Intelligence-Driven Precipitation Control for Enhanced Alumina Productivity with Consistent Quality: A Machine Learning Approach

Paul Gupta[1], Subhadeep Bhattacharya[2], Keshav Karn[3] and Salman Hussain[4]
1. Assistant Vice President, Utkal Alumina International Limited, Rayagada, India
2. Manager, Asset Care Alumina Refinery
3. AGM, Alumina Refinery
4. General Manager, Digital
Hindalco Industries, Renukoot, Uttar Pradesh, India
Corresponding author: paul.gupta@adityabirla.com
https://doi.org/10.71659/icsoba2024-aa020

Abstract

Alumina extraction from boehmite in bauxite, such as in Central India, poses a unique challenge due to its inherently high energy requirements. Global refineries employing such energy-intensive processes must have an acute focus on energy efficiency to ensure competitive production costs. A pivotal efficiency metric is precipitation productivity or yield, indicative of refinery operational efficiency. Traditionally, maintaining optimum yield involves manual analysis of process sample results, where technical analysts meticulously correlate deviations in process conditions with lab values to identify influential factors. This approach is sensitive and heavily relies on human experience and expertise, underscoring the necessity for automated and intelligent solutions.

In response to this challenge, this article presents a pioneering endeavour to develop a mathematical model utilizing machine learning (ML) algorithms, coupled with Bayer process principles and noise factor considerations, to optimize yield in alumina precipitation. Notably, the model partitions the precipitation process into three segments: agglomeration, new growth, and existing growth. This facilitates a granular understanding and targeted intervention. Crucially, our model integrates constraints within each segment, ensuring that the quality of the resultant alumina remains uncompromised while optimizing yield. A transformative shift towards data-driven precipitation control is anticipated by leveraging the power of ML algorithms to optimize productivity while consistently delivering quality alumina. This intrinsic balance between productivity and product quality constitutes a hallmark of our approach, which not only enhances operational excellence but also underscores the disruptive potential of AI-driven solutions in alumina refining.

Keywords: Alumina Production, Precipitation Control, Machine Learning, Artificial Intelligence, Operational Efficiency.

1. Introduction

Precipitation productivity stands as the key performance indicator for an alumina refinery, measuring the amount of trihