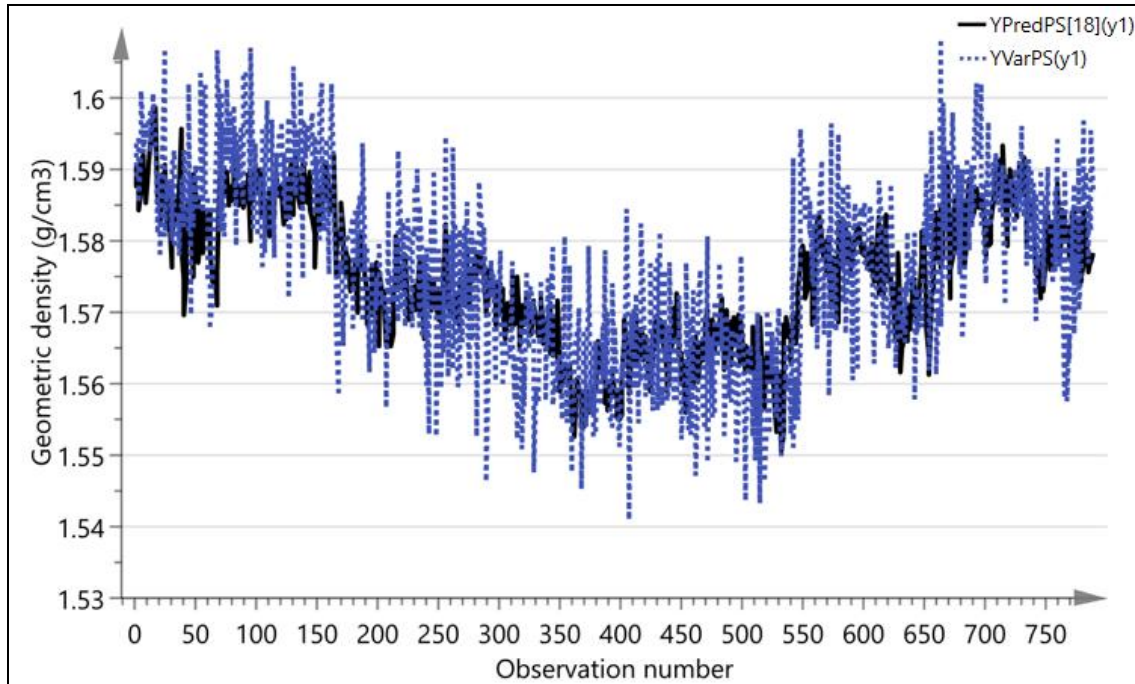


**Figure 20: Comparison between observed versus predicted Geometric Density ( $y_1$ )**

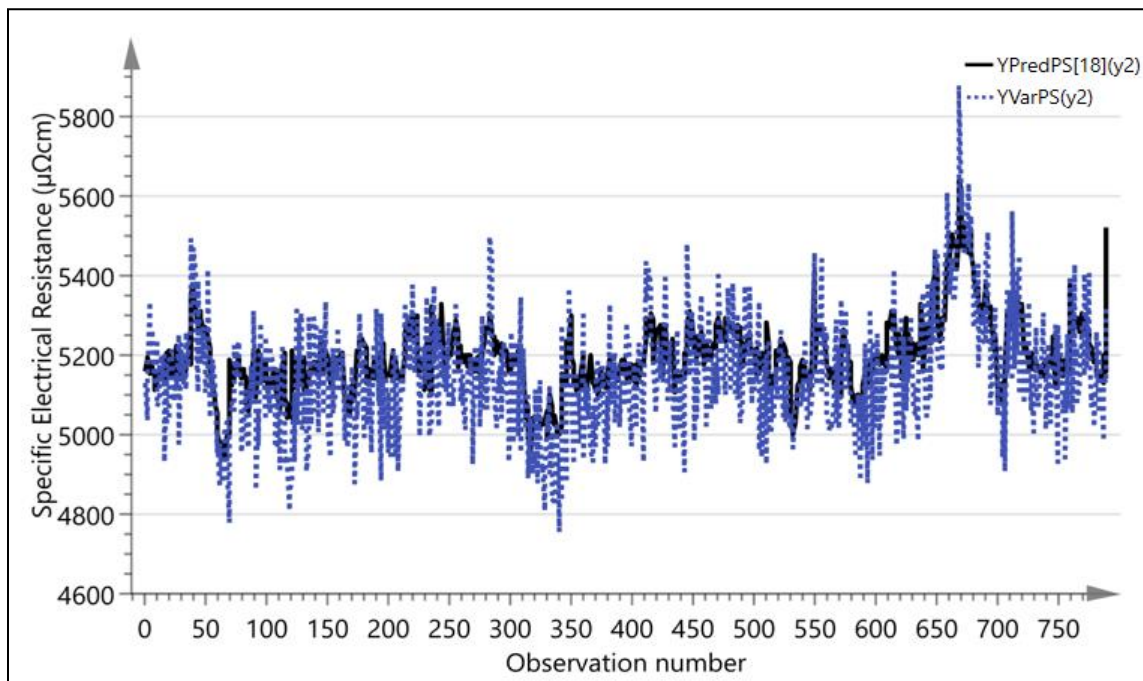


**Figure 21: Comparison between observed versus predicted Specific Electrical Resistance ( $y_2$ )**

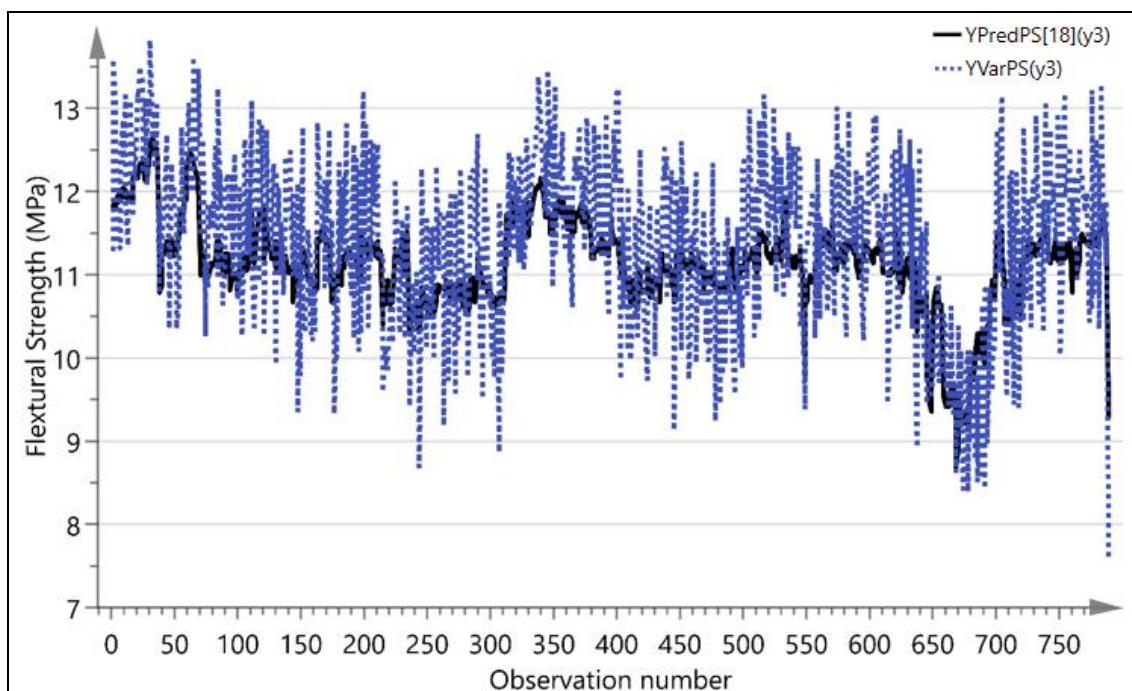


**Figure 22: Comparison between observed versus predicted Flexural Strength ( $y_3$ )**

The three figures show that prediction ability is good as the observed and predicted values follow the same pattern of increasing and decreasing throughout the observations used for prediction. Moreover, the specific electrical resistance and flexural strength show some observations of periods of sudden increase and decrease, which the model succeeded in predicting. However, the spread of observed data is higher than the observed, which suggests that PLS could give better predictivity when data is less variable.

#### **4.4 Chemical and Reactivity Properties Model**

The second model developed to improve anode quality is the chemical and reactivity properties model, which includes chemical and reactivity variables. As previously mentioned, the collected dataset cannot be used to build a single comprehensive model of anode quality, as the anode cores tested

(observations) with physical properties data do not include chemical and reactivity properties data. The final dataset used for the chemical and reactivity model is provided in Appendix (7B). This dataset was tested and filtered to remove invalid observations. Additionally, independent variables that do not contribute to or influence the model were excluded from the analysis.

#### **4.4.1 Discrete Independent Variables of Chemical and Reactivity Properties Model**

Like the physical properties model, discrete variables in the chemical and reactivity model are ordered numerically based on their values. However, the variable values in the chemical and reactivity model dataset differ from those in the physical properties model, as the observations for the two models are distinct. The discrete variables, their values, and represented ordinal levels are presented in Table 24, covering only three discrete variables: fine particles percentage, coarse particles percentage, and anode forming time. These are the only discrete variables whose values differ from physical properties model values. Therefore, they need to represent ordinal levels in different ways.

**Table 24: Fine, Coarse Particles Percentage, and Anode Forming Time variables values and represented levels of the chemical and reactivity properties model**

| Variable                                 | Variable value | Represented level |
|------------------------------------------|----------------|-------------------|
| Fine particles percentage ( $x_{10}$ )   | 18             | 1                 |
|                                          | 19             | 2                 |
|                                          | 20             | 3                 |
|                                          | 21             | 4                 |
| Coarse particles percentage ( $x_{12}$ ) | 24             | 1                 |
|                                          | 25             | 2                 |
|                                          | 26             | 3                 |
|                                          | 27             | 4                 |
|                                          | 28             | 5                 |
|                                          | 29             | 6                 |
|                                          | 30             | 7                 |
|                                          | 31             | 8                 |
|                                          | 33             | 10                |
|                                          | 35             | 12                |
| Anode forming time ( $x_{14}$ )          | 65             | 1                 |
|                                          | 67             | 3                 |
|                                          | 68             | 4                 |

#### 4.4.2 Chemical and Reactivity Model Preparation

The independent variables were evaluated to assess the strength of their relationships with the dependent variables. Cooling power ( $x_4$ ) demonstrated a weak correlation with the dependent variables. Table 25 presents the variance values between cooling power and the dependent variables, which indicates that cooling power has been excluded from the model based on its weak contribution. As cooling power is excluded from both the physical and

chemical models, the carbon plant controllers should consider removing this variable from monitoring and controlling the process; even in literature, no such plant or researcher has highlighted the importance of this variable in controlling and improving anode quality. The significance of the remaining 31 variables on the dependent variables will be tested after building the model.

**Table 25: Cooling power relationship strength with dependent variables in the chemical and reactivity model**

| <b>X-variable</b> | <b>Y-variable</b>                  | <b>Variance (<math>R^2</math>)</b> |
|-------------------|------------------------------------|------------------------------------|
| Cooling power     | CO <sub>2</sub> Reactivity Residue | 2.84%                              |
|                   | Sulfur (S) Element Content         | 0.75%                              |
|                   | Iron (Fe) Element Content          | 3.91%                              |

The final dataset for the chemical and reactivity properties model has fewer observations than that of the physical properties model. This is due to chemical and reactivity variables being analyzed less frequently than physical variables during the data collection. The total number of observations in the final dataset was 587, which was reduced to 541 after invalid observations were removed. The model was then built using 75% of this reduced dataset, which is 405 observations.

Outlier observations were detected using the same methods as previously: DmodX and Hotelling's ( $T^2$ ). Figures 23 and 24 show the DmodX and Hotelling's plots, respectively. The same procedure was followed to identify outliers, using the critical values ( $D_{crit}$  and  $T^2_{crit}$ ), which were determined to be 1.19 and 30.67, respectively. After identifying and removing the outlier observations based on these methods, the number of observations in the model was reduced to 388.

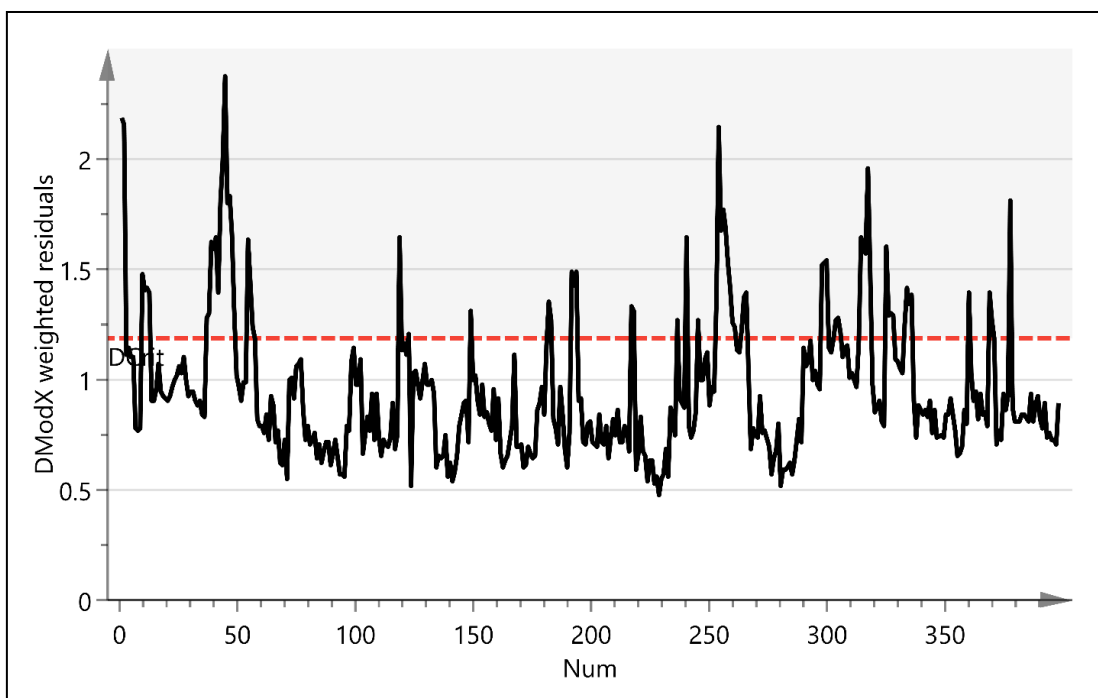


**Figure 23: DmodX plot of the chemical and reactivity model**

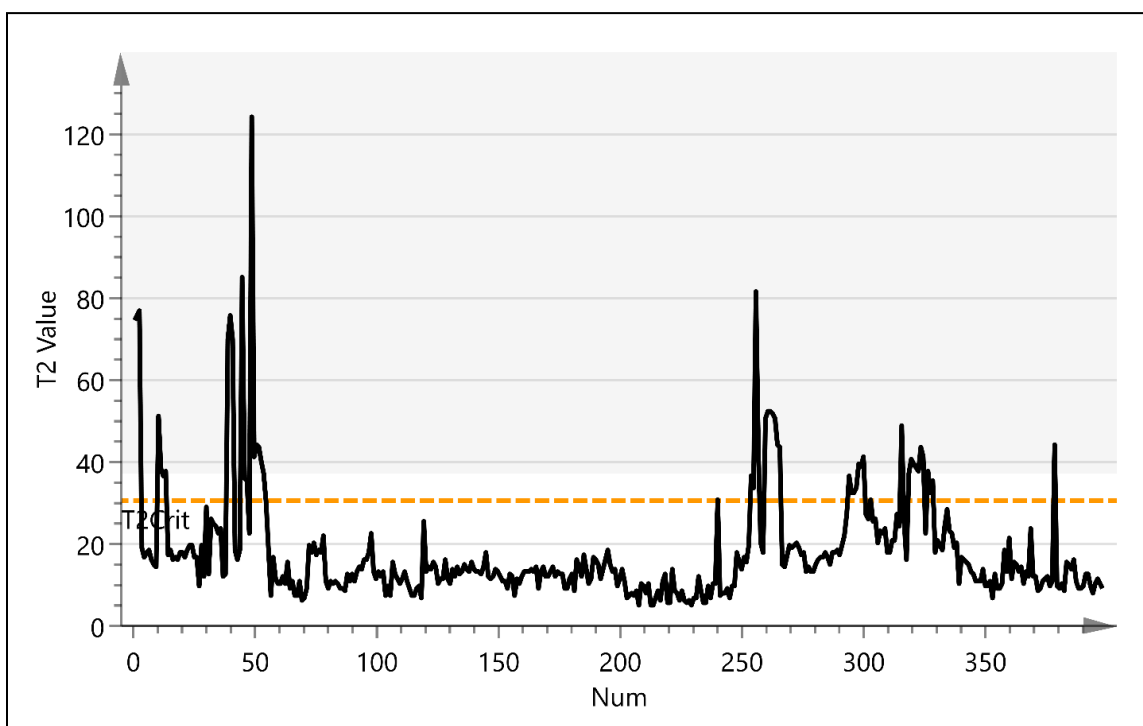


**Figure 24: Hotelling's ( $T^2$ ) plot of the chemical and reactivity model**

### 4.4.3 Chemical and Reactivity Model Main Findings

After preparing the dataset, the chemical and reactivity model was ready to be constructed. Like the physical properties model, Table 26 summarizes the chemical and reactivity model's results. The results highlight clear differences between the two models, such as the number of observations and latent variables used. In this case, only nine latent variables were required. Since the number of observations in this model is much smaller, fewer latent variables were needed to capture the dataset without capturing noise.

**Table 26: Summary of the chemical and reactivity properties model**

| Model Performance Metrics                         | Symbol  | Result |
|---------------------------------------------------|---------|--------|
| Number of observations                            | N       | 388    |
| Total number of X variables                       | -       | 31     |
| Number of discrete X variables                    | -       | 10     |
| Number of continuous X variables                  | -       | 21     |
| Number of Y variables                             | -       | 3      |
| Number of latent variables                        | A       | 9      |
| Independent variables variance (%)                | $R_x^2$ | 80.9   |
| Dependent variables variance (%)                  | $R_y^2$ | 64.1   |
| Cross-validation dependent variables variance (%) | $Q^2$   | 60.0   |

The chemical and reactivity properties model demonstrated a strong relationship, with the dependent variables accounting for 64.1% of the variance. This is almost the same as the variance explained by the physical properties model. The variance of independent variables is lower than that of the physical properties model. This can be attributed to the high variability in the X-variables dataset, as many X-variables primarily influence physical properties.

The relationship between the number of LVs used in the chemical and reactivity properties model and the independent variables' variance ( $R_y^2$ ) is illustrated in Figure 25. It shows that adding the last three LVs increases the variance by less than 3%, indicating that the optimal number of LVs has been nearly reached. Additionally, Figure 26 presents the cross-validated variance ( $Q^2$ ) behavior as more LVs are added. The optimal  $Q^2$  value is reached after 7 LVs; however, two additional LVs were included, as the slight decrease in  $Q^2$  is justifiable given the corresponding increase in  $R_y^2$ . Adding more than 9 LVs would result in a negligible increase in  $R_y^2$  and a significant decrease in  $Q^2$ , suggesting that the model would begin to capture noise rather than meaningful data.

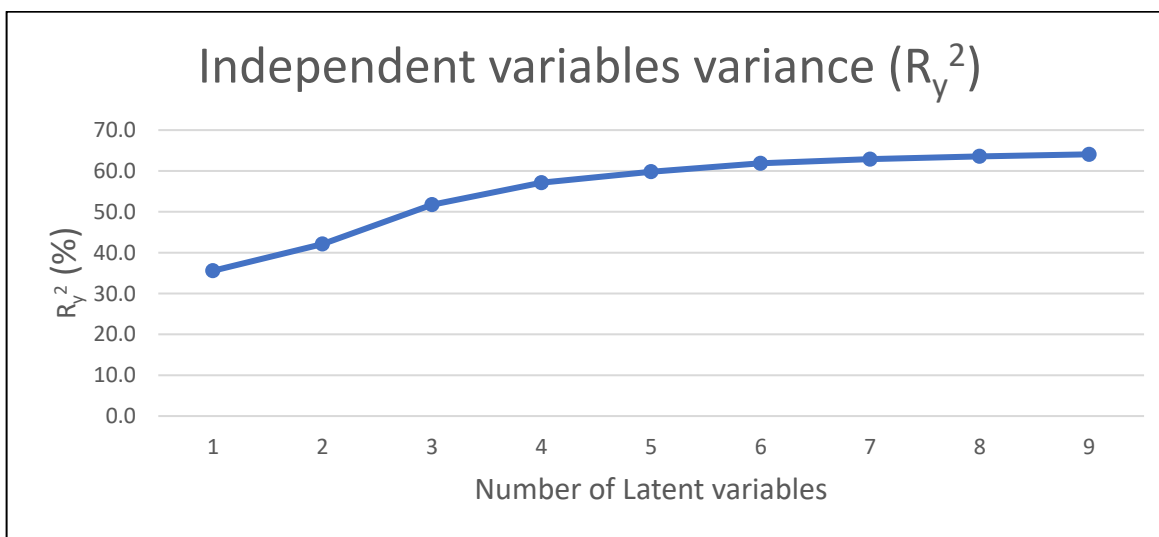


**Figure 25: Independent variables variance ( $R_y^2$ ) explained by model versus number of latent variables used of the chemical and reactivity properties model**

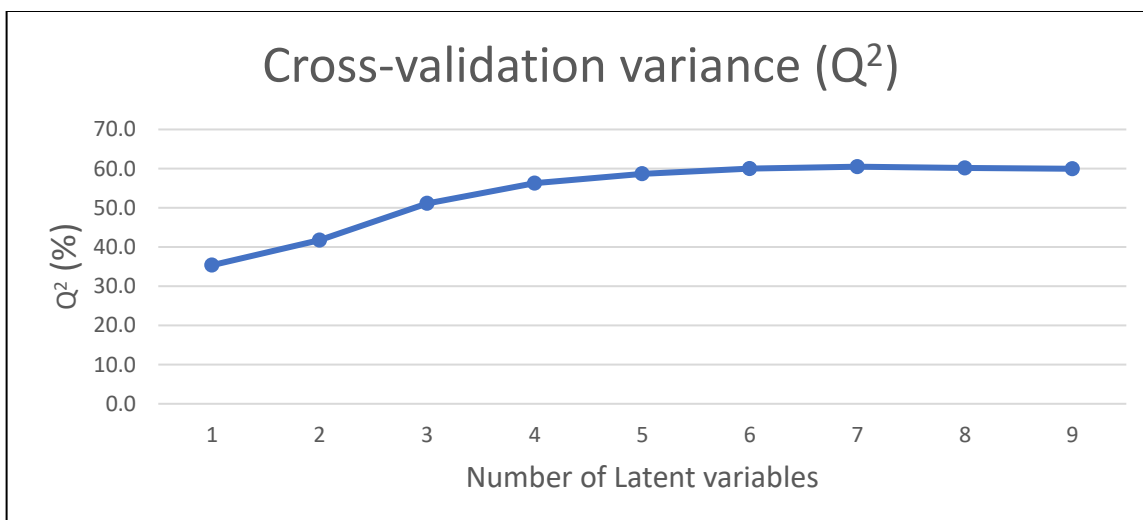


**Figure 26: Cross-validation variance ( $Q^2$ ) explained by model versus number of latent variables used of the chemical and reactivity properties model**

The PLS model for the chemical and reactivity properties has been used to derive a mathematical model equation. Table 27 presents the coefficients of the independent variables for each of the three Y variables. Similar to the physical properties model, these coefficients form the mathematical expression in Equation (4). However, it is important to note that these coefficients are based on scaled and mean-centered variables for the continuous variables and unstandardized variables for the discrete variables. As with the physical properties model, the intercepts are not included as data are standardized. Moreover, non-zero values are consistent with the literature.

**Table 27: Model 2 - Chemical and reactivity properties model mathematical equation coefficients**

| Coefficient | CO <sub>2</sub> Reactivity Residue ( $y_4$ ) | Sulfur (S) Element Content ( $y_5$ ) | Iron (Fe) Element Content ( $y_6$ ) |
|-------------|----------------------------------------------|--------------------------------------|-------------------------------------|
| Intercept   | 0                                            | 0                                    | 0                                   |

|          |        |        |        |
|----------|--------|--------|--------|
| $x_1$    | 0.049  | -0.018 | -0.034 |
| $x_2$    | 0.037  | 0.084  | 0.007  |
| $x_3$    | 0.030  | 0.074  | -0.029 |
| $x_5$    | 0.051  | -0.025 | -0.041 |
| $x_6$    | -0.084 | -0.053 | -0.021 |
| $x_7$    | 0.109  | 0.068  | 0.040  |
| $x_8$    | 0.007  | -0.090 | 0.013  |
| $x_9$    | -0.104 | -0.061 | 0.055  |
| $x_{10}$ | 0.133  | -0.060 | -0.186 |
| $x_{11}$ | 0.102  | 0.036  | 0.024  |
| $x_{12}$ | -0.017 | 0.062  | -0.022 |
| $x_{13}$ | 0.082  | -0.024 | -0.088 |
| $x_{14}$ | -0.046 | 0.031  | 0.025  |
| $x_{15}$ | 0.017  | -0.030 | 0.049  |
| $x_{16}$ | -0.003 | 0.012  | 0.013  |
| $x_{17}$ | 0.004  | 0.053  | 0.024  |
| $x_{18}$ | 0.000  | 0.055  | 0.004  |
| $x_{19}$ | 0.001  | -0.014 | -0.057 |
| $x_{20}$ | -0.039 | -0.050 | 0.028  |
| $x_{21}$ | 0.028  | -0.217 | -0.025 |
| $x_{22}$ | -0.024 | 0.035  | -0.008 |
| $x_{23}$ | -0.043 | 0.272  | 0.115  |
| $x_{24}$ | -0.004 | 0.039  | 0.015  |
| $x_{25}$ | -0.111 | 0.057  | 0.104  |
| $x_{26}$ | -0.005 | 0.076  | 0.054  |
| $x_{27}$ | -0.023 | -0.035 | -0.033 |
| $x_{28}$ | -0.026 | 0.041  | 0.023  |
| $x_{29}$ | 0.106  | 0.033  | 0.001  |
| $x_{30}$ | 0.203  | -0.232 | -0.117 |
| $x_{31}$ | -0.012 | 0.054  | 0.053  |
| $x_{32}$ | 0.017  | -0.055 | -0.097 |

As expected, GAD which was the most important parameter in the physical properties model, becomes less significant in this model since it is a physical parameter rather than a chemical one. For the first dependent variable, CO<sub>2</sub> reactivity residue, the primary observation is that CPC sulfur content is the most negatively correlated variable. This is because sulfur acts as an inhibitor to the CO<sub>2</sub> reaction, leading to higher residue in anodes with higher sulfur content. Additionally, pitch sulfur content has an inverse effect on CO<sub>2</sub> reactivity residue, but its influence is smaller than CPC sulfur content. This is due to the lower sulfur content in pitch and its smaller proportion in the anode compared to CPC. Furthermore, among the baking kiln parameters, longer soaking times result in higher CO<sub>2</sub> reactivity residue, as the anodes undergo reactive treatment for a longer period, soaking at peak temperature for an extended time.

For the second dependent variable, sulfur content in the anode, CPC sulfur content is the most influential factor, which is expected. Additionally, pitch sulfur content is directly proportional to the sulfur content in the anode. Moreover, increasing the soaking time inversely affects the anode's sulfur content, as more sulfur is burned off with longer soaking times. The third variable in the model, iron content in the anode, is directly proportional to the iron content in all CPC, fine, and butt, as anticipated.

#### **4.4.4 Test the Significance of the Independent Variables of Model 2**

As with the first model, independent variables with near-zero coefficients were tested to evaluate their statistical significance. Similarly, the H<sub>0</sub> states that the independent variable ( $x_i$ ) has no significant effect on the dependent variable ( $y_j$ ) ( $\beta_i=0$ ). In contrast, H<sub>1</sub> states that the independent

variable ( $x_i$ ) has a significant effect on the dependent variable ( $y_j$ ) ( $\beta_i \neq 0$ ). The SSE values for the dependent variables  $y_4$ ,  $y_5$ , and  $y_6$  were 339.11, 259.50, and 340.32, respectively. Tables 28, 29, and 30 summarize the key statistical parameters used to analyze the significance of independent variables with near-zero coefficients. For detailed statistical parameters, refer to Appendix (6). The degree of freedom for the statistical tests performed for this model was calculated as 356.

**Table 28: CO<sub>2</sub> Reactivity Residue ( $y_4$ ) - test the significance of  $x$ -variables with near-zero coefficients**

| <b>Variable</b> | <b>Coefficient (<math>\beta_i</math>)</b> | <b>t-value</b> | <b>p-value</b> | <b>Significance</b> | <b>Hypothesis decision (H0)</b> |
|-----------------|-------------------------------------------|----------------|----------------|---------------------|---------------------------------|
| $x_8$           | 0.007                                     | 0.185          | 0.853          | $0.853 > 0.05$      | Accepted                        |
| $x_{12}$        | -0.017                                    | -0.673         | 0.501          | $0.501 > 0.05$      | Accepted                        |
| $x_{15}$        | 0.017                                     | 0.318          | 0.751          | $0.751 > 0.05$      | Accepted                        |
| $x_{16}$        | -0.003                                    | -0.056         | 0.955          | $0.955 > 0.05$      | Accepted                        |
| $x_{17}$        | 0.004                                     | 0.075          | 0.941          | $0.941 > 0.05$      | Accepted                        |
| $x_{18}$        | 0.000                                     | 0.000          | 1.000          | $1.000 > 0.05$      | Accepted                        |
| $x_{19}$        | 0.001                                     | 0.019          | 0.985          | $0.985 > 0.05$      | Accepted                        |
| $x_{24}$        | -0.004                                    | -0.075         | 0.940          | $0.940 > 0.05$      | Accepted                        |
| $x_{26}$        | -0.005                                    | -0.093         | 0.926          | $0.926 > 0.05$      | Accepted                        |
| $x_{32}$        | 0.017                                     | 0.318          | 0.751          | $0.751 > 0.05$      | Accepted                        |

**Table 29: Sulfur (S) Element Content ( $y_5$ ) - test the significance of x-variables with near-zero coefficients**

| Variable | Coefficient ( $\beta_i$ ) | t-value | p-value | Significance   | Hypothesis decision ( $H_0$ ) |
|----------|---------------------------|---------|---------|----------------|-------------------------------|
| $x_1$    | -0.018                    | -0.385  | 0.701   | $0.701 > 0.05$ | Accepted                      |
| $x_{16}$ | 0.012                     | 0.256   | 0.798   | $0.798 > 0.05$ | Accepted                      |
| $x_{19}$ | -0.014                    | -0.299  | 0.765   | $0.765 > 0.05$ | Accepted                      |

**Table 30: Iron (Fe) Element Content ( $y_6$ ) - test the significance of x-variables with near-zero coefficients**

| Variable | Coefficient ( $\beta_i$ ) | t-value | p-value | Significance   | Hypothesis decision ( $H_0$ ) |
|----------|---------------------------|---------|---------|----------------|-------------------------------|
| $x_2$    | 0.007                     | 0.270   | 0.788   | $0.788 > 0.05$ | Accepted                      |
| $x_8$    | 0.013                     | 0.344   | 0.731   | $0.731 > 0.05$ | Accepted                      |
| $x_{18}$ | 0.004                     | 0.075   | 0.941   | $0.941 > 0.05$ | Accepted                      |
| $x_{22}$ | -0.008                    | -0.149  | 0.881   | $0.881 > 0.05$ | Accepted                      |
| $x_{29}$ | 0.001                     | 0.023   | 0.982   | $0.982 > 0.05$ | Accepted                      |

As with the first model,  $X_{18}$  does not significantly affect two dependent variables,  $Y_4$  and  $Y_6$ . Moreover, the acceptance or rejection of the null hypothesis is subject to an error percentage ( $\alpha = 5\%$ ). However,  $Y_5$  is affected by  $X_{18}$  as the Silicon content in fines and Sulfur content in anode ( $Y_5$ ) depend on supplier sources, and changing supplier sources would change all impurities in fines and anodes. Equations (36) to (38) present the expressions for the second model's three dependent variables, which are the  $CO_2$  reactivity residue, Sulfur (S) element content, and Iron (Fe) element content, respectively. Due to data preprocessing, the intercept is not included in the

expressions, as with the first model. Regarding the independent variables that are treated as discrete, the same concern about significance in a broader range of levels in the first model applies to this model.

$$\begin{aligned}\hat{y}_4 = & 0.049x_1 + 0.037x_2 + 0.030x_3 + 0.051x_5 - 0.084x_6 + 0.109x_7 - 0.104x_9 \\ & + 0.133x_{10} + 0.102x_{11} + 0.082x_{13} - 0.046x_{14} - 0.039x_{20} \\ & + 0.028x_{21} - 0.024x_{22} - 0.043x_{23} - 0.111x_{25} - 0.023x_{27} \\ & - 0.026x_{28} + 0.106x_{29} + 0.203x_{30} - 0.012x_{31}\end{aligned}\quad (36)$$

$$\begin{aligned}\hat{y}_5 = & 0.084x_2 + 0.074x_3 - 0.025x_5 - 0.053x_6 + 0.068x_7 - 0.090x_8 - 0.061x_9 \\ & - 0.060x_{10} + 0.036x_{11} + 0.062x_{12} - 0.024x_{13} + 0.031x_{14} \\ & - 0.030x_{15} + 0.053x_{17} + 0.055x_{18} - 0.050x_{20} - 0.217x_{21} \\ & + 0.035x_{22} + 0.272x_{23} + 0.039x_{24} + 0.057x_{25} + 0.076x_{26} \\ & - 0.035x_{27} + 0.041x_{28} + 0.033x_{29} - 0.232x_{30} + 0.054x_{31} \\ & - 0.055x_{32}\end{aligned}\quad (37)$$

$$\begin{aligned}\hat{y}_6 = & -0.034x_1 - 0.029x_3 - 0.041x_5 - 0.021x_6 + 0.040x_7 + 0.055x_9 - 0.186x_{10} \\ & + 0.024x_{11} - 0.022x_{12} - 0.088x_{13} + 0.025x_{14} + 0.049x_{15} \\ & + 0.013x_{16} + 0.024x_{17} - 0.057x_{19} + 0.028x_{20} - 0.025x_{21} \\ & + 0.115x_{23} + 0.015x_{24} + 0.104x_{25} + 0.054x_{26} - 0.033x_{27} \\ & + 0.023x_{28} - 0.117x_{30} + 0.053x_{31} - 0.097x_{32}\end{aligned}\quad (38)$$

#### 4.4.5 Chemical and Reactivity Properties Model Interpretation

The chemical and reactivity properties model was interpreted using the same methods applied to the physical properties model: the score plot, loading plot, and VIP method. This model did not utilize the contribution plot, as it would not provide additional insights. Since the datasets for both models cover the same period, the same blends were used, and the same variables

would explain the shifts between data groups. The only difference lies in the magnitude of the dataset.

The score plot, the first interpretation tool used, is shown in Figure 27. The plot explains 22.1% of the independent variables' variance using the first two latent variables, accounting for approximately 30% of the total variability the model explains, which is 80.9%. Like the first model, the score plot reveals five distinct data groups, each exhibiting similar patterns in raw material properties or process parameters. The transitions between groups are likely due to the same reasons as in the first model. However, this plot indicates that data from Group 1 and some from Group 5 are not fully explained by the first two LVs, suggesting that additional LVs are needed to capture more data space.

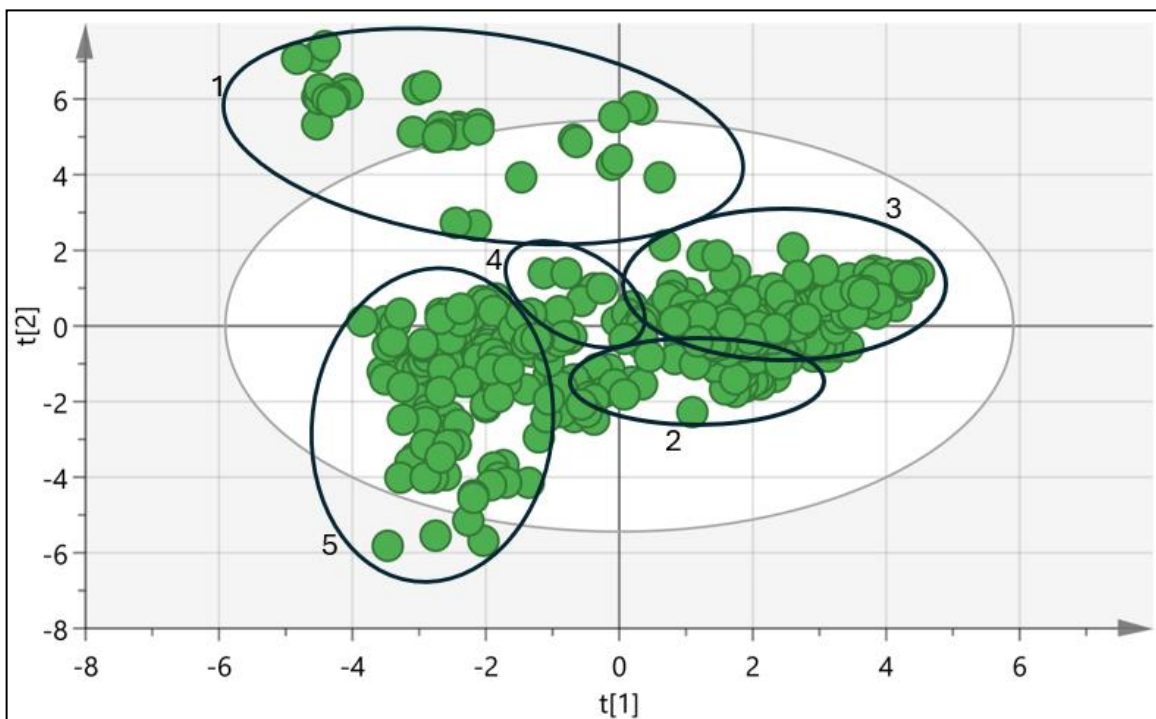


**Figure 27: The 2D score plot for the first two latent variables of the chemical and reactivity properties model**

After adding the third LV to the data space, the 3D score plot is presented in Figure 28. The third LV contributed an additional 7.9% to the explained variance of the X-variables, bringing the total explained X-variable variance to 37.9%, which accounts for 46.8% of the total variability explained by the 9 LVs of the model. With the inclusion of the third LV, some data from group 1 entered the explained X-variable space. However, most of Group 1 data and some from Group 5 remain outside the space explained by the first three LVs. This highlights the need for more LVs to capture more of the dataset, justifying the use of 9 LVs in the model.

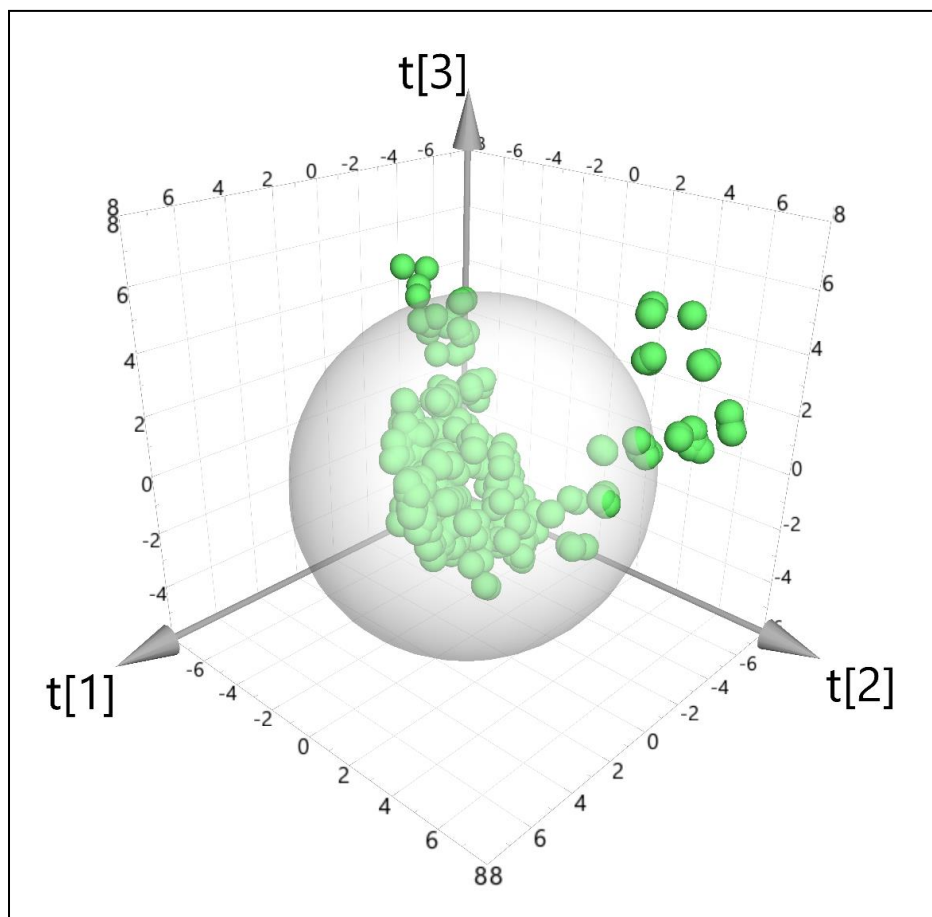


**Figure 28: The 3D score plot for the first two latent variables of the chemical and reactivity properties model**

The second interpretation method is the loading plot, which helps explain the PLS model's relationships between X and Y variables. The loading plot for the chemical and reactivity properties model, based on the first two LVs, is shown in Figure 29. Given that the number of variables is the same in both models, the same challenge of visualizing all correlations between variables has arisen. However, the most significant relationships and correlations are visible and can be thoroughly analyzed.

Regarding loading weights, fire soaking time, CPC iron content, CPC sulfur content, and pitch coking value are the most influential variables in the model for the first LV, as they have the largest absolute loading values. For the second LV, the paste plant process parameters significantly influence the model. This indicates that raw material properties have a greater impact on this model than the first model, as most raw material variables represent chemical properties.

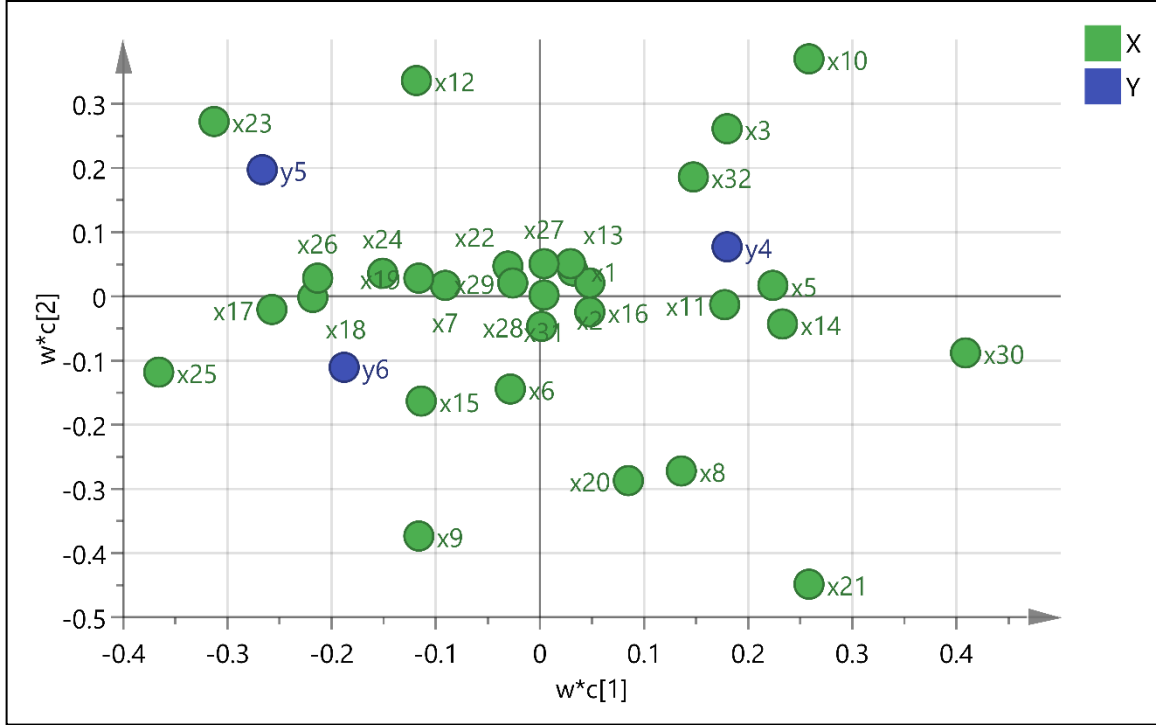


**Figure 29: Loading plot of the first two latent variables of the chemical and reactivity properties model**

The two closely aligned variables are anode sulfur content and CPC sulfur content, as the CPC supplier sources used in anode production play a vital role in determining the sulfur content of the carbon anodes. Similarly, anode iron content is positively correlated with CPC iron content in both LVs' loading values. Additionally, CO<sub>2</sub> reactivity residue is inversely affected by CPC sulfur content, which is consistent with findings in the literature. Furthermore, CO<sub>2</sub> reactivity residue is positively influenced by fire soaking time as per the first LV.

In addition, the third LV loading weights are presented in Figure 30, where the vertical axis represents LV 3 in a plot of LV 1 versus LV 3. The plot shows that fine particles percentage, anode forming speed, and fine sodium content are the most influential variables based on the third LV.

Moreover, final anode temperature also emerges as a key variable in the third LV. The plot further indicates that anode sulfur content falls within the same quadrant of the scatter plot as the CPC and pitch sulfur content variables, as they share the same sign for both LV 1 and LV 3. This provides additional evidence of the positive correlations between these variables.

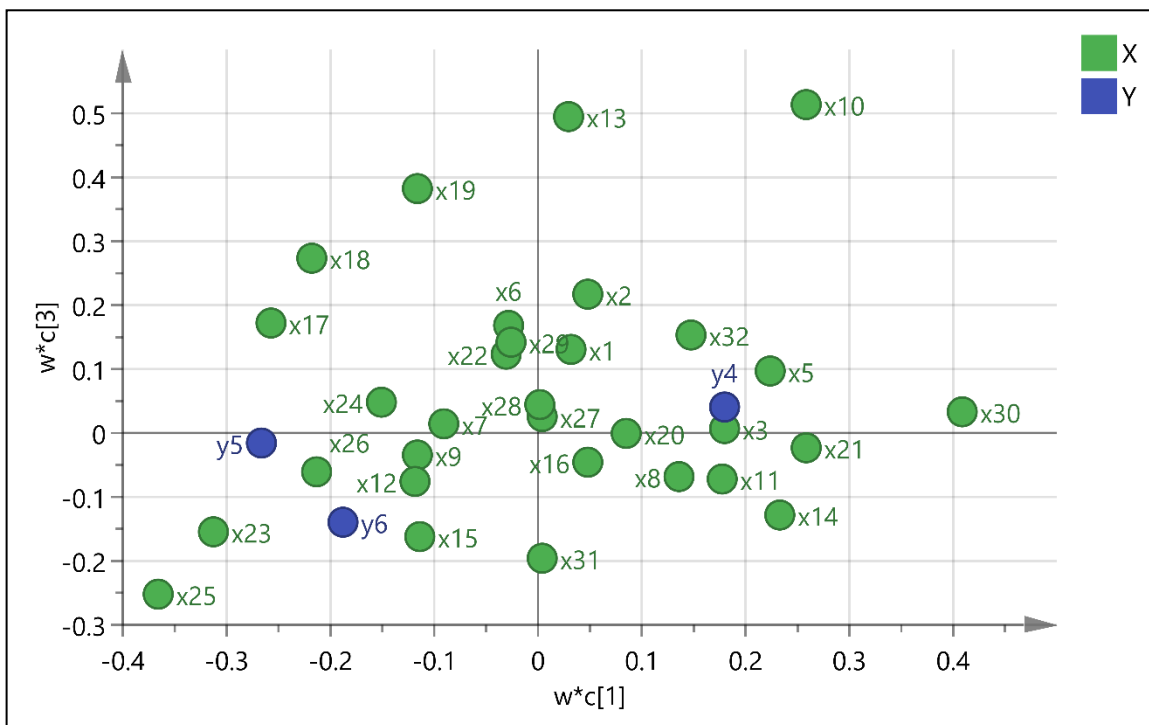


**Figure 30: Loading plot of the latent variables 1 and 3 of the chemical and reactivity properties model**

The VIP method was the last interpretation method to investigate the chemical and reactivity properties model. Table 31 presents the most important X-variable VIP values, their ranking, and the process they represent. The table includes the most influential and important variables for predicting Y-variables. The full list of VIP values for all X-variables of the chemical and reactivity properties model is provided in Appendix (7B).

The fire soaking time is the most important variable in the chemical and reactivity model, and this is expected and consistent with the methods discussed previously. The raw material properties significantly impact the chemical and reactivity properties model. Therefore, these properties play an essential role in process variability and quality. Moreover, their impact is higher compared with the first model. Furthermore, fine particles percentage ( $X_{10}$ ) has a significant effect on the chemical and reactivity properties model with fourth rank, and this is consistent with literature as fine particles have a high percentage of impurity elements which in the end would affect the chemical properties of the anodes produced.

**Table 31: VIP values for X-variables of the chemical and reactivity properties model**

| <b>Rank</b> | <b>VIP Value</b> | <b>X variable ID</b> | <b>X variable</b>            | <b>Process</b> |
|-------------|------------------|----------------------|------------------------------|----------------|
| 1           | 1.936            | $x_{30}$             | Fire soaking time            | Baking kiln    |
| 2           | 1.679            | $x_{25}$             | CPC Iron element content     | Raw material   |
| 3           | 1.583            | $x_{23}$             | CPC Sulfur element content   | Raw material   |
| 4           | 1.516            | $x_{10}$             | Fine particles percentage    | Paste plant    |
| 5           | 1.484            | $x_{21}$             | Pitch coking value           | Raw material   |
| 6           | 1.206            | $x_{17}$             | Fine Iron element content    | Raw material   |
| 7           | 1.177            | $x_{14}$             | Anode forming time           | Paste plant    |
| 8           | 1.076            | $x_{18}$             | Fine Silicon element content | Raw material   |
| 9           | 1.038            | $x_3$                | Mixing power                 | Paste plant    |
| 10          | 1.032            | $x_{11}$             | Medium particles percentage  | Paste plant    |

#### 4.4.6 Prediction of New Chemical and Reactivity Properties Dataset

As with the physical properties model, a dataset is left to predict the dependent variables. However, the prediction dataset for the chemical and reactivity properties is 136 observations as the volume of the chemical and reactivity model data is much less than the physical properties model, but still 25% of the total dataset. Table 32 presents the variance of predicted versus observed values of the three Y-variables. The average variance is 57.15%, higher than the first model prediction variance by 6.02%. This is because the prediction dataset is less variable than the first model dataset. This difference is normal, as both models demonstrate good predictive ability, with insignificant variance differences.

**Table 32: Observed versus predicted dependent variables variance values of the chemical and reactivity properties model**

| Y-Variable ID | Variable                           | Variance ( $R^2$ ) |
|---------------|------------------------------------|--------------------|
| $y_4$         | CO <sub>2</sub> reactivity residue | 57.92%             |
| $y_5$         | Sulfur (S) element content         | 60.63%             |
| $y_6$         | Iron (Fe) element content          | 52.91%             |

Figures 31, 32, and 33 present the predicted versus observed results of CO<sub>2</sub> reactivity residue, sulfur element content, and iron element content, respectively. The black line represents the predicted values, and the blue dotted line represents the observed values. These figures are clearer compared with the first model figures due to the low number of observations of the prediction set. The prediction ability is very good, and the model captures the pattern of increasing and decreasing dependent variables.

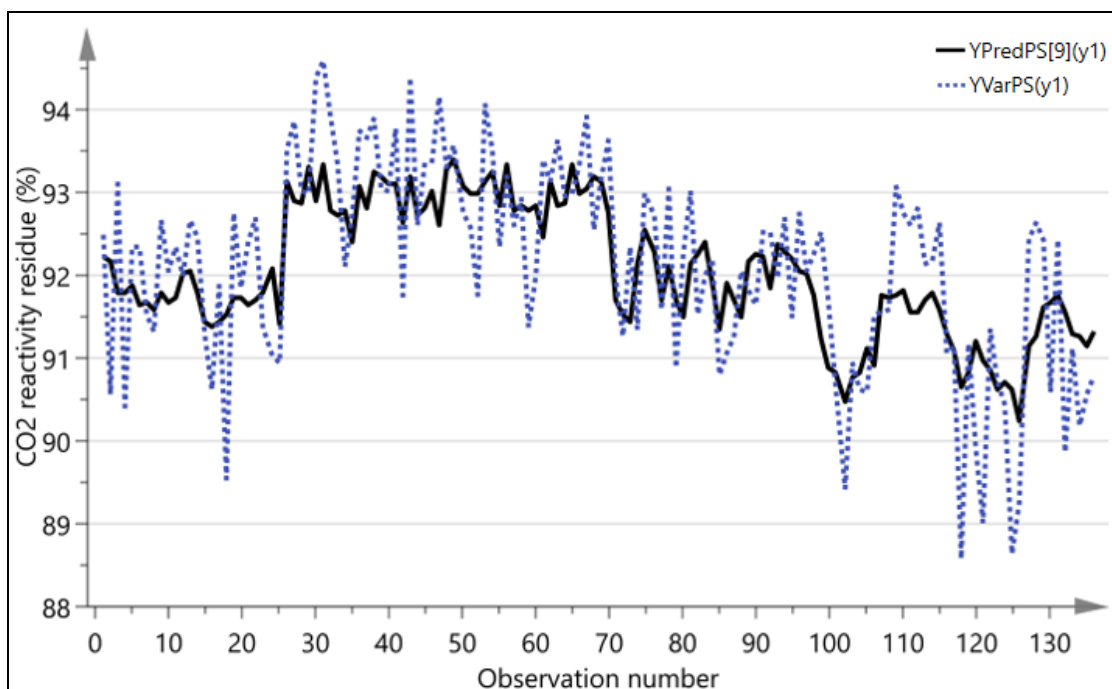


**Figure 31: Comparison between observed versus predicted CO<sub>2</sub> reactivity residue (y<sub>4</sub>)**

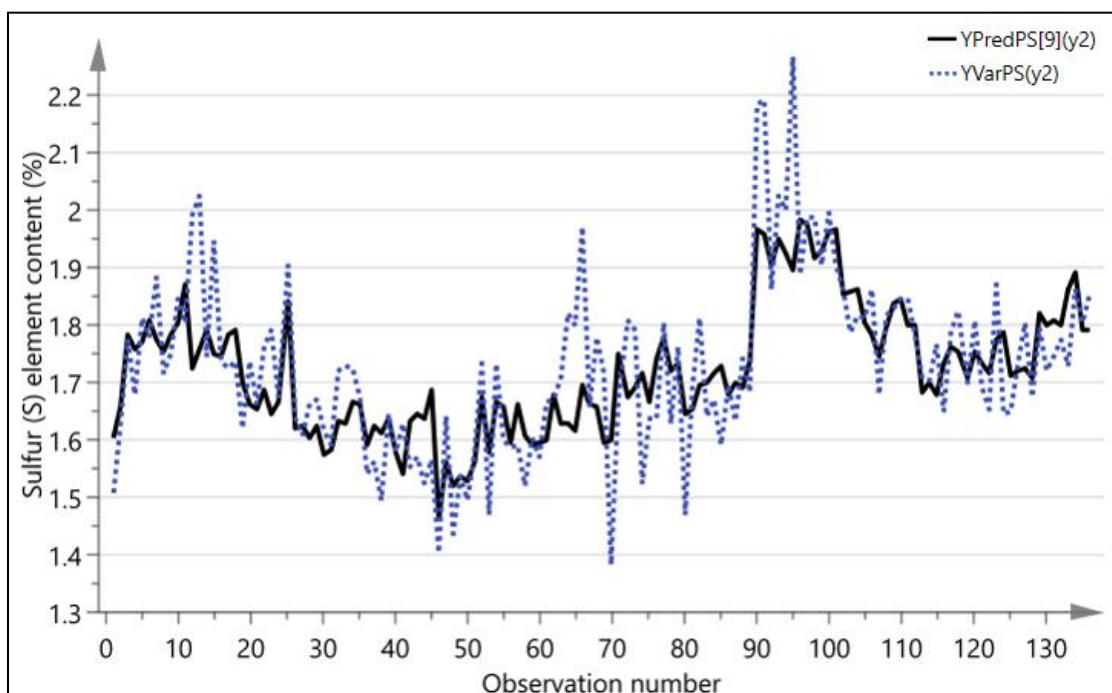


**Figure 32: Comparison between observed versus predicted sulfur (S) element content (y<sub>5</sub>)**

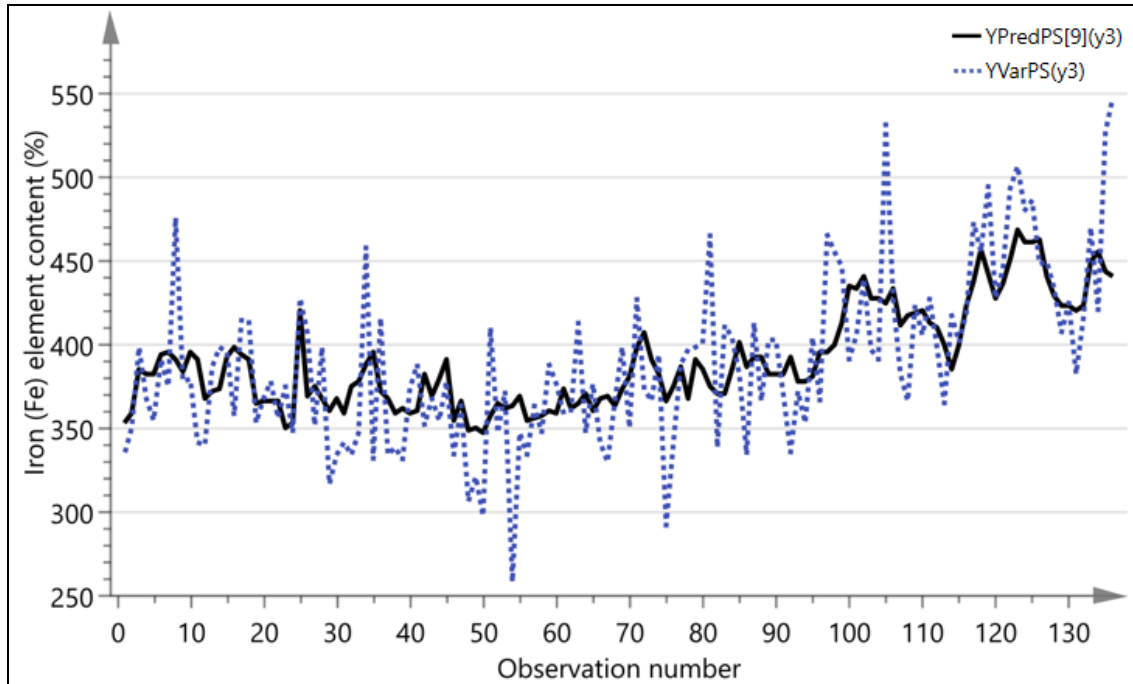


**Figure 33: Comparison between observed versus predicted iron (Fe) element content ( $y_6$ )**

## 4.5 Net Carbon Consumption of PLS Model Built

The last step in the research study is calculating the NCC reduction due to the model built, which could be used to estimate the cost reduction using the PLS model. The actual NCC values of Carbon Plant 2 for the study period were collected and found to be 412.17 kg per ton of aluminum produced.

Equation (32) was used to calculate the NCC reduction between actual NCC and estimated NCC using the PLS model. Appendix (8) summarized the calculation of the PLS model's NCC reduction, which was found to be 1.65 kg per ton of aluminum. After that, the cost reduction was estimated based on the literature correlation between cost reduction and NCC value, in which reducing NCC by 1 kg would result in a cost reduction of half a Million US

dollars. The cost reduction estimation was found to be 825,000 US dollars annually.

## **4.6 Chapter Conclusion**

This chapter illustrates the results for the two models built: the physical properties model (model 1) and the chemical and reactivity properties model (model 2). Before model building, different steps were taken to prepare the dataset, such as assigning the type of X-variables, excluding non-contributory X-variables, and removing outlier observations. Then, the results of the two models are illustrated, analyzed, and discussed. The models' findings have been interpreted using the main PLS interpretation methods: score plot, contribution plot, loading plot, and VIP method. The most important X-variables for predicting Y-variables have been found. Moreover, the NCC value and cost reduction have been estimated for the PLS model built. Furthermore, the prediction portion of the dataset was used to evaluate the models' prediction ability.

## **Chapter 5: Conclusion and Recommendations**

**5.1 Conclusion ..... 125**

**5.2 Recommendations ..... 127**

## 5.1 Conclusion

The research study investigated the relationship between raw material properties and process parameters against carbon anode manufactured quality in an aluminum smelting industry. The complexity of anodes' manufacturing -many processes and variables- results in variations in the quality of anodes, which explains the absence of a mathematical prediction model of the anode quality.

Hence, a multivariate mathematical model of the anode quality was developed based on actual data from a carbon plant. The model was built using SIMCA software, which is popular for multivariate modeling, such as PCA and PLS.

Not all anode cores are tested with all physical, chemical, and reactivity properties; the dataset collected has been used to build two separate PLS models. Moreover, the non-contributory independent variables and invalid and outlier observations were removed from the dataset before the model was built. The dataset has been divided into two sets: a training set to build the model and a prediction set to validate the model's prediction ability.

After preparing the dataset, two models were developed. The physical properties model showed strong explanatory power, with  $R_y^2$  of 64.0%, using 85.2% of the X-variables variance ( $R_x^2$ ). The regression coefficients highlighted a strong dependency on green anode density, with baking process parameters also significantly influencing physical properties.

The second model, focused on chemical and reactivity properties, also demonstrated strong explanatory power, with  $R_y^2$  of 64.1%, similar to the first model. However, as expected, its  $R_x^2$  was lower since most X-variables

represent physical properties. The most influential variables for this model were the chemical properties of the raw materials.

The prediction ability of both models was high, as the variances of observed versus predicted Y-variables were 51.1% and 57.2% for the first and second models, respectively. The prediction values followed the same pattern of increasing and decreasing as the observed values, suggesting a very good prediction ability of the models.

The primary source of variation in both models was the variability in raw material properties, particularly CPC properties. For the first model, the most influential variables were green anode density and CPC sulfur content. CPC properties and baking kiln process parameters had a significant impact on the first PLS model. Based on VIP values, green anode density was the most important variable, followed by heating rate. Additionally, many baking kiln and paste plant process parameters played important roles in the model's projection.

The CPC iron content and paste plant process parameters significantly influence the second PLS model. Moreover, raw material properties have a greater impact on the second model than on the first model. Based on VIP values, the most important variable of the second model was fire soaking time. Furthermore, the raw material properties variables are highly important to the model as their VIP values are the highest.

Lastly, the NCC reduction value between the actual and estimated using the PLS model built has been calculated and found to be 1.65 kg per ton of aluminum. Accordingly, the cost reduction is estimated based on the NCC reduction value and found to be 825,000 US dollars per year.

## 5.2 Recommendations

The research findings suggest several recommendations to improve the anode quality and prediction ability of the PLS model:

- Focus on green anode density as it is the most important variable and the core of the quality of paste plant process parameters.
- Reduce the CPC supplier sources because increasing sources produces variation in anode quality.
- Study the impact of some variables not included in the PLS model to increase the model's predictability.
- Consider removing cooling power ( $x_4$ ) from the variables measured to control and improve the quality of anodes as it does not affect both physical and chemical properties.
- Retrain the model using an updated dataset whenever new process parameters are introduced or additional supplier sources for raw materials are incorporated. This ensures the model remains accurate and relevant by reflecting the system's changes.

The research results are limited to some constraints on which the PLS model has been built. The model was built for Carbon 2 in ALBA according to the Aluminum Pechiney technology of that plant, so the variables were based on this technology. Moreover, the baking kiln process parameters were available weekly instead of continuous tracking and were limited to four variables only. Furthermore, data synchronization for anode core quality was based on process control expertise rather than system tracking. Lastly, the dataset anode cores were not tested with all properties for all observations. Therefore, future studies are recommended to overcome the mentioned research limitations:

- To have an advanced system of synchronizing data between anode core quality, raw material properties, and baking kiln process parameters.
- To collect more baking kiln data as the four variables representing long process are much lower.
- A data system for baking kilns, like paste plants, would collect process parameters continuously instead of weekly.
- Reduce the number of CPC supplier sources to reduce variation in anode quality produced and improve the model's predictivity.
- To measure all properties of anode cores, to have one comprehensive PLS model for all properties instead of having two models representing different property types.
- To analyze a dataset with a broader range of levels for the independent variables treated as discrete in order to investigate their significance on the dependent variables.
- To investigate the interaction effect between X-variables, which could help explain some Y-variables that are not captured by assuming independence.
- To screen the X variables to remove less relevant predictors and enhance model performance using the PLS method and other modeling techniques.

## List of References

- Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley interdisciplinary reviews: computational statistics*, 2(1), 97-106. <https://doi.org/10.1002/wics.51>
- Alexopoulos, E. C. (2010). Introduction to multivariate regression analysis. *Hippokratia*, 14 (Suppl 1), 23. PMID: 21487487; PMCID: PMC3049417.
- Aluminium Bahrain. (2020). *Baked Anode Core properties and Lab analysis*. Aluminium Bahrain Manuals.
- Aluminium Bahrain. (2020). *Baking Kiln Process and baked anode quality*. Aluminium Bahrain Manuals.
- Aluminium Bahrain. (2020). *Paste Plant Process and Operation*. Aluminium Bahrain Manuals.
- Amrani, S., Kocaefe, D., Kocaefe, Y., Bhattacharyay, D., Bouazara, M., & Côté, J. (2021). Effect of carbon anode production parameters on anode cracking. *SN Applied Sciences*, 3, 1-19. <https://doi.org/10.1007/s42452-020-04133-8>
- Bertrand, F., Magnanensi, J., Meyer, N., & Maumy-Bertrand, M. (2014). plsRglm: Algorithmic insights and applications. *Edited: June*.
- Bhattacharyay, D., Kocaefe, D., Kocaefe, Y., & Morais, B. (2016). Comparison of linear multivariable, partial least square regression, and artificial neural network analyses to study the effect of different parameters on anode properties. *Light Metals* 2015, 1129-1134. [https://doi.org/10.1007/978-3-319-48248-4\\_189](https://doi.org/10.1007/978-3-319-48248-4_189)

Bhattacharyay, D., Kocaefe, D., Kocaefe, Y., & Morais, B. (2017). An artificial neural network model for predicting the CO<sub>2</sub> reactivity of carbon anodes used in the primary aluminum production. *Neural computing and Applications*, 28, 553-563. <https://doi.org/10.1007/s00521-015-2093-7>

Bhattacharyay, D., Kocaefe, D., Kocaefe, Y., Morais, B., & Gagnon, M. (2016). Application of the artificial neural network (ANN) in predicting anode properties. *Light Metals 2013*, 1189-1194. [https://doi.org/10.1007/978-3-319-65136-1\\_201](https://doi.org/10.1007/978-3-319-65136-1_201)

Breitkopf, C., & Swider-Lyons, K. (Eds.). (2016). *Springer handbook of electrochemical energy*. Springer.

Burnham, A. J., MacGregor, J. F., & Viveros, R. (1999). Latent variable multivariate regression modeling. *Chemometrics and Intelligent Laboratory Systems*, 48(2), 167-180. [https://doi.org/10.1016/S0169-7439\(99\)00018-0](https://doi.org/10.1016/S0169-7439(99)00018-0)

Chamam, Y., Kocaefe, D., Kocaefe, Y., Bhattacharyay, D., & Morais, B. (2016). Effect of heating rate during baking on the properties of carbon anodes used in aluminum industry. *Light Metals 2016*, 947-951. <https://doi.org/10.1002/9781119274780.ch159>

Danks, N. P., & Ray, S. (2018). Predictions from partial least squares models. In *Applying partial least squares in tourism and hospitality research* (pp. 35-52). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78756-699-620181003>

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change*, 173, 121092. <https://doi.org/10.1016/j.techfore.2021.121092>

DeGroot, M. H., & Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley.

Dufour, A., Tessier, J., Gosselin, L., & Duchesne, C. (2016). Empirical modeling of the baking furnace to predict baked anode properties. *Light Metals 2016*, 953-958. [https://doi.org/10.1007/978-3-319-48251-4\\_160](https://doi.org/10.1007/978-3-319-48251-4_160)

Dunn, K. (2019). Process improvement using data. *Retrieved on*, 11-2016.

Ergon, R. (2004). Informative PLS score-loading plots for process understanding and monitoring. *Journal of Process Control*, 14(8), 889-897. <https://doi.org/10.1016/j.jprocont.2004.02.004>

Eriksson, L., Trygg, J., & Wold, S. (2008). CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 22(11-12), 594-600. <https://doi.org/10.1002/cem.1187>

Farrés, M., Platikanov, S., Tsakovski, S., & Tauler, R. (2015). Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics*, 29(10), 528-536. <https://doi.org/10.1002/cem.2736>

Fischer, W. K., Keller, F., Perruchoud, R. C., & Oderbolz, S. (2016). Baking parameters and the resulting anode quality. In *Essential Readings in Light Metals: Volume 4 Electrode Technology for Aluminum Production* (pp. 427-433). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-48200-2\\_59](https://doi.org/10.1007/978-3-319-48200-2_59)

Flåten, G. R., Grung, B., & Kvalheim, O. M. (2004). A method for validation of reference sets in SIMCA modelling. *Chemometrics and Intelligent*

*Laboratory Systems*, 72(1), 101-109.  
<https://doi.org/10.1016/j.chemolab.2004.03.003>

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Holland, S. M. (2008). Principal components analysis (PCA). *Department of Geology, University of Georgia, Athens, GA, 30602*, 2501.

Hulse, K. L. (2000). *Anode manufacture: raw materials, formulation and processing parameters*. R & D Carbon Limited.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.  
<https://doi.org/10.1007/978-3-031-38747-0>

Joseph, F. (2010). *Multivariate Data Analysis*. Pearson Education Limited.

Juster, T. C. (2013). Using Excel to develop random number sense. *Spreadsheets in Education*, 6(2).  
<http://epublications.bond.edu.au/ejsie/vol6/iss2/4>

Keller, F., & Sulger, P. O. (2006). *Anodes for the Aluminium Industry*. R & D Carbon Limited.

Kettaneh, N., Berglund, A., & Wold, S. (2005). PCA and PLS with very large data sets. *Computational Statistics & Data Analysis*, 48(1), 69-85.  
<https://doi.org/10.1016/j.csda.2003.11.027>

Khaji, K., & Al Qassemi, M. (2016). The role of anode manufacturing processes in net carbon consumption. *Metals*, 6(6), 128.  
<https://doi.org/10.3390/met6060128>

Khaji, K., & Qassemi, M. A. (2016). Factors influencing baked anode properties. *Light Metals 2015*, 1135-1140. [https://doi.org/10.1007/978-3-319-48248-4\\_190](https://doi.org/10.1007/978-3-319-48248-4_190)

Kocaefe, D., Kocaefe, Y., Bhattacharyay, D., Sanogo, B., Ahoutou, Y., Sun, H., & Coulombe, P. (2017). Development of techniques and tools for the determination of carbon anode quality. In *Light Metals 2017* (pp. 1247-1254). Springer International Publishing. [https://doi.org/10.1007/978-3-319-51541-0\\_149](https://doi.org/10.1007/978-3-319-51541-0_149)

Kocaefe, D., Sarkar, A., Lu, Y., Bhattacharyay, D., Kocaefe, Y. S., & Morais, B. (2014). Utilization of artificial neural network to analyze and predict the influence of different parameters on anode properties.

Lauzon-Gauthier, J. (2011). Multivariate latent variable modelling of the pre-baked anode manufacturing process used in aluminum smelting. *Mémoire, Université Laval*.

Lauzon-Gauthier, J., Duchesne, C., & Tessier, J. (2012). A Database approach for predicting and monitoring baked anode properties. *Jom*, 64, 1334-1342. <https://doi.org/10.1007/s11837-012-0433-y>

Lauzon-Gauthier, J., Duchesne, C., Tessier, J., Cantin, K., & Petit, I. (2016). Multivariate monitoring of the prebaked anode manufacturing process and anode quality. *Light Metals 2011*, 967-972. [https://doi.org/10.1007/978-3-319-48160-9\\_165](https://doi.org/10.1007/978-3-319-48160-9_165)

Lazraq, A., & Cléroux, R. (2001). The PLS multivariate regression model: testing the significance of successive PLS components. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 15(6), 523-536. <https://doi.org/10.1002/cem.641>

- Logan, J. D. (2013). *Applied mathematics*. John Wiley & Sons.
- Mehmood, T. (2016). Hotelling  $T^2$  based variable selection in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 154, 23-28. <https://doi.org/10.1016/j.chemolab.2016.03.001>
- Manolescu, P., Duchesne, C., Tessier, J., & Saevarsdottir, G. (2020). Estimation of net carbon consumption in aluminum electrolysis using multivariate analysis. *Journal of Sustainable Metallurgy*, 6, 739-750. <https://doi.org/10.1007/s40831-020-00310-6>
- Meier, H. M. (1996). *Cracking Behaviour of Anodes*. R & D Carbon Limited.
- Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers*. John wiley & sons.
- Nayak, S., & Katharbatcha, M. (2021). Baking Furnace Optimizations at Aditya to Maintain Consistent Quality and Productivity. In *Light Metals 2021: 50th Anniversary Edition* (pp. 984-991). Springer International Publishing. [https://doi.org/10.1007/978-3-030-65396-5\\_129](https://doi.org/10.1007/978-3-030-65396-5_129)
- Ndjom, B., Malik, M. S., Al Marzouqi, A., Sahu, T. K., & Rabba, S. A. (2016). Improving anode baked density and air permeability through process optimization and coke blending. *Light Metals 2013*, 1105-1110. [https://doi.org/10.1007/978-3-319-65136-1\\_187](https://doi.org/10.1007/978-3-319-65136-1_187)
- Orme, J. G., & Combs-Orme, T. (2009). *Multiple regression with discrete dependent variables*. Oxford University Press.
- Sijabat, S. S., Ashad, F., Buandra, A., & Mugiono, E. (2017). Anode Quality Improvement at Inalum Smelter. In *Light Metals 2017* (pp. 1213-1219).

Springer International Publishing. [https://doi.org/10.1007/978-3-319-51541-0\\_145](https://doi.org/10.1007/978-3-319-51541-0_145)

Silva, A. F., Vercruysse, J., Vervaet, C., Remon, J. P., Lopes, J. A., De Beer, T., & Sarraguça, M. C. (2019). In-depth evaluation of data collected during a continuous pharmaceutical manufacturing process: a multivariate statistical process monitoring approach. *Journal of pharmaceutical sciences*, 108(1), 439-450. <https://doi.org/10.1016/j.xphs.2018.07.033>

Smith, L. I. (2002). A tutorial on principal components analysis.

Sommerseth, C. (2016). The effect of production parameters on the performance of carbon anodes for aluminium production. <http://hdl.handle.net/11250/2387203>

Stockburger, D. W. (1998). *Multivariate statistics: Concepts, models, and applications*. David W. Stockburger. <https://dwstockburger.com/Multibook/Mlt07.htm>

Tessier, J., Duchesne, C., Tarcy, G. P., Gauthier, C., & Dufour, G. (2012). Multivariate analysis and monitoring of the performance of aluminum reduction cells. *Industrial & engineering chemistry research*, 51(3), 1311-1323. <https://doi.org/10.1021/ie201258b>

Wichern, D. (2014). *Applied Multivariate Statistical Analysis*. Pearson Education Limited.

Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, 20(4), 397-405. <https://doi.org/10.1080/00401706.1978.10489693>

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), 109-130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

Worley, B., Halouska, S., & Powers, R. (2013). Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical biochemistry*, 433(2), 102-104. <https://doi.org/10.1016/j.ab.2012.10.011>

# **List of Appendices**

**Appendix 1:** Sample of data collected on anode quality, raw material properties, and process parameters.

**Appendix 2:** Anode quality properties after preparation.

**Appendix 3:** Data collected preparation after adding paste plant process parameters.

**Appendix 4:** Final data collected ready to use for building model.

**Appendix 5:** SIMCA software output.

**Appendix 6:** Additional details on the statistical testing parameters.

**Appendix 7:** Full VIP plot values.

**Appendix 8:** Calculation of NCC based on the model built.

## **Appendix 1:**

### **Sample of data collected of anode quality, raw material properties, and process parameters.**

**This Appendix contains a sample of the data collected from the carbon plant. The following figures represent the data of each subset:**

**Section A: Anode quality properties data (Laboratory analysis).**

**Section B: Calcined petroleum coke (CPC) properties data.**

**Section C: Butts properties data.**

**Section D: Fines properties data.**

**Section E: Coal tar pitch (CTP) properties data.**

**Section F: Paste plant process parameters data.**

**Section G: Baking kiln process parameters data.**

## Section A: Anode quality properties data (Laboratory analysis).

Note: The sample of data represents the first week of February 2021, and the data represent each anode's core quality properties.

| Week | Kiln | Date       | Sample ID         | Anode Code | Geometrical Density | Elect. Resis. Sample Average | Flexural strength | Youngs Modulus | Coefficient Thermal expansion | TSR   | Critical Strain | Thermal Conductivity | Air permeability | Ash   | CO2 Reactivity Residue | CO2 Reactivity Dust | CO2 Reactivity Loss | Air Reactivity Residue | Air Reactivity Dust | Air Reactivity Loss | Real Density | Crystallite size | Sulphur % XRF percentage | Vanadium ppm XRF | Iron ppm XRF | Calcium ppm XRF | Silicone ppm XRF | Nickel ppm XRF | Sodium ppm XRF |
|------|------|------------|-------------------|------------|---------------------|------------------------------|-------------------|----------------|-------------------------------|-------|-----------------|----------------------|------------------|-------|------------------------|---------------------|---------------------|------------------------|---------------------|---------------------|--------------|------------------|--------------------------|------------------|--------------|-----------------|------------------|----------------|----------------|
|      |      |            |                   |            | G/CM3               | UOHMM                        | N/MM2             | GPA            | 1/K                           | 1/K   |                 | W/MK                 | NPM              | %     | %                      | %                   | %                   | %                      | %                   | %                   | G/CM3        | Å                | %                        | ppm              | ppm          | ppm             | ppm              | ppm            | ppm            |
| 1    | 3    | 01/02/2021 | S-20210202-000121 | 121010967  | 1.586               | 5000                         | 13.69             | 8.5            |                               |       | 1.612           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 33.9             | 1.76                     | 198              | 346          | 115             | 30               | 102            | 137            |
| 1    | 3    | 01/02/2021 | S-20210202-000122 | 221011435  | 1.594               | 5580                         | 12.6              | 8.5            | 4.364                         | 1.579 | 1.489           | 4.625                | 0.330            |       |                        |                     |                     |                        |                     |                     |              | 31.7             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 01/02/2021 | S-20210202-000123 | 121020712  | 1.584               | 5863                         | 13.51             | 8.7            |                               |       | 1.548           |                      |                  |       |                        |                     |                     | 62.3                   | 6.46                | 31.3                | 2.11         | 32.8             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 01/02/2021 | S-20210202-000124 | 221021167  | 1.582               | 5270                         | 11.94             | 8.3            | 4.364                         | 1.222 | 1.432           | 3.724                | 0.290            |       |                        |                     |                     |                        |                     |                     |              | 34.3             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 01/02/2021 | S-20210202-000125 | 221020166  | 1.583               | 5270                         | 12.74             | 8.5            |                               |       | 1.502           |                      |                  | 0.330 | 91.6                   | 1.75                | 6.64                |                        |                     |                     | 2.11         | 34.1             | 1.7                      | 204              | 324          | 79              | 26               | 100            | 151            |
| 1    | 3    | 02/02/2021 | S-20210202-000740 | 121011124  | 1.594               | 5383                         | 12.57             | 8.7            |                               |       | 1.445           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 33.7             | 1.82                     | 215              | 408          | 103             | 38               | 102            | 202            |
| 1    | 3    | 02/02/2021 | S-20210202-000741 | 121013212  | 1.584               | 5287                         | 11.84             | 8.5            | 4.364                         | 1.326 | 1.388           | 4.168                | 0.610            |       |                        |                     |                     |                        |                     |                     |              | 32.6             |                          |                  |              |                 |                  |                |                |
| 1    | 3    | 02/02/2021 | S-20210202-000742 | 121013637  | 1.595               | 5253                         | 12.36             | 8.8            |                               |       | 1.401           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 32.9             | 1.77                     | 207              | 361          | 101             | 28               | 98             | 165            |
| 1    | 3    | 02/02/2021 | S-20210202-000743 | 221011123  | 1.589               | 5327                         | 11.91             | 8.6            | 4.364                         | 1.242 | 1.383           | 3.918                | 0.380            |       |                        |                     |                     |                        |                     |                     |              | 33.8             |                          |                  |              |                 |                  |                |                |
| 1    | 3    | 02/02/2021 | S-20210202-000744 | 221011167  | 1.582               | 5470                         | 11.12             | 8.6            |                               |       | 1.301           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 35.7             | 1.76                     | 208              | 341          | 70              | 43               | 102            | 126            |
| 1    | 3    | 03/02/2021 | S-20210203-000308 | 121010910  | 1.579               | 5180                         | 13.56             | 8.5            | 4.364                         | 1.479 | 1.597           | 4.040                | 1.550            |       |                        |                     |                     |                        |                     |                     |              | 33.7             |                          |                  |              |                 |                  |                |                |
| 1    | 3    | 03/02/2021 | S-20210203-000309 | 221010915  | 1.579               | 5127                         | 12.47             | 7.9            |                               |       | 1.580           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 35.4             | 1.81                     | 212              | 373          | 106             | 31               | 102            | 176            |
| 1    | 3    | 03/02/2021 | S-20210203-000310 | 221010921  | 1.590               | 4990                         | 13.08             | 8.1            | 4.364                         | 1.746 | 1.607           | 4.742                | 0.380            |       |                        |                     |                     |                        |                     |                     |              | 35.9             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 03/02/2021 | S-20210203-000311 | 121021461  | 1.586               | 5170                         | 12.04             | 8.1            | 4.364                         | 1.086 | 1.494           | 3.173                | 0.470            |       |                        |                     |                     |                        |                     |                     |              | 35.9             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 03/02/2021 | S-20210203-000312 | 221021316  | 1.594               | 5260                         | 12.96             | 8.5            |                               |       | 1.521           |                      |                  |       |                        |                     |                     | 72.5                   | 4.94                | 22.5                | 2.11         | 34.1             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 04/02/2021 | S-20210204-000221 | 121021621  | 1.591               | 4967                         | 11.5              | 8.1            | 4.364                         | 1.194 | 1.416           | 3.679                | 0.540            |       |                        |                     |                     |                        |                     |                     |              | 35.4             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 04/02/2021 | S-20210204-000222 | 121021619  | 1.590               | 4767                         | 10.94             | 9              |                               |       | 1.220           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 34.1             | 1.82                     | 229              | 348          | 111             | 57               | 111            | 130            |
| 2    | 3    | 04/02/2021 | S-20210204-000223 | 121021144  | 1.594               | 5013                         | 12.96             | 9.1            | 4.364                         | 1.177 | 1.423           | 3.612                | 0.610            |       |                        |                     |                     |                        |                     |                     |              | 32.8             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 04/02/2021 | S-20210204-000224 | 221021620  | 1.606               | 4717                         | 12.78             | 9.8            |                               |       | 1.307           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 32.4             | 1.76                     | 218              | 333          | 99              | 53               | 106            | 110            |
| 2    | 3    | 04/02/2021 | S-20210204-000225 | 221021619  | 1.582               | 4977                         | 12.17             | 8.3            | 4.364                         | 1.419 | 1.463           | 4.232                | 0.440            |       |                        |                     |                     |                        |                     |                     |              | 34.6             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 06/02/2021 | S-20210206-000743 | 121021557  | 1.591               | 5107                         | 11.83             | 8.2            |                               |       | 1.443           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 35.2             | 1.63                     | 202              | 386          | 113             | 37               | 103            | 173            |
| 2    | 3    | 06/02/2021 | S-20210206-000744 | 221021561  | 1.586               | 5160                         | 11.73             | 8.6            | 4.364                         | 1.478 | 1.369           | 4.712                | 0.590            |       |                        |                     |                     |                        |                     |                     |              | 35.1             |                          |                  |              |                 |                  |                |                |
| 3    | 3    | 06/02/2021 | S-20210206-000745 | 121030486  | 1.602               | 5117                         | 11.74             | 8.7            |                               |       | 1.349           |                      |                  | 0.508 | 90.7                   | 1.84                | 7.48                |                        |                     |                     | 2.12         | 35.4             | 1.63                     | 198              | 536          | 108             | 31               | 102            | 186            |
| 3    | 3    | 06/02/2021 | S-20210206-000746 | 121030327  | 1.584               | 5190                         | 12.77             | 8.7            | 4.364                         | 1.390 | 1.473           | 4.119                | 0.400            |       |                        |                     |                     |                        |                     |                     |              | 35.1             |                          |                  |              |                 |                  |                |                |
| 3    | 3    | 06/02/2021 | S-20210206-000747 | 121030234  | 1.603               | 5097                         | 11.97             | 8.9            |                               |       | 1.353           |                      |                  |       |                        |                     |                     | 69.7                   | 4.60                | 25.7                | 2.12         | 35.3             |                          |                  |              |                 |                  |                |                |
| 3    | 3    | 06/02/2021 | S-20210206-000748 | 221030649  | 1.581               | 5473                         | 9.87              | 8.1            | 4.364                         | 1.413 | 1.216           | 5.074                | 0.490            |       |                        |                     |                     |                        |                     |                     |              | 35.6             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 07/02/2021 | S-20210207-000779 | 121021583  | 1.585               | 6530                         | 6.03              | 6.5            |                               |       | 0.922           |                      |                  |       |                        |                     |                     | 68.9                   | 4.14                | 27.0                | 2.10         | 32.0             |                          |                  |              |                 |                  |                |                |
| 2    | 3    | 07/02/2021 | S-20210207-000780 | 221021581  | 1.585               | 5413                         | 10.34             | 8.4            |                               |       | 1.232           |                      |                  |       |                        |                     |                     |                        |                     |                     |              | 36.6             | 1.86                     | 217              | 478          | 120             | 32               | 107            | 255            |
| 2    | 3    | 07/02/2021 | S-20210207-000781 | 221021830  | 1.585               | 5043                         | 10.65             | 8.5            | 4.364                         | 1.190 | 1.259           | 4.127                | 0.700            |       |                        |                     |                     |                        |                     |                     |              | 36.1             |                          |                  |              |                 |                  |                |                |
| 3    | 3    | 07/02/2021 | S-20210207-000777 | 221030416  | 1.585               | 6530                         | 6.03              | 6.5            |                               |       | 0.922           |                      |                  |       |                        |                     |                     |                        |                     |                     | 2.10         | 32.0             |                          |                  |              |                 |                  |                |                |
| 3    | 3    | 07/02/2021 | S-20210207-000778 | 121030704  | 1.600               | 5320                         | 9.34              | 7.8            | 4.364                         | 1.073 | 1.193           | 3.924                | 0.530            |       |                        |                     |                     |                        |                     |                     |              | 33.9             |                          |                  |              |                 |                  |                |                |

## Section B: Calcined petroleum coke (CPC) properties data.

Note: The sample of data represents the first week of February 2021, and data are presented on a daily basis.

| Blend No. | Sampling Date | S    | V   | Fe  | CO <sub>2</sub> Rec. | +4.75 mm | Lc   |
|-----------|---------------|------|-----|-----|----------------------|----------|------|
|           |               | %    | ppm | ppm | %                    | %        | Å    |
| 30091     | 01-Feb-21     | 1.81 | 290 | 161 | 5.7                  | 30.5     | 26.9 |
| 40149     | 02-Feb-21     | 1.87 | 251 | 140 | 5.5                  | 36.1     | 26.4 |
| 40149     | 03-Feb-21     | 1.90 | 238 | 134 | 4.8                  | 29.2     | 26.3 |
| 40149     | 04-Feb-21     | 1.93 | 248 | 126 | 5.1                  | 28.2     | 28.0 |
| 40149     | 05-Feb-21     | 1.86 | 237 | 105 | 4.9                  | 28.7     | 26.1 |
| 40149     | 06-Feb-21     | 1.79 | 238 | 116 | 4.4                  | 30.5     | 26.2 |
| 40149     | 07-Feb-21     | 1.85 | 266 | 106 | 5.8                  | 27.3     | 26.8 |

### Section C: Butts properties data.

Note: The sample of data represents the year 2021, and data are presented weekly.

| Week NO.       | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          |
|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Receiving Date | 05.01.2021 | 12.01.2021 | 19.01.2021 | 26.01.2021 | 02.02.2021 | 09.02.2021 | 16.02.2021 | 23.02.2021 | 02.03.2021 |
| Fe ppm         | 590        | 411        | 380        | 389        | 402        | 422        | 415        | 425        | 482        |
| Na ppm         | 391        | 115        | 149        | 93         | 94         | 112        | 16         | 146        | 187        |
| Week NO.       | 10         | 11         | 12         | 13         | 14         | 15         | 16         | 17         | 18         |
| Receiving Date | 09.03.2021 | 16.03.2021 | 23.03.2021 | 30.03.2021 | 06.04.2021 | 13.04.2021 | 20.04.2021 | 27.04.2021 | 04.05.2021 |
| Fe ppm         | 518        | 393        | 409        | 501        | 480        | 323        | 443        | 463        | 425        |
| Na ppm         | 86         | 282        | 214        | 158        | 152        | 140        | 362        | 193        | 232        |
| Week NO.       | 19         | 20         | 21         | 22         | 23         | 24         | 25         | 26         | 27         |
| Receiving Date | 11.05.2021 | 18.05.2021 | 25.05.2021 | 01.06.2021 | 08.06.2021 | 15.06.2021 | 22-Jun-21  | 03-Jun-21  | 06-Jul-21  |
| Fe ppm         | 486        | 383        | 425        | 521        | 462        | 346        | 402        | 310        | 317        |
| Na ppm         | 112        | 136        | 143        | 126        | 127        | 174        | 275        | 119        | 20         |
| Week NO.       | 28         | 29         | 30         | 31         | 32         | 33         | 34         | 35         | 36         |
| Receiving Date | 13-Jul-21  | 20-Jul-21  | 27-Jul-21  | 03-Aug-21  | 10-Aug-21  | 17-Aug-21  | 24-Aug-21  | 31-Aug-21  | 07-Sep-21  |
| Fe ppm         | 444        | 426        | 402        | 398        | 334        | 698        | 517        | 467        | 582        |
| Na ppm         | 252        | 148        | 134        | 166        | 95         | 333        | 418        | 302        | 277        |
| Week NO.       | 37         | 38         | 39         | 40         | 41         | 42         | 43         | 44         | 45         |
| Receiving Date | 14-Sep-21  | 21-Sep-21  | 28-Sep-21  | 05-Oct-21  | 12-Oct-21  | 19-Oct-21  | 26-Oct-21  | 02-Nov-21  | 09-Nov-21  |
| Fe ppm         | 379        | 1570       | 1108       | 458        | 640        | 382        | 471        | 447        | 466        |
| Na ppm         | 172        | 1049       | 155        | 787        | 223        | 193        | 272        | 334        | 239        |
| Week NO.       | 46         | 47         | 48         | 49         | 50         | 51         | 52         |            |            |
| Receiving Date | 16-Nov-21  | 23-Nov-21  | 30-Nov-21  | 07-Dec-21  | 14-Dec-21  | 21-Dec-21  | 28-Dec-21  |            |            |
| Fe ppm         | 461        | 385        | 411        | 396        | 426        | 386        | 422        |            |            |
| Na ppm         | 197        | 216        | 188        | 148        | 179        | 146        | 171        |            |            |

## Section D: Fines properties data.

Note: The sample of data represents the year 2021, and data are presented weekly.

| Week NO.       | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          |
|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Receiving Date | 05.01.2021 | 12.01.2021 | 19.01.2021 | 26.01.2022 | 02.02.2021 | 09.02.2021 | 16.02.2021 | 23.02.2021 | 02.03.2021 |
| Fe ppm         | 490        | 373        | 864        | 597        | 343        | 364        | 590        | 405        | 824        |
| Si ppm         | 110        | 54         | 275        | 185        | 54         | 54         | 30         | 65         | 493        |
| Na ppm         | 88         | 17         | 285        | 132        | 20         | 20         | 149        | 19         | 197        |
| Week NO.       | 10         | 11         | 12         | 13         | 14         | 15         | 16         | 17         | 18         |
| Receiving Date | 09.03.2021 | 16.03.2021 | 23.03.2021 | 30.03.2021 | 06.04.2021 | 13.04.2021 | 20.04.2021 | 27.04.2021 | 04.05.2021 |
| Fe ppm         | 386        | 568        | 375        | 665        | 427        | 263        | 332        | 323        | 292        |
| Si ppm         | 106        | 131        | 98         | 933        | 65         | 70         | 84         | 62         | 59         |
| Na ppm         | 50         | 129        | 20         | 124        | 23         | 20         | 20         | 20         | 25         |
| Week NO.       | 19         | 20         | 21         | 22         | 23         | 24         | 25         | 26         | 27         |
| Receiving Date | 11.05.2021 | 18.05.2021 | 25.05.2021 | 01-Jun-21  | 08-Jun-21  | 15-Jun-21  | 22-Jun-21  | 29-Jun-21  | 06-Jul-21  |
| Fe ppm         | 374        | 319        | 329        | 348        | 293        | 345        | 324        | 350        | 335        |
| Si ppm         | 76         | 76         | 50         | 92         | 57         | 64         | 68         | 88         | 32         |
| Na ppm         | 40         | 20         | 305        | 55         | 20         | 23         | 25         | 18         | 247        |
| Week NO.       | 28         | 29         | 30         | 31         | 32         | 33         | 34         | 35         | 36         |
| Receiving Date | 13-Jul-21  | 20-Jul-21  | 27-Jul-21  | 03-Aug-21  | 10-Aug-21  | 17-Aug-21  | 24-Aug-21  | 31-Aug-21  | 07-Sep-21  |
| Fe ppm         | 686        | 345        | 354        | 328        | 399        | 409        | 247        | 424        | 426        |
| Si ppm         | 359        | 85         | 85         | 72         | 118        | 113        | 67         | 109        | 259        |
| Na ppm         | 452        | 20         | 25         | 30         | 20         | 37         | 20         | 54         | 98         |
| Week NO.       | 37         | 38         | 39         | 40         | 41         | 42         | 43         | 44         | 45         |
| Receiving Date | 14-Sep-21  | 21-Sep-21  | 28-Sep-21  | 05-Oct-21  | 12-Oct-21  | 19-Oct-21  | 26-Oct-21  | 02-Nov-21  | 09-Nov-21  |
| Fe ppm         | 1856       | 452        | 297        | 481        | 326        | 441        | 310        | 378        | 448        |
| Si ppm         | 277        | 69         | 74         | 119        | 87         | 96         | 81         | 85         | 75         |
| Na ppm         | 1348       | 469        | 29         | 105        | 44         | 35         | 45         | 27         | 19         |
| Week NO.       | 46         | 47         | 48         | 49         | 50         | 51         | 52         |            |            |
| Receiving Date | 16-Nov-21  | 23-Nov-21  | 30-Nov-21  | 07-Dec-21  | 14-Dec-21  | 21-Dec-21  | 28-Dec-21  |            |            |
| Fe ppm         | 448        | 483        | 365        | 338        | 376        | 367        | 368        |            |            |
| Si ppm         | 99         | 42         | 74         | 53         | 58         | 57         | 60         |            |            |
| Na ppm         | 48         | 202        | 17         | 20         | 13         | 10         | 20         |            |            |

## Section E: Coal tar pitch (CTP) properties data.

Note: The sample of data represents the year 2021, and data are presented weekly.

|                 |    |                   |                   |                   |                   |                   |                   |                   |                   |                   |
|-----------------|----|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Week No.        |    | <b>1</b>          | <b>2</b>          | <b>3</b>          | <b>4</b>          | <b>5</b>          | <b>6</b>          | <b>7</b>          | <b>8</b>          | <b>9</b>          |
| Date Received   |    | <b>05.01.2021</b> | <b>12.01.2021</b> | <b>19.01.2021</b> | <b>26.01.2021</b> | <b>02.02.2021</b> | <b>09.02.2021</b> | <b>16.02.2021</b> | <b>23.02.2021</b> | <b>02.03.2021</b> |
| Softening Point | °C | 112.4             | 113.9             | 112.7             | 112.1             | 112.9             | 113.5             | 113.3             | 113.1             | 112.4             |
| Coking Value    | %  | 57.2              | 56.8              | 56.7              | 57.6              | 57.1              | 56.9              | 56.6              | 57.2              | 57.0              |
| Sulphur         | %  | 0.64              | 0.62              | 0.61              | 0.59              | 0.7               | 0.75              | 0.701             | 0.584             | 0.625             |
| Week No.        |    | <b>10</b>         | <b>11</b>         | <b>12</b>         | <b>13</b>         | <b>14</b>         | <b>15</b>         | <b>16</b>         | <b>17</b>         | <b>18</b>         |
| Date Received   |    | <b>09.03.2021</b> | <b>16.03.2021</b> | <b>23.03.2021</b> | <b>30.03.2021</b> | <b>06.04.2021</b> | <b>13.04.2021</b> | <b>20.04.2021</b> | <b>27.04.2021</b> | <b>04.05.2021</b> |
| Softening Point | °C | 112.7             | 113.5             | 113.0             | 113.2             | 112.3             | 113.4             | 112.7             | 112.2             | 112.9             |
| Coking Value    | %  | 56.8              | 57.4              | 57.1              | 57.3              | 56.8              | 56.9              | 57.1              | 56.8              | 57.4              |
| Sulphur         | %  | 0.65              | 0.64              | 0.69              | 0.66              | 0.63              | 0.68              | 0.634             | 0.601             | 0.6               |
| Week No.        |    | <b>19</b>         | <b>20</b>         | <b>21</b>         | <b>22</b>         | <b>23</b>         | <b>24</b>         | <b>25</b>         | <b>26</b>         | <b>27</b>         |
| Date Received   |    | <b>11.05.2021</b> | <b>18.05.2021</b> | <b>25.05.2021</b> | <b>01.06.2021</b> | <b>08.06.2021</b> | <b>15.06.2021</b> | <b>22.06.2021</b> | <b>29.06.2021</b> | <b>06.07.2021</b> |
| Softening Point | °C | 113.3             | 112.5             | 112.1             | 111.9             | 111.3             | 111.6             | 112.5             | 112.9             | 112.2             |
| Coking Value    | %  | 56.6              | 57.2              | 56.9              | 57.0              | 56.2              | 57.0              | 56.7              | 57.3              | 57.1              |
| Sulphur         | %  | 0.661             | 0.56              | 0.6               | 0.59              | 0.574             | 0.61              | 0.6               | 0.59              | 0.62              |
| Week No.        |    | <b>28</b>         | <b>29</b>         | <b>30</b>         | <b>31</b>         | <b>32</b>         | <b>33</b>         | <b>34</b>         | <b>35</b>         | <b>36</b>         |
| Date Received   |    | <b>13.07.2021</b> | <b>20.07.2021</b> | <b>26.07.2021</b> | <b>03.08.2021</b> | <b>10.08.2021</b> | <b>17.08.2021</b> | <b>24.08.2021</b> | <b>31.08.2021</b> | <b>07.09.2021</b> |
| Softening Point | °C | 111.9             | 112.4             | 113.1             | 113.4             | 113.1             | 113.3             | 112.6             | 113.1             | 112.3             |
| Coking Value    | %  | 56.9              | 57.0              | 56.5              | 56.2              | 56.7              | 56.9              | 57.1              | 56.9              | 56.7              |
| Sulphur         | %  | 0.66              | 0.64              | 0.67              | 0.62              | 0.58              | 0.571             | 0.534             | 0.614             | 0.62              |
| Week No.        |    | <b>37</b>         | <b>38</b>         | <b>39</b>         | <b>40</b>         | <b>41</b>         | <b>42</b>         | <b>43</b>         | <b>44</b>         | <b>45</b>         |
| Date Received   |    | <b>14.09.2021</b> | <b>21.09.2021</b> | <b>28.09.2021</b> | <b>05.10.2021</b> | <b>12.10.2021</b> | <b>19.10.2021</b> | <b>26.10.2021</b> | <b>02.11.2021</b> | <b>09.11.2021</b> |
| Softening Point | °C | 113.4             | 112.6             | 112.3             | 112.8             | 111.6             | 112.9             | 112.1             | 113.2             | 112.4             |
| Coking Value    | %  | 56.4              | 57.1              | 56.7              | 57.0              | 55.9              | 56.1              | 56.6              | 57.0              | 56.4              |
| Sulphur         | %  | 0.601             | 0.613             | 0.62              | 0.54              | 0.63              | 0.66              | 0.66              | 0.66              | 0.67              |
| Week No.        |    | <b>46</b>         | <b>47</b>         | <b>48</b>         | <b>49</b>         | <b>50</b>         | <b>51</b>         | <b>52</b>         |                   |                   |
| Date Received   |    | <b>16.11.2021</b> | <b>23.11.2021</b> | <b>30.11.2021</b> | <b>07.12.2021</b> | <b>14.12.2021</b> | <b>21.12.2021</b> | <b>28.12.2021</b> |                   |                   |
| Softening Point | °C | 112.9             | 113.2             | 113.5             | 112.9             | 111.9             | 112.4             | 112.9             |                   |                   |
| Coking Value    | %  | 56.2              | 56.8              | 57.0              | 56.6              | 56.2              | 56.9              | 57.1              |                   |                   |
| Sulphur         | %  | 0.608             | 0.532             | 0.62              | 0.61              | 0.56              | 0.505             | 0.509             |                   |                   |

## Section F: Paste plant process parameters data.

Note: The sample of data represents the first few green anodes produced on May 31, 2021. Data are presented continuously for every anode produced, along with its process parameters.

| Date and Time      | Anode No | Green anode density (GAD) | Pitch level | Mixing Power | Cooling Power | Mixing temperature | Dry aggregate temperature | Anode forming Temperature | Green scrap content | Butts (very coarse) percentage | Fine particles percentage | Medium particles percentage | Coarse particles percentage | K1.040 Speed | K1.040 Time | K1.050 Speed | K1.050 Time |
|--------------------|----------|---------------------------|-------------|--------------|---------------|--------------------|---------------------------|---------------------------|---------------------|--------------------------------|---------------------------|-----------------------------|-----------------------------|--------------|-------------|--------------|-------------|
| 31-Mar-21 00:01:18 | 209      | 1.623                     | 13.20       | 312.000      | 246.190       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:02:58 | 206      | 1.639                     | 13.20       | 305.000      | 247.143       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:04:48 | 210      | 1.624                     | 13.20       | 309.000      | 247.636       | 203                | 183                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:06:28 | 207      | 1.634                     | 13.20       | 303.000      | 245.818       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:08:18 | 211      | 1.627                     | 13.20       | 300.000      | 244.083       | 203                | 183                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:09:58 | 208      | 1.637                     | 13.20       | 299.000      | 244.917       | 203                | 183                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:11:58 | 212      | 1.625                     | 13.20       | 298.000      | 245.700       | 203                | 182                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:13:48 | 209      | 1.634                     | 13.20       | 307.000      | 242.667       | 203                | 182                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:15:38 | 213      | 1.628                     | 13.20       | 300.000      | 241.100       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:17:18 | 210      | 1.637                     | 13.20       | 307.000      | 241.818       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:19:08 | 214      | 1.623                     | 13.20       | 303.000      | 241.000       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:20:48 | 211      | 1.634                     | 13.20       | 312.000      | 241.500       | 203                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:22:38 | 215      | 1.628                     | 13.20       | 302.000      | 241.727       | 203                | 181                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:24:28 | 212      | 1.634                     | 13.20       | 309.000      | 239.182       | 203                | 181                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:26:18 | 216      | 1.620                     | 13.20       | 308.000      | 240.818       | 203                | 182                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:28:08 | 213      | 1.635                     | 13.20       | 302.000      | 240.000       | 203                | 182                       | 147                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:29:48 | 217      | 1.622                     | 13.20       | 309.000      | 241.417       | 202                | 182                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:31:48 | 214      | 1.634                     | 13.20       | 306.000      | 246.200       | 202                | 181                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:33:28 | 218      | 1.622                     | 13.20       | 305.000      | 247.667       | 202                | 181                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:35:28 | 215      | 1.625                     | 13.20       | 291.000      | 244.399       | 202                | 181                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |
| 31-Mar-21 00:37:09 | 219      | 1.629                     | 13.20       | 309.000      | 244.000       | 202                | 181                       | 146                       | 2.00                | 24.00                          | 20.00                     | 28.00                       | 28.00                       | 1101         | 65          | 1101         | 68          |

## Section G: Baking kiln process parameters data.

Note: The sample of data represents the year 2021, and data are presented weekly.

| Parameters              | Unit  | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      |
|-------------------------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Fire Cycle time         | hr    | 24     | 24     | 24     | 24     | 24     | 26     | 26     | 26     | 26     |
| Fire Soaking time       | hr    | 41     | 45     | 45     | 45     | 46     | 50     | 50     | 50     | 50     |
| Final Anode Temperature | °C    | 1106.1 | 1105.3 | 1103.8 | 1107.3 | 1118.6 | 1124.6 | 1127.1 | 1130.6 | 1118.0 |
| Heating Rate            | °C/hr | 19.8   | 19.5   | 19.6   | 19.7   | 17.9   | 17.7   | 17.7   | 17.9   | 18.0   |
| Parameters              | Unit  | 10     | 11     | 12     | 13     | 14     | 15     | 16     | 17     | 18     |
| Fire Cycle time         | hr    | 26     | 26     | 26     | 26     | 26     | 26     | 26     | 26     | 25     |
| Fire Soaking time       | hr    | 50     | 50     | 50     | 50     | 50     | 50     | 50     | 50     | 50     |
| Final Anode Temperature | °C    | 1125.4 | 1120.9 | 1119.5 | 1122.8 | 1119.2 | 1122.8 | 1123.3 | 1117.0 | 1123.5 |
| Heating Rate            | °C/hr | 17.8   | 18.1   | 18.1   | 18.0   | 17.7   | 17.9   | 18.0   | 18.9   | 19.2   |
| Parameters              | Unit  | 19     | 20     | 21     | 22     | 23     | 24     | 25     | 26     | 27     |
| Fire Cycle time         | hr    | 25     | 25     | 25     | 25     | 25     | 25     | 25     | 25     | 25     |
| Fire Soaking time       | hr    | 50     | 50     | 50     | 50     | 50     | 50     | 50     | 50     | 50     |
| Final Anode Temperature | °C    | 1127.3 | 1124.6 | 1123.0 | 1121.4 | 1128.1 | 1118.8 | 1111.5 | 1120.9 | 1120.5 |
| Heating Rate            | °C/hr | 18.9   | 19.0   | 18.8   | 18.7   | 18.7   | 18.5   | 18.6   | 18.9   | 18.8   |
| Parameters              | Unit  | 28     | 29     | 30     | 31     | 32     | 33     | 34     | 35     | 36     |
| Fire Cycle time         | hr    | 25     | 25     | 25     | 25     | 25     | 25     | 25     | 25     | 24     |
| Fire Soaking time       | hr    | 50     | 50     | 50     | 50     | 50     | 50     | 50     | 50     | 45     |
| Final Anode Temperature | °C    | 1125.5 | 1128.5 | 1128.1 | 1126.1 | 1126.9 | 1125.5 | 1125.8 | 1129.9 | 1117.5 |
| Heating Rate            | °C/hr | 18.8   | 18.7   | 18.8   | 18.9   | 18.4   | 18.7   | 18.7   | 18.8   | 19.3   |
| Parameters              | Unit  | 37     | 38     | 39     | 40     | 41     | 42     | 43     | 44     | 45     |
| Fire Cycle time         | hr    | 24     | 24     | 24     | 24     | 24     | 24     | 24     | 24     | 24     |
| Fire Soaking time       | hr    | 45     | 45     | 45     | 45     | 45     | 45     | 45     | 45     | 45     |
| Final Anode Temperature | °C    | 1112.8 | 1113.4 | 1115.0 | 1114.1 | 1117.9 | 1118.7 | 1118.7 | 1120.1 | 1115.1 |
| Heating Rate            | °C/hr | 19.4   | 19.5   | 19.6   | 18.6   | 19.0   | 19.0   | 19.0   | 19.3   | 19.1   |
| Parameters              | Unit  | 46     | 47     | 48     | 49     | 50     | 51     | 52     |        |        |
| Fire Cycle time         | hr    | 24     | 24     | 24     | 24     | 24     | 24     | 24     |        |        |
| Fire Soaking time       | hr    | 45     | 45     | 45     | 45     | 45     | 45     | 45     |        |        |
| Final Anode Temperature | °C    | 1120.4 | 1113.2 | 1119.1 | 1115.5 | 1117.5 | 1120.2 | 1117.9 |        |        |
| Heating Rate            | °C/hr | 19.3   | 19.4   | 19.3   | 19.4   | 19.2   | 19.1   | 19.2   |        |        |

## **Appendix 2:**

### **Anode quality properties after preparation.**

**This Appendix contains the following two sections:**

**Section A: Anode quality properties after preparation of Model 1.**

**Section B: Anode quality properties after preparation of Model 2.**

## Section A: Anode quality properties after preparation of Model 1.

Note: The sample of data represents a few observations of November 2020.

| Date       | Sample ID         | Anode Code | Anode Quality Properties |                                |                    |
|------------|-------------------|------------|--------------------------|--------------------------------|--------------------|
|            |                   |            | Geometric Density        | Specific Electrical Resistance | Flextural Strength |
| 01/11/2020 | S-20201101-000669 | 220401242  | 1.594                    | 5303                           | 14.07              |
| 01/11/2020 | S-20201101-000668 | 120401270  | 1.593                    | 5027                           | 12.82              |
| 01/11/2020 | S-20201101-000670 | 220410129  | 1.581                    | 4933                           | 11.48              |
| 01/11/2020 | S-20201101-000671 | 220410202  | 1.574                    | 5150                           | 14.34              |
| 02/11/2020 | S-20201102-000693 | 120401433  | 1.591                    | 4933                           | 11.75              |
| 03/11/2020 | S-20201103-000664 | 120400846  | 1.579                    | 5033                           | 12.29              |
| 03/11/2020 | S-20201103-000665 | 120400854  | 1.583                    | 5167                           | 9.52               |
| 03/11/2020 | S-20201103-000663 | 220401483  | 1.571                    | 4940                           | 11.59              |
| 04/11/2020 | S-20201104-000629 | 220400576  | 1.583                    | 5133                           | 13.77              |
| 04/11/2020 | S-20201104-000628 | 120400701  | 1.594                    | 4943                           | 14.33              |
| 04/11/2020 | S-20201104-000630 | 220410320  | 1.586                    | 5570                           | 11.27              |
| 04/11/2020 | S-20201104-000632 | 220420077  | 1.575                    | 5297                           | 12.11              |
| 04/11/2020 | S-20201104-000631 | 220420099  | 1.572                    | 6673                           | 2.21               |
| 05/11/2020 | S-20201105-000294 | 220400557  | 1.573                    | 4840                           | 13.41              |

## Section B: Anode quality properties after preparation of Model 2.

Note: The sample of data represents a few observations of May 2021.

| Date       | Sample ID         | Anode Code | Anode Quality Properties |                            |                           |
|------------|-------------------|------------|--------------------------|----------------------------|---------------------------|
|            |                   |            | CO2 Reactivity Residue   | Sulfur (S) Element Content | Iron (Fe) Element Content |
| 01/05/2021 | S-20210501-000434 | 221140554  | 93.4                     | 1.402                      | 334                       |
| 01/05/2021 | S-20210501-000440 | 121140389  | 93.3                     | 1.372                      | 275                       |
| 02/05/2021 | S-20210502-000790 | 121160290  | 95.7                     | 1.566                      | 312                       |
| 04/05/2021 | S-20210504-000802 | 121140395  | 93.3                     | 1.390                      | 295                       |
| 05/05/2021 | S-20210505-000686 | 221151208  | 93.6                     | 1.636                      | 362                       |
| 05/05/2021 | S-20210505-000690 | 221160157  | 94.1                     | 1.643                      | 364                       |
| 06/05/2021 | S-20210506-000363 | 221150817  | 91.3                     | 1.451                      | 251                       |
| 08/05/2021 | S-20210508-000439 | 121150160  | 93.9                     | 1.270                      | 305                       |
| 08/05/2021 | S-20210508-000443 | 221150400  | 93.6                     | 1.380                      | 309                       |
| 08/05/2021 | S-20210508-000445 | 121160896  | 94.0                     | 1.329                      | 300                       |
| 09/05/2021 | S-20210509-001338 | 121170117  | 91.5                     | 1.381                      | 293                       |
| 10/05/2021 | S-20210510-000753 | 121160946  | 93.7                     | 1.407                      | 394                       |
| 11/05/2021 | S-20210511-000726 | 121161136  | 93.3                     | 1.436                      | 306                       |
| 15/05/2021 | S-20210515-000418 | 121170005  | 93.6                     | 1.535                      | 321                       |

# **Appendix 3:**

## **Data collected preparation after adding paste plant process parameters.**

**This Appendix contains the following two sections:**

**Section A: Data collected preparation, including paste plant process  
parameters of Model 1.**

**Section B: Data collected preparation, including paste plant process  
parameters of Model 2.**

## Section A: Data collected preparation, including paste plant process parameters of Model 1.

Note: The sample of data represents a few observations of November 2020.

| Date       | Sample ID         | Anode Code | Anode Quality Properties |                                |                    | Paste Plant Parameters    |             |              |               |                    |                           |                           |                     |                                |                           |                             |                             |                     |                    |
|------------|-------------------|------------|--------------------------|--------------------------------|--------------------|---------------------------|-------------|--------------|---------------|--------------------|---------------------------|---------------------------|---------------------|--------------------------------|---------------------------|-----------------------------|-----------------------------|---------------------|--------------------|
|            |                   |            | Geometric Density        | Specific Electrical Resistance | Flextural Strength | Green anode density (GAD) | Pitch level | Mixing Power | Cooling Power | Mixing temperature | Dry aggregate temperature | Anode forming Temperature | Green scrap content | Butts (very coarse) percentage | Fine particles percentage | Medium particles percentage | Coarse particles percentage | Anode forming speed | Anode forming time |
| 01/11/2020 | S-20201101-000669 | 220401242  | 1.594                    | 5303                           | 14.07              | 1.628                     | 13.80       | 279          | 252           | 205                | 191                       | 145                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1122                | 65                 |
| 01/11/2020 | S-20201101-000668 | 120401270  | 1.593                    | 5027                           | 12.82              | 1.626                     | 13.80       | 277          | 260           | 204                | 190                       | 144                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1120                | 65                 |
| 01/11/2020 | S-20201101-000670 | 220410129  | 1.581                    | 4933                           | 11.48              | 1.623                     | 13.80       | 280          | 253           | 201                | 191                       | 145                       | 3                   | 28                             | 21                        | 27                          | 24                          | 1122                | 67                 |
| 01/11/2020 | S-20201101-000671 | 220410202  | 1.574                    | 5150                           | 14.34              | 1.633                     | 13.80       | 279          | 255           | 200                | 188                       | 127                       | 3                   | 28                             | 21                        | 27                          | 24                          | 1122                | 67                 |
| 02/11/2020 | S-20201102-000693 | 120401433  | 1.591                    | 4933                           | 11.75              | 1.637                     | 13.80       | 278          | 262           | 204                | 193                       | 145                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1120                | 65                 |
| 03/11/2020 | S-20201103-000664 | 120400846  | 1.579                    | 5033                           | 12.29              | 1.627                     | 13.80       | 280          | 254           | 205                | 191                       | 144                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1120                | 65                 |
| 03/11/2020 | S-20201103-000665 | 120400854  | 1.583                    | 5167                           | 9.52               | 1.627                     | 13.80       | 278          | 257           | 205                | 189                       | 144                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1120                | 65                 |
| 03/11/2020 | S-20201103-000663 | 220401483  | 1.571                    | 4940                           | 11.59              | 1.639                     | 13.80       | 280          | 272           | 203                | 189                       | 144                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1122                | 67                 |
| 04/11/2020 | S-20201104-000629 | 220400576  | 1.583                    | 5133                           | 13.77              | 1.621                     | 13.50       | 281          | 255           | 206                | 192                       | 145                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1122                | 65                 |
| 04/11/2020 | S-20201104-000628 | 120400701  | 1.594                    | 4943                           | 14.33              | 1.624                     | 13.70       | 281          | 259           | 204                | 191                       | 144                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1120                | 65                 |
| 04/11/2020 | S-20201104-000630 | 220410320  | 1.586                    | 5570                           | 11.27              | 1.628                     | 13.80       | 280          | 256           | 201                | 191                       | 123                       | 3                   | 28                             | 21                        | 27                          | 24                          | 1122                | 67                 |
| 04/11/2020 | S-20201104-000632 | 220420077  | 1.575                    | 5297                           | 12.11              | 1.630                     | 13.80       | 286          | 258           | 205                | 191                       | 148                       | 3                   | 28                             | 21                        | 27                          | 24                          | 1122                | 67                 |
| 04/11/2020 | S-20201104-000631 | 220420099  | 1.572                    | 6673                           | 2.21               | 1.620                     | 13.80       | 282          | 263           | 205                | 191                       | 147                       | 3                   | 28                             | 21                        | 27                          | 24                          | 1127                | 67                 |
| 05/11/2020 | S-20201105-000294 | 220400557  | 1.573                    | 4840                           | 13.41              | 1.619                     | 13.60       | 278          | 256           | 205                | 192                       | 145                       | 2                   | 28                             | 21                        | 27                          | 24                          | 1122                | 65                 |

## Section B: Data collected preparation, including paste plant process parameters of Model 2.

Note: The sample of data represents a few observations of May 2021.

| Date       | Sample ID         | Anode Code | Anode Quality Properties |                            |                           | Paste Plant Parameters    |             |              |               |                    |                           |                           |                     |                                |                           |                             |                             |                     |                    |  |
|------------|-------------------|------------|--------------------------|----------------------------|---------------------------|---------------------------|-------------|--------------|---------------|--------------------|---------------------------|---------------------------|---------------------|--------------------------------|---------------------------|-----------------------------|-----------------------------|---------------------|--------------------|--|
|            |                   |            | CO2 Reactivity Residue   | Sulfur (S) Element Content | Iron (Fe) Element Content | Green anode density (GAD) | Pitch level | Mixing Power | Cooling Power | Mixing temperature | Dry aggregate temperature | Anode forming Temperature | Green scrap content | Butts (very coarse) percentage | Fine particles percentage | Medium particles percentage | Coarse particles percentage | Anode forming speed | Anode forming time |  |
|            |                   |            |                          |                            |                           |                           |             |              |               |                    |                           |                           |                     |                                |                           |                             |                             |                     |                    |  |
|            |                   |            |                          |                            |                           |                           |             |              |               |                    |                           |                           |                     |                                |                           |                             |                             |                     |                    |  |
| 01/05/2021 | S-20210501-000434 | 221140554  | 93.4                     | 1.402                      | 334                       | 1.620                     | 13.20       | 305          | 255           | 209                | 186                       | 148                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 68                 |  |
| 01/05/2021 | S-20210501-000440 | 121140389  | 93.3                     | 1.372                      | 275                       | 1.631                     | 13.20       | 303          | 246           | 204                | 185                       | 148                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 02/05/2021 | S-20210502-000790 | 121160290  | 95.7                     | 1.566                      | 312                       | 1.632                     | 13.30       | 304          | 253           | 204                | 180                       | 148                       | 1                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 04/05/2021 | S-20210504-000802 | 121140395  | 93.3                     | 1.390                      | 295                       | 1.634                     | 13.20       | 304          | 252           | 207                | 186                       | 148                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 05/05/2021 | S-20210505-000686 | 221151208  | 93.6                     | 1.636                      | 362                       | 1.627                     | 13.30       | 300          | 260           | 205                | 182                       | 147                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 68                 |  |
| 05/05/2021 | S-20210505-000690 | 221160157  | 94.1                     | 1.643                      | 364                       | 1.631                     | 13.30       | 313          | 234           | 203                | 177                       | 143                       | 1                   | 24                             | 20                        | 28                          | 28                          | 1101                | 68                 |  |
| 06/05/2021 | S-20210506-000363 | 221150817  | 91.3                     | 1.451                      | 251                       | 1.629                     | 13.30       | 313          | 252           | 207                | 183                       | 146                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 68                 |  |
| 08/05/2021 | S-20210508-000439 | 121150160  | 93.9                     | 1.270                      | 305                       | 1.636                     | 13.30       | 305          | 252           | 208                | 185                       | 146                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 08/05/2021 | S-20210508-000443 | 221150400  | 93.6                     | 1.380                      | 309                       | 1.630                     | 13.30       | 306          | 258           | 207                | 186                       | 146                       | 2                   | 24                             | 20                        | 28                          | 28                          | 1101                | 68                 |  |
| 08/05/2021 | S-20210508-000445 | 121160896  | 94.0                     | 1.329                      | 300                       | 1.640                     | 13.30       | 300          | 268           | 202                | 178                       | 148                       | 3                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 09/05/2021 | S-20210509-001338 | 121170117  | 91.5                     | 1.381                      | 293                       | 1.641                     | 13.30       | 305          | 261           | 199                | 177                       | 147                       | 3                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 10/05/2021 | S-20210510-000753 | 121160946  | 93.7                     | 1.407                      | 394                       | 1.646                     | 13.20       | 303          | 257           | 199                | 174                       | 147                       | 3                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 11/05/2021 | S-20210511-000726 | 121161136  | 93.3                     | 1.436                      | 306                       | 1.633                     | 13.30       | 306          | 260           | 200                | 176                       | 147                       | 3                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |
| 15/05/2021 | S-20210515-000418 | 121170005  | 93.6                     | 1.535                      | 321                       | 1.648                     | 13.30       | 301          | 256           | 203                | 178                       | 147                       | 3                   | 24                             | 20                        | 28                          | 28                          | 1101                | 65                 |  |

## **Appendix 4:**

### **Final data collected ready to use for building model.**

**This Appendix contains the following two sections:**

**Section A: Final data collected from Model 1.**

**Section B: Final data collected from Model 2.**

## Section A: Final data collected from Model 1.

Note: The sample of data represents the first 30 observations of the total dataset.

| Observation | y1    | y2   | y3   | x1    | x2    | x3  | x4  | x5  | x6  | x7  | x8 | x9 | x10 | x11 | x12 | x13  | x14 | x15 | x16 | x17  | x18 | x19 | x20   | x21  | x22  | x23  | x24 | x25 | x26 | x27  | x28  | x29 | x30 | x31    | x32  |
|-------------|-------|------|------|-------|-------|-----|-----|-----|-----|-----|----|----|-----|-----|-----|------|-----|-----|-----|------|-----|-----|-------|------|------|------|-----|-----|-----|------|------|-----|-----|--------|------|
| 1           | 1.576 | 5227 | 11   | 1.629 | 13.8  | 278 | 262 | 199 | 188 | 147 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 331 | 20  | 1025 | 59  | 147 | 112.6 | 57.4 | 0.61 | 2.21 | 237 | 144 | 4.1 | 28.0 | 26.3 | 26  | 41  | 1119.7 | 16.2 |
| 2           | 1.589 | 5137 | 11.3 | 1.627 | 13.80 | 295 | 256 | 200 | 183 | 146 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 1.87 | 214 | 99  | 4.7 | 27.4 | 26.9 | 24  | 41  | 1106.3 | 19.7 |
| 3           | 1.576 | 4870 | 14.5 | 1.630 | 13.80 | 271 | 255 | 201 | 184 | 147 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.08 | 218 | 125 | 4.1 | 27.4 | 27.7 | 24  | 41  | 1106.3 | 19.7 |
| 4           | 1.594 | 5037 | 13.5 | 1.623 | 13.80 | 286 | 253 | 201 | 184 | 148 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.08 | 218 | 125 | 4.1 | 27.4 | 27.7 | 24  | 41  | 1106.3 | 19.7 |
| 5           | 1.584 | 5327 | 9.47 | 1.625 | 13.80 | 282 | 261 | 201 | 185 | 148 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.08 | 218 | 125 | 4.1 | 27.4 | 27.7 | 24  | 41  | 1106.3 | 19.7 |
| 6           | 1.586 | 4953 | 13.8 | 1.624 | 13.80 | 307 | 261 | 201 | 183 | 148 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.08 | 218 | 125 | 4.1 | 27.4 | 27.7 | 24  | 41  | 1100.7 | 19.7 |
| 7           | 1.589 | 5383 | 9.86 | 1.619 | 13.80 | 308 | 260 | 201 | 183 | 148 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.08 | 218 | 125 | 4.1 | 27.4 | 27.7 | 24  | 41  | 1100.7 | 19.7 |
| 8           | 1.599 | 4973 | 11.3 | 1.628 | 13.80 | 305 | 259 | 201 | 182 | 148 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.08 | 218 | 125 | 4.1 | 27.4 | 27.7 | 24  | 41  | 1100.7 | 19.6 |
| 9           | 1.595 | 4987 | 12.6 | 1.612 | 13.80 | 303 | 262 | 201 | 184 | 149 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.13 | 234 | 128 | 4.3 | 32.3 | 28.2 | 24  | 41  | 1106.3 | 19.7 |
| 10          | 1.594 | 5150 | 13   | 1.634 | 13.80 | 305 | 256 | 202 | 184 | 147 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.13 | 234 | 128 | 4.3 | 32.3 | 28.2 | 24  | 41  | 1105.2 | 19.6 |
| 11          | 1.585 | 5337 | 12.4 | 1.615 | 13.80 | 302 | 259 | 201 | 183 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.04 | 236 | 145 | 4.5 | 28.2 | 28.1 | 24  | 41  | 1100.7 | 19.6 |
| 12          | 1.592 | 5217 | 12.1 | 1.620 | 13.80 | 306 | 258 | 200 | 181 | 147 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.04 | 236 | 145 | 4.5 | 28.2 | 28.1 | 24  | 41  | 1100.7 | 19.6 |
| 13          | 1.601 | 5263 | 11.3 | 1.634 | 13.80 | 307 | 262 | 202 | 182 | 148 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.04 | 236 | 145 | 4.5 | 28.2 | 28.1 | 24  | 41  | 1105.2 | 19.6 |
| 14          | 1.595 | 5220 | 11.3 | 1.636 | 13.80 | 314 | 262 | 199 | 181 | 149 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.04 | 236 | 145 | 4.5 | 28.2 | 28.1 | 24  | 41  | 1105.2 | 19.7 |
| 15          | 1.592 | 5260 | 10.9 | 1.620 | 13.80 | 308 | 256 | 200 | 180 | 148 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.04 | 236 | 145 | 4.5 | 28.2 | 28.1 | 24  | 41  | 1105.2 | 19.7 |
| 16          | 1.592 | 5123 | 12.5 | 1.627 | 13.80 | 309 | 258 | 199 | 180 | 148 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.04 | 236 | 145 | 4.5 | 28.2 | 28.1 | 24  | 41  | 1105.2 | 19.7 |
| 17          | 1.585 | 5050 | 10.2 | 1.627 | 13.80 | 309 | 267 | 199 | 179 | 147 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.14 | 233 | 118 | 4.6 | 27.5 | 28.0 | 24  | 41  | 1105.2 | 19.7 |
| 18          | 1.595 | 5123 | 12.9 | 1.630 | 13.80 | 306 | 267 | 199 | 179 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.14 | 233 | 118 | 4.6 | 27.5 | 28.0 | 24  | 41  | 1105.2 | 19.7 |
| 19          | 1.596 | 5230 | 12.3 | 1.618 | 13.80 | 304 | 254 | 199 | 180 | 146 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.14 | 233 | 118 | 4.6 | 27.5 | 28.0 | 24  | 41  | 1105.2 | 19.7 |
| 20          | 1.594 | 5073 | 13.2 | 1.631 | 13.80 | 306 | 258 | 200 | 180 | 147 | 3  | 28 | 20  | 27  | 25  | 1120 | 65  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.14 | 233 | 118 | 4.6 | 27.5 | 28.0 | 24  | 41  | 1105.2 | 19.7 |
| 21          | 1.597 | 5087 | 11.8 | 1.628 | 13.80 | 301 | 251 | 200 | 182 | 147 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 372 | 347 | 634  | 537 | 377 | 112.3 | 56.5 | 0.63 | 2.14 | 233 | 118 | 4.6 | 27.5 | 28.0 | 24  | 41  | 1105.2 | 19.7 |
| 22          | 1.598 | 5170 | 11.4 | 1.630 | 13.80 | 306 | 261 | 201 | 175 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 527 | 188 | 389  | 53  | 35  | 112.2 | 57.0 | 0.65 | 1.99 | 191 | 115 | 4.9 | 31.6 | 26.7 | 24  | 41  | 1101.3 | 19.8 |
| 23          | 1.592 | 5200 | 12   | 1.628 | 13.80 | 307 | 260 | 202 | 175 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 527 | 188 | 389  | 53  | 35  | 112.2 | 57.0 | 0.65 | 1.99 | 191 | 115 | 4.9 | 31.6 | 26.7 | 24  | 41  | 1101.3 | 19.8 |
| 24          | 1.599 | 5133 | 12   | 1.632 | 13.80 | 309 | 253 | 201 | 175 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 527 | 188 | 389  | 53  | 35  | 112.2 | 57.0 | 0.65 | 1.99 | 191 | 115 | 4.9 | 31.6 | 26.7 | 24  | 41  | 1101.3 | 19.8 |
| 25          | 1.601 | 5200 | 13.1 | 1.640 | 13.80 | 306 | 262 | 201 | 174 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 527 | 188 | 389  | 53  | 35  | 112.2 | 57.0 | 0.65 | 1.99 | 191 | 115 | 4.9 | 31.6 | 26.7 | 24  | 41  | 1101.3 | 19.8 |
| 26          | 1.595 | 4930 | 11.8 | 1.642 | 13.80 | 303 | 263 | 202 | 176 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 527 | 188 | 389  | 53  | 35  | 112.2 | 57.0 | 0.65 | 1.99 | 191 | 115 | 4.9 | 31.6 | 26.7 | 24  | 41  | 1101.3 | 19.8 |
| 27          | 1.594 | 5337 | 11.4 | 1.639 | 13.80 | 307 | 259 | 202 | 176 | 145 | 3  | 28 | 20  | 27  | 25  | 1122 | 68  | 527 | 188 | 389  | 53  | 35  | 112.2 | 57.0 | 0.65 | 1.99 | 191 | 115 | 4.9 | 31.6 | 26.7 | 24  | 41  | 1101.3 | 19.8 |
| 28          | 1.579 | 4963 | 12.3 | 1.632 | 13.80 | 302 | 261 | 202 | 177 | 144 | 3  | 25 | 20  | 27  | 28  | 1122 | 68  | 558 | 380 | 372  | 60  | 44  | 113.6 | 57.4 | 0.68 | 2.03 | 199 | 130 | 4.7 | 26.5 | 26.9 | 24  | 41  | 1106.1 | 19.8 |
| 29          | 1.577 | 5330 | 11.1 | 1.628 | 13.80 | 307 | 264 | 202 | 177 | 145 | 3  | 25 | 20  | 29  | 26  | 1122 | 68  | 558 | 380 | 372  | 60  | 44  | 113.6 | 57.4 | 0.68 | 2.03 | 199 | 130 | 4.7 | 26.5 | 26.9 | 24  | 41  | 1106.1 | 19.8 |
| 30          | 1.581 | 5300 | 11.1 | 1.625 | 13.80 | 306 | 259 | 202 | 177 | 145 | 3  | 25 | 20  | 29  | 26  | 1122 | 68  | 558 | 380 | 372  | 60  | 44  | 113.6 | 57.4 | 0.68 | 1.99 | 187 | 140 | 4.6 | 28.0 | 27.6 | 24  | 41  | 1106.1 | 19.8 |

## Section B: Final data collected from Model 2.

Note: The sample of data represents the first 30 observations of the total dataset.

| Observation | y4   | y5    | y6  | x1    | x2    | x3  | x4  | x5  | x6  | x7  | x8 | x9 | x10 | x11 | x12 | x13  | x14 | x15 | x16 | x17 | x18 | x19 | x20   | x21  | x22  | x23  | x24 | x25 | x26 | x27  | x28  | x29 | x30 | x31    | x32  |
|-------------|------|-------|-----|-------|-------|-----|-----|-----|-----|-----|----|----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-------|------|------|------|-----|-----|-----|------|------|-----|-----|--------|------|
| 1           | 93.4 | 1.402 | 334 | 1.620 | 13.2  | 305 | 255 | 209 | 186 | 148 | 2  | 24 | 20  | 28  | 28  | 1101 | 68  | 480 | 152 | 427 | 65  | 23  | 112.3 | 56.8 | 0.63 | 1.11 | 256 | 118 | 3.5 | 30.8 | 24.5 | 26  | 50  | 1117.0 | 18.9 |
| 2           | 93.3 | 1.372 | 275 | 1.631 | 13.20 | 303 | 246 | 204 | 185 | 148 | 2  | 24 | 20  | 28  | 28  | 1101 | 65  | 480 | 152 | 427 | 65  | 23  | 112.3 | 56.8 | 0.63 | 1.11 | 256 | 118 | 3.5 | 30.8 | 24.5 | 26  | 50  | 1117.0 | 18.9 |
| 3           | 95.7 | 1.566 | 312 | 1.632 | 13.30 | 304 | 253 | 204 | 180 | 148 | 1  | 24 | 20  | 28  | 28  | 1101 | 65  | 443 | 362 | 332 | 84  | 20  | 112.7 | 57.1 | 0.63 | 1.79 | 165 | 47  | 2.2 | 24.3 | 28.3 | 26  | 50  | 1117.0 | 18.9 |
| 4           | 93.3 | 1.390 | 295 | 1.634 | 13.20 | 304 | 252 | 207 | 186 | 148 | 2  | 24 | 20  | 28  | 28  | 1101 | 65  | 480 | 152 | 427 | 65  | 23  | 112.3 | 56.8 | 0.63 | 1.11 | 256 | 118 | 3.5 | 30.8 | 24.5 | 25  | 50  | 1123.5 | 18.9 |
| 5           | 93.6 | 1.636 | 362 | 1.627 | 13.30 | 300 | 260 | 205 | 182 | 147 | 2  | 24 | 20  | 28  | 28  | 1101 | 68  | 323 | 140 | 263 | 70  | 20  | 113.4 | 56.9 | 0.68 | 1.78 | 171 | 45  | 3.1 | 27.9 | 26.0 | 25  | 50  | 1123.5 | 18.9 |
| 6           | 94.1 | 1.643 | 364 | 1.631 | 13.30 | 313 | 234 | 203 | 177 | 143 | 1  | 24 | 20  | 28  | 28  | 1101 | 68  | 443 | 362 | 332 | 84  | 20  | 112.7 | 57.1 | 0.63 | 1.73 | 204 | 107 | 3.1 | 31.4 | 26.9 | 25  | 50  | 1123.5 | 18.9 |
| 7           | 91.3 | 1.451 | 251 | 1.629 | 13.30 | 313 | 252 | 207 | 183 | 146 | 2  | 24 | 20  | 28  | 28  | 1101 | 68  | 323 | 140 | 263 | 70  | 20  | 113.4 | 56.9 | 0.68 | 1.66 | 157 | 50  | 3.2 | 23.6 | 23.4 | 25  | 50  | 1123.5 | 18.9 |
| 8           | 93.9 | 1.270 | 305 | 1.636 | 13.30 | 305 | 252 | 208 | 185 | 146 | 2  | 24 | 20  | 28  | 28  | 1101 | 65  | 323 | 140 | 263 | 70  | 20  | 113.4 | 56.9 | 0.68 | 1.69 | 179 | 62  | 3.4 | 16.8 | 24.9 | 25  | 50  | 1123.5 | 19.2 |
| 9           | 93.6 | 1.380 | 309 | 1.630 | 13.30 | 306 | 258 | 207 | 186 | 146 | 2  | 24 | 20  | 28  | 28  | 1101 | 68  | 323 | 140 | 263 | 70  | 20  | 113.4 | 56.9 | 0.68 | 1.91 | 154 | 117 | 3.4 | 24.0 | 24.3 | 25  | 50  | 1123.5 | 19.2 |
| 10          | 94.0 | 1.329 | 300 | 1.640 | 13.30 | 300 | 268 | 202 | 178 | 148 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 443 | 362 | 332 | 84  | 20  | 112.7 | 57.1 | 0.63 | 1.70 | 156 | 120 | 2.5 | 30.0 | 26.1 | 25  | 50  | 1123.5 | 19.2 |
| 11          | 91.5 | 1.381 | 293 | 1.641 | 13.30 | 305 | 261 | 199 | 177 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 463 | 193 | 323 | 62  | 20  | 112.2 | 56.8 | 0.6  | 1.84 | 163 | 67  | 2.2 | 34.1 | 27.4 | 25  | 50  | 1123.5 | 19.2 |
| 12          | 93.7 | 1.407 | 394 | 1.646 | 13.20 | 303 | 257 | 199 | 174 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 443 | 362 | 332 | 84  | 20  | 112.7 | 57.1 | 0.63 | 1.70 | 156 | 120 | 2.5 | 30.0 | 26.1 | 25  | 50  | 1123.5 | 19.2 |
| 13          | 93.3 | 1.436 | 306 | 1.633 | 13.30 | 306 | 260 | 200 | 176 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 443 | 362 | 332 | 84  | 20  | 112.7 | 57.1 | 0.63 | 1.70 | 156 | 120 | 2.5 | 30.0 | 26.1 | 25  | 50  | 1127.3 | 19.2 |
| 14          | 93.6 | 1.535 | 321 | 1.648 | 13.30 | 301 | 256 | 203 | 178 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 463 | 193 | 323 | 62  | 20  | 112.2 | 56.8 | 0.6  | 1.74 | 153 | 79  | 2.6 | 25.9 | 27.3 | 25  | 50  | 1127.3 | 18.9 |
| 15          | 94.9 | 1.547 | 281 | 1.641 | 13.30 | 311 | 260 | 203 | 180 | 148 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 463 | 193 | 323 | 62  | 20  | 112.2 | 56.8 | 0.6  | 1.96 | 171 | 121 | 2.2 | 35.0 | 26.4 | 25  | 50  | 1124.6 | 18.9 |
| 16          | 92.1 | 1.609 | 312 | 1.623 | 13.30 | 311 | 275 | 195 | 172 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 68  | 425 | 232 | 292 | 59  | 25  | 112.9 | 57.4 | 0.6  | 2.16 | 181 | 71  | 1.9 | 18.5 | 28.2 | 25  | 50  | 1124.6 | 18.9 |
| 17          | 93.7 | 1.423 | 288 | 1.625 | 13.30 | 296 | 265 | 200 | 179 | 148 | 3  | 24 | 21  | 27  | 28  | 1101 | 68  | 425 | 232 | 292 | 59  | 25  | 112.9 | 57.4 | 0.6  | 2.16 | 181 | 71  | 1.9 | 18.5 | 28.2 | 25  | 50  | 1124.6 | 18.9 |
| 18          | 91.9 | 1.533 | 486 | 1.645 | 13.30 | 282 | 261 | 202 | 178 | 148 | 3  | 24 | 20  | 28  | 28  | 1101 | 65  | 425 | 232 | 292 | 59  | 25  | 112.9 | 57.4 | 0.6  | 2.27 | 175 | 64  | 3.2 | 29.7 | 28.1 | 25  | 50  | 1124.6 | 19.0 |
| 19          | 91.9 | 1.493 | 530 | 1.635 | 13.30 | 316 | 259 | 201 | 179 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 68  | 425 | 232 | 292 | 59  | 25  | 112.9 | 57.4 | 0.6  | 2.27 | 175 | 64  | 3.2 | 29.7 | 28.1 | 25  | 50  | 1124.6 | 19.0 |
| 20          | 92.7 | 1.496 | 298 | 1.632 | 13.30 | 300 | 259 | 202 | 179 | 147 | 3  | 24 | 20  | 28  | 28  | 1101 | 68  | 463 | 193 | 323 | 62  | 20  | 112.2 | 56.8 | 0.6  | 1.79 | 150 | 69  | 2.8 | 33.9 | 26.8 | 25  | 50  | 1124.6 | 19.0 |
| 21          | 92.6 | 1.582 | 410 | 1.626 | 13.30 | 302 | 255 | 204 | 182 | 148 | 3  | 24 | 20  | 28  | 28  | 1101 | 68  | 425 | 232 | 292 | 59  | 25  | 112.9 | 57.4 | 0.6  | 2.27 | 175 | 64  | 3.2 | 29.7 | 28.1 | 25  | 50  | 1124.6 | 19.0 |
| 22          | 92.3 | 1.750 | 292 | 1.634 | 13.20 | 310 | 262 | 204 | 184 | 147 | 3  | 26 | 20  | 27  | 27  | 1101 | 65  | 486 | 112 | 374 | 76  | 40  | 113.3 | 56.6 | 0.66 | 2.41 | 198 | 100 | 2.7 | 27.4 | 26.6 | 25  | 50  | 1123.0 | 19.0 |
| 23          | 93.0 | 1.681 | 302 | 1.635 | 13.30 | 310 | 261 | 202 | 179 | 146 | 3  | 24 | 20  | 28  | 28  | 1101 | 68  | 463 | 193 | 323 | 62  | 20  | 112.2 | 56.8 | 0.6  | 1.96 | 171 | 121 | 2.2 | 35.0 | 26.4 | 25  | 50  | 1123.0 | 18.8 |
| 24          | 91.5 | 1.529 | 292 | 1.631 | 13.20 | 309 | 250 | 205 | 183 | 148 | 3  | 26 | 20  | 27  | 27  | 1101 | 68  | 486 | 112 | 374 | 76  | 40  | 113.3 | 56.6 | 0.66 | 2.41 | 198 | 100 | 2.7 | 27.4 | 26.6 | 25  | 50  | 1123.0 | 18.8 |
| 25          | 93.5 | 1.541 | 295 | 1.634 | 13.20 | 305 | 266 | 204 | 183 | 150 | 3  | 26 | 20  | 27  | 27  | 1101 | 65  | 486 | 112 | 374 | 76  | 40  | 113.3 | 56.6 | 0.66 | 2.41 | 198 | 100 | 2.7 | 27.4 | 26.6 | 25  | 50  | 1123.0 | 18.8 |
| 26          | 92.5 | 1.507 | 384 | 1.626 | 13.10 | 292 | 261 | 201 | 179 | 148 | 3  | 26 | 20  | 26  | 28  | 1101 | 69  | 486 | 112 | 374 | 76  | 40  | 113.3 | 56.6 | 0.66 | 2.87 | 191 | 67  | 2.6 | 28.6 | 27.9 | 25  | 50  | 1121.4 | 18.8 |
| 27          | 91.8 | 1.733 | 348 | 1.631 | 13.20 | 305 | 257 | 205 | 181 | 149 | 3  | 26 | 20  | 27  | 27  | 1101 | 68  | 486 | 112 | 374 | 76  | 40  | 113.3 | 56.6 | 0.66 | 2.87 | 191 | 67  | 2.6 | 28.6 | 27.9 | 25  | 50  | 1121.4 | 18.8 |
| 28          | 95.0 | 1.683 | 295 | 1.627 | 13.20 | 302 | 264 | 205 | 182 | 149 | 3  | 26 | 20  | 28  | 26  | 1101 | 68  | 383 | 136 | 319 | 76  | 20  | 112.5 | 57.2 | 0.56 | 2.13 | 189 | 92  | 3.3 | 29.5 | 27.6 | 25  | 50  | 1121.4 | 18.8 |
| 29          | 91.1 | 1.622 | 392 | 1.632 | 13.20 | 306 | 264 | 201 | 176 | 147 | 3  | 26 | 20  | 28  | 26  | 1101 | 68  | 383 | 136 | 319 | 76  | 20  | 112.5 | 57.2 | 0.56 | 2.31 | 203 | 104 | 2.7 | 29.8 | 27.9 | 25  | 50  | 1121.4 | 18.8 |
| 30          | 89.6 | 1.697 | 340 | 1.629 | 13.20 | 315 | 265 | 204 | 182 | 148 | 3  | 26 | 20  | 26  | 28  | 1101 | 68  | 383 | 136 | 319 | 76  | 20  | 112.5 | 57.2 | 0.56 | 2.35 | 194 | 85  | 2.4 | 31.4 | 28.2 | 25  | 50  | 1121.4 | 18.8 |

## **Appendix 5:**

### **SIMCA software output.**

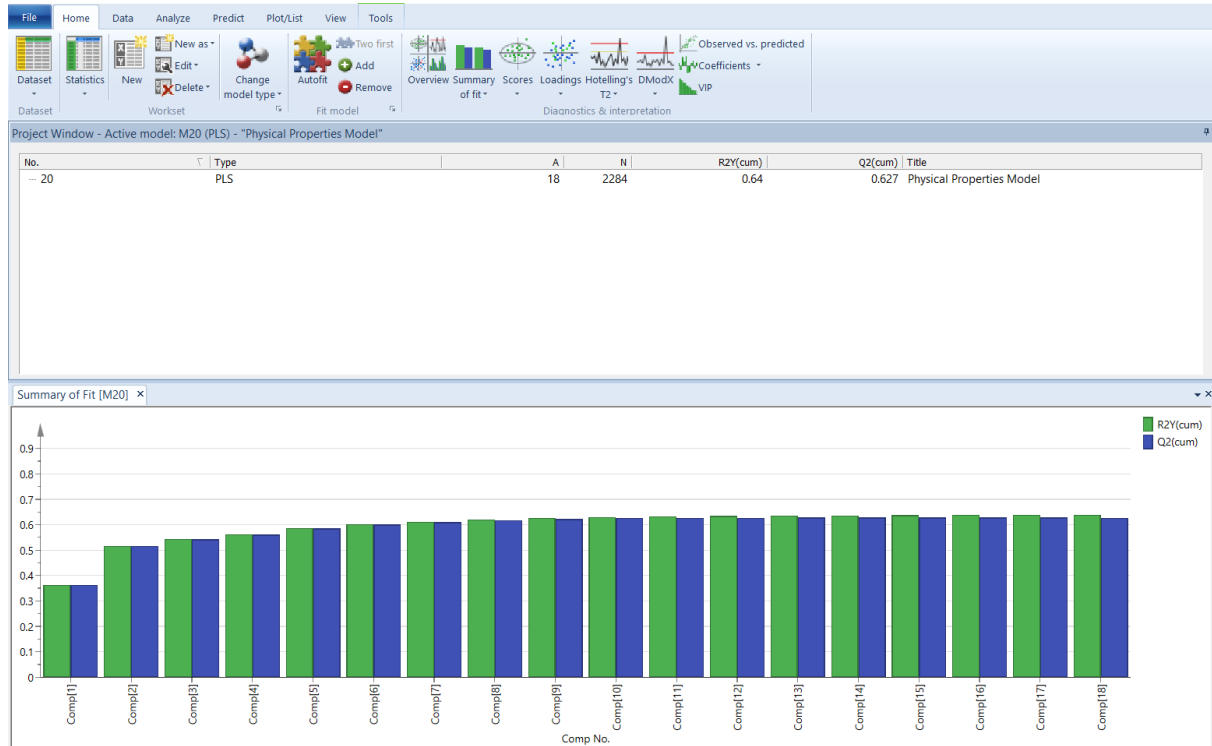
**This Appendix contains the following two sections:**

**Section A: SIMCA software output from Model 1.**

**Section B: SIMCA software output from Model 2.**

## Section A: SIMCA software output from Model 1.

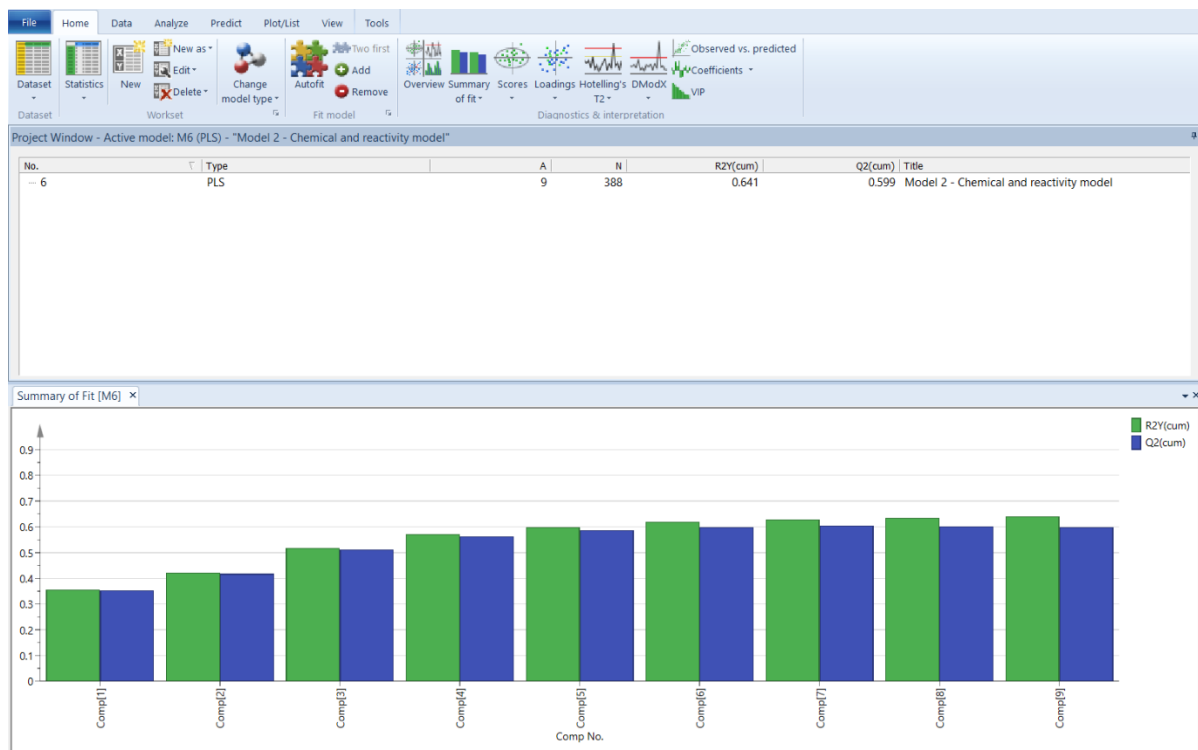
The first figure is the summary of the PLS model results. The second figure is the CV-ANOVA results.



|    | 1                                     | 2       | 3    | 4        | 5       | 6          | 7        |
|----|---------------------------------------|---------|------|----------|---------|------------|----------|
| 1  | <b>M20(Physical Properties Model)</b> |         |      |          |         |            |          |
| 2  | <b>y1</b>                             |         |      |          |         |            |          |
| 3  | <b>Total corr.</b>                    | 2283    | 2284 | 0.999562 |         |            | 0.999781 |
| 4  | <b>Regression</b>                     | 901.91  | 36   | 25.053   | 40.7788 | 0          | 5.0053   |
| 5  | <b>Residual</b>                       | 1381.09 | 2248 | 0.614364 |         |            | 0.783814 |
| 6  | <b>y2</b>                             |         |      |          |         |            |          |
| 7  | <b>Total corr.</b>                    | 2283    | 2284 | 0.999562 |         |            | 0.999781 |
| 8  | <b>Regression</b>                     | 50.459  | 36   | 1.40164  | 1.41134 | 0.0537559  | 1.18391  |
| 9  | <b>Residual</b>                       | 2232.54 | 2248 | 0.993123 |         |            | 0.996556 |
| 10 | <b>y3</b>                             |         |      |          |         |            |          |
| 11 | <b>Total corr.</b>                    | 2283    | 2284 | 0.999562 |         |            | 0.999781 |
| 12 | <b>Regression</b>                     | 134.227 | 36   | 3.72852  | 3.90077 | 4.479e-014 | 1.93094  |
| 13 | <b>Residual</b>                       | 2148.77 | 2248 | 0.95586  |         |            | 0.977681 |

## Section B: SIMCA software output from Model 2.

The first figure is the summary of the PLS model results. The second figure is the CV-ANOVA results.



|    | 1                                           | 2       | 3   | 4        | 5       | 6            | 7        |
|----|---------------------------------------------|---------|-----|----------|---------|--------------|----------|
| 1  | M6(Model 2 - Chemical and reactivity model) | SS      | DF  | MS       | F       | p            | SD       |
| 2  | y4                                          |         |     |          |         |              |          |
| 3  | Total corr.                                 | 387     | 388 | 0.997423 |         |              | 0.998711 |
| 4  | Regression                                  | 47.885  | 18  | 2.66028  | 2.90257 | 7.7829e-005  | 1.63104  |
| 5  | Residual                                    | 339.115 | 370 | 0.916527 |         |              | 0.957354 |
| 6  | y5                                          |         |     |          |         |              |          |
| 7  | Total corr.                                 | 387     | 388 | 0.997423 |         |              | 0.998711 |
| 8  | Regression                                  | 127.502 | 18  | 7.08343  | 10.0997 | 4.84377e-023 | 2.66147  |
| 9  | Residual                                    | 259.498 | 370 | 0.701347 |         |              | 0.837465 |
| 10 | y6                                          |         |     |          |         |              |          |
| 11 | Total corr.                                 | 387     | 388 | 0.997423 |         |              | 0.998711 |
| 12 | Regression                                  | 46.6781 | 18  | 2.59323  | 2.81937 | 0.000123822  | 1.61035  |
| 13 | Residual                                    | 340.322 | 370 | 0.919789 |         |              | 0.959056 |

## **Appendix 6:**

### **Additional details on the statistical testing parameters.**

**This Appendix contains the following two sections:**

**Section A: Statistical testing parameters for Model 1.**

**Section B: Statistical testing parameters for Model 2.**

### Section A: Statistical testing parameters for Model 1.

| Variables                         | $SE_{\beta}$ | $\sum (x_i - \bar{x})^2$ | t-value | p-value |
|-----------------------------------|--------------|--------------------------|---------|---------|
| y <sub>1</sub> VS X <sub>18</sub> | 0.016        | 2283.0                   | -0.305  | 0.760   |
| y <sub>1</sub> VS X <sub>20</sub> | 0.003        | 86961.2                  | 4.143   | 0.000   |
| y <sub>1</sub> VS X <sub>21</sub> | 0.016        | 2283.0                   | 1.160   | 0.246   |
| y <sub>1</sub> VS X <sub>26</sub> | 0.016        | 2283.0                   | -0.427  | 0.669   |
| y <sub>1</sub> VS X <sub>6</sub>  | 0.016        | 2283.0                   | -0.183  | 0.855   |
| y <sub>2</sub> VS X <sub>10</sub> | 0.028        | 1240.2                   | -0.071  | 0.944   |
| y <sub>2</sub> VS X <sub>13</sub> | 0.006        | 27100.9                  | 2.811   | 0.005   |
| y <sub>2</sub> VS X <sub>14</sub> | 0.015        | 4602.5                   | -0.341  | 0.733   |
| y <sub>2</sub> VS X <sub>16</sub> | 0.021        | 2283.0                   | 0.816   | 0.415   |
| y <sub>2</sub> VS X <sub>18</sub> | 0.021        | 2283.0                   | 0.096   | 0.924   |
| y <sub>2</sub> VS X <sub>19</sub> | 0.021        | 2283.0                   | -0.144  | 0.886   |
| y <sub>2</sub> VS X <sub>22</sub> | 0.021        | 2283.0                   | 0.528   | 0.598   |
| y <sub>2</sub> VS X <sub>23</sub> | 0.021        | 2283.0                   | -0.624  | 0.533   |
| y <sub>2</sub> VS X <sub>25</sub> | 0.021        | 2283.0                   | 0.768   | 0.443   |
| y <sub>2</sub> VS X <sub>27</sub> | 0.021        | 2283.0                   | -0.048  | 0.962   |
| y <sub>2</sub> VS X <sub>28</sub> | 0.021        | 2283.0                   | 0.672   | 0.502   |
| y <sub>2</sub> VS X <sub>29</sub> | 0.017        | 3498.0                   | -0.832  | 0.406   |
| y <sub>2</sub> VS X <sub>3</sub>  | 0.021        | 2283.0                   | 0.432   | 0.666   |
| y <sub>2</sub> VS X <sub>7</sub>  | 0.021        | 2283.0                   | 0.240   | 0.810   |
| y <sub>3</sub> VS X <sub>1</sub>  | 0.020        | 2283.0                   | 0.098   | 0.922   |
| y <sub>3</sub> VS X <sub>13</sub> | 0.006        | 27100.9                  | 3.203   | 0.001   |
| y <sub>3</sub> VS X <sub>14</sub> | 0.014        | 4602.5                   | 0.069   | 0.945   |
| y <sub>3</sub> VS X <sub>16</sub> | 0.020        | 2283.0                   | -0.783  | 0.434   |
| y <sub>3</sub> VS X <sub>18</sub> | 0.020        | 2283.0                   | -0.342  | 0.732   |
| y <sub>3</sub> VS X <sub>19</sub> | 0.020        | 2283.0                   | 0.245   | 0.807   |
| y <sub>3</sub> VS X <sub>20</sub> | 0.003        | 86961.2                  | -1.208  | 0.227   |
| y <sub>3</sub> VS X <sub>22</sub> | 0.020        | 2283.0                   | -0.685  | 0.493   |
| y <sub>3</sub> VS X <sub>25</sub> | 0.020        | 2283.0                   | -0.196  | 0.845   |
| y <sub>3</sub> VS X <sub>27</sub> | 0.020        | 2283.0                   | -0.147  | 0.883   |
| y <sub>3</sub> VS X <sub>28</sub> | 0.020        | 2283.0                   | 0.538   | 0.591   |
| y <sub>3</sub> VS X <sub>3</sub>  | 0.020        | 2283.0                   | -0.440  | 0.660   |
| y <sub>3</sub> VS X <sub>31</sub> | 0.020        | 2283.0                   | 0.636   | 0.525   |
| y <sub>3</sub> VS X <sub>7</sub>  | 0.020        | 2283.0                   | 0.440   | 0.660   |

**Section B: Statistical testing parameters for Model 2.**

| Variables                         | $SE_{\beta}$ | $\sum (x_i - \bar{x})^2$ | t-value | p-value |
|-----------------------------------|--------------|--------------------------|---------|---------|
| y <sub>4</sub> VS x <sub>12</sub> | 0.025        | 1737.1                   | -0.673  | 0.501   |
| y <sub>4</sub> VS x <sub>15</sub> | 0.054        | 387.0                    | 0.318   | 0.751   |
| y <sub>4</sub> VS x <sub>16</sub> | 0.054        | 387.0                    | -0.056  | 0.955   |
| y <sub>4</sub> VS x <sub>17</sub> | 0.054        | 387.0                    | 0.075   | 0.941   |
| y <sub>4</sub> VS x <sub>18</sub> | 0.054        | 387.0                    | 0.000   | 1.000   |
| y <sub>4</sub> VS x <sub>19</sub> | 0.054        | 387.0                    | 0.019   | 0.985   |
| y <sub>4</sub> VS x <sub>24</sub> | 0.054        | 387.0                    | -0.075  | 0.940   |
| y <sub>4</sub> VS x <sub>26</sub> | 0.054        | 387.0                    | -0.093  | 0.926   |
| y <sub>4</sub> VS x <sub>32</sub> | 0.054        | 387.0                    | 0.318   | 0.751   |
| y <sub>4</sub> VS x <sub>8</sub>  | 0.038        | 776.5                    | 0.185   | 0.853   |
| y <sub>5</sub> VS x <sub>1</sub>  | 0.047        | 387.0                    | -0.385  | 0.701   |
| y <sub>5</sub> VS x <sub>16</sub> | 0.047        | 387.0                    | 0.256   | 0.798   |
| y <sub>5</sub> VS x <sub>19</sub> | 0.047        | 387.0                    | -0.299  | 0.765   |
| y <sub>6</sub> VS x <sub>18</sub> | 0.054        | 776.5                    | 0.075   | 0.941   |
| y <sub>6</sub> VS x <sub>2</sub>  | 0.026        | 1648.7                   | 0.270   | 0.788   |
| y <sub>6</sub> VS x <sub>22</sub> | 0.054        | 387.0                    | -0.149  | 0.881   |
| y <sub>6</sub> VS x <sub>29</sub> | 0.044        | 566.0                    | 0.023   | 0.982   |
| y <sub>6</sub> VS x <sub>8</sub>  | 0.038        | 776.5                    | 0.344   | 0.731   |

## **Appendix 7:**

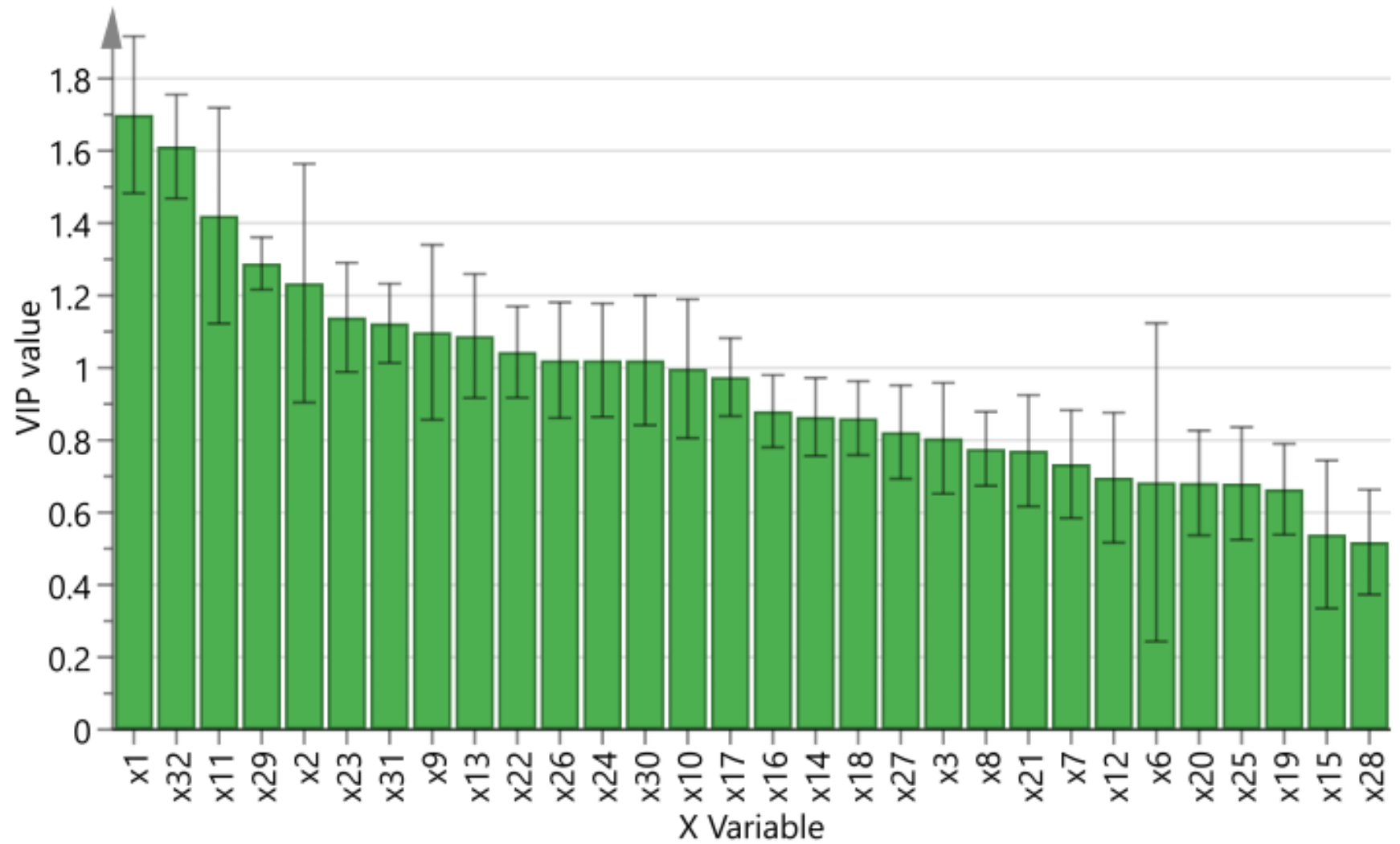
### **Full VIP plot values.**

**This Appendix contains the following two sections:**

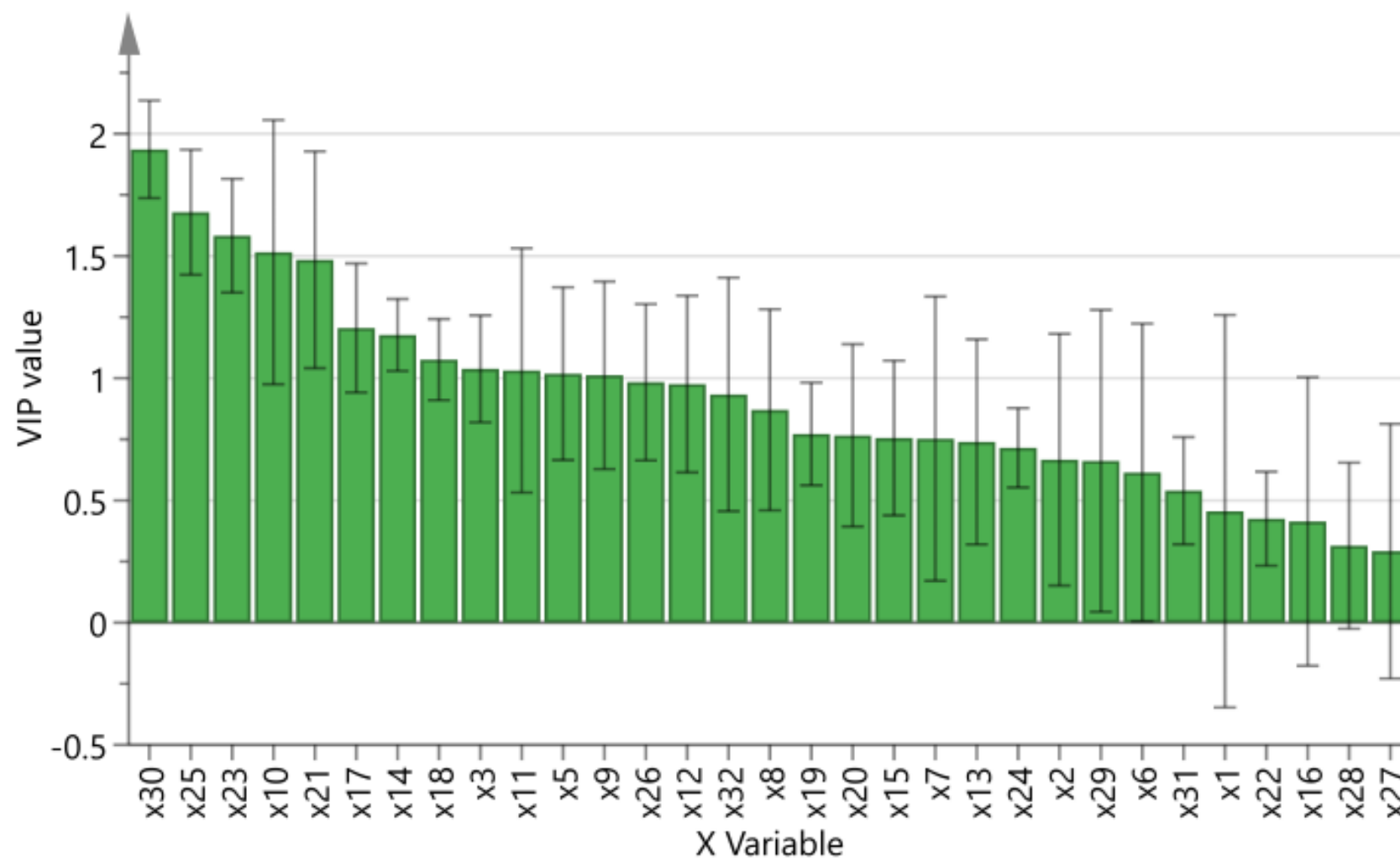
**Section A: Full VIP plot values of Model 1.**

**Section B: Full VIP plot values of Model 2.**

**Section A: Full VIP plot values of Model 1.**



### Section B: Full VIP plot values of Model 2.



## **Appendix 8:**

**Calculation of NCC based on model built.**

### **Calculation of NCC based on model built:**

- The average real NCC of Carbon 2 during the research period from November 2020 to October 2022 is 412.17 kg/ton of aluminum.
- Average anode geometric density during the same period (average observed geometric density) is 1.576 g/cm<sup>3</sup>.
- Anode gematric density obtained from the best variables is 1.610 g/cm<sup>3</sup>.

First, the empirical equation constant is calculated as follows:

$$\alpha = \rho_a(NCC - 334) = 1.576(412.17 - 334) = 123.2$$

Then, NCC based on the model is calculated as follows:

$$NCC = 334 + 123.2 \left( \frac{1}{1.610} \right) = 410.52 \text{ kg/ton of Aluminum}$$

Finally, the reduced NCC value is calculated:

$$\text{Reduced NCC} = 412.17 - 410.52 = 1.65 \text{ kg/ton of Aluminum}$$

## ملخص الدراسة

يُعد إنتاج قطب الأنود الكربوني عملية حاسمة في إنتاج الألمنيوم ويشكل نسبة عالية من تكاليف الإنتاج. لذلك، فإن تحسين قطب الأنود أمر ضروري لتقليل التكاليف وزيادة الإنتاج السنوي. تتأثر جودة قطب الأنود الكربوني بشكل كبير بخصائص المواد الخام ومعلمات التشغيل. المشكلة التي سيتم تناولها في هذه الدراسة البحثية هي عدم وجود علاقة رياضية معروفة أو نموذج للتحكم في جودة أقطاب الأنود المنتجة. الأهداف الرئيسية للدراسة هي تطوير نموذج رياضي لجودة قطب الأنود يعتمد على معلمات التشغيل الخاصة به، وتحديد معلمات التشغيل المهمة التي تحدد جودة قطب الأنود، ودراسة الأسباب الرئيسية لتغير جودة قطب الأنود، وتحديد العلاقة بين جودة قطب الأنود ومعلمات التشغيل. تم تحقيق أهداف الدراسة باستخدام طريقة إحصائية متعددة المتغيرات (نمذجة المتغيرات الكامنة) لبناء نموذج رياضي يبين خصائص جودة قطب الأنود ومعلمات التشغيل. ثم تم استخدام النموذج المبني للتنبؤ بمجموعة بيانات جديدة للتحقق من قدرة التنبؤية. بالإضافة إلى ذلك، تم استخدام طرق تفسير مختلفة للحصول على معلمات التشغيل المهمة ودراسة الأسباب الرئيسية لتغير جودة قطب الأنود. البيانات التي تم جمعها لتحقيق هذه الأهداف تشمل خصائص جودة قطب الأنود الفعلية ومعلمات التشغيل وبيانات خصائص المواد الخام المستخرجة من مصنع الكربون 2 في شركة ألمنيوم البحرين (ألبا). تم بناء نموذجين (PLS) لتحسين الخصائص الفيزيائية والخصائص الكيميائية والتفاعلية لقطب الأنود باستخدام برنامج (SIMCA). قدمت الدراسة نماذج جيدة لجودة المصعد مع نسبة تباين بلغت 64.0% و 64.1% لنموذج الخصائص الفيزيائية (النموذج 1) ونموذج الخصائص الكيميائية والتفاعلية (النموذج 2) على التوالي. علاوة على ذلك، كانت قدرة التنبؤ للنموذج مرتفعة، حيث بلغت نسبة التباين 51.1% و 57.2% للنموذجين الأول والثاني على التوالي. كما أظهرت الدراسة أن المصدر الرئيسي لتغير جودة قطب الأنود كان التباين في خصائص المواد الخام، وتحديدًا خصائص فحم الكوك البترولي المحمص (CPC). بالإضافة إلى ذلك، أوضحت الدراسة أن المتغير الأكثر تأثيراً للنموذج الأول كان كثافة قطب الأنود الأخضر (GAD)، والنموذج الثاني كان محتوى الحديد في فحم الكوك البترولي ومعلمات التشغيل في مصنع العجينة. علاوة على ذلك، كانت المتغيرات الأهم هي كثافة قطب الأنود الأخضر ووقت التثريب الحراري للنموذجين الأول والثاني على التوالي. وقد قدرت الدراسة خفض الاستهلاك النوعي للكربون باستخدام النموذج المبني ووجدت أنه يبلغ 1.65 كجم لكل طن من الألمنيوم. وبناءً على ذلك، تم تقدير تخفيض التكلفة بحوالي 825,000 دولار أمريكي سنويًا.

جامعة البحرين  
كلية الهندسة  
قسم الهندسة الميكانيكية



# تحسين جودة قطب الأنود بناءً على معلمات التشغيل في صناعة الألمنيوم

(دراسة حالة من مملكة البحرين)

أطروحة مقدمة كجزء من متطلبات الحصول على درجة الماجستير في الإدارة الهندسية

إعداد:

احمد محمد يوسف مشكور

20132683

إشراف

د. أسامة الجمل

(أستاذ مساعد)

د. ياسر العلوي

(أستاذ مساعد)

مملكة البحرين

فبراير 2025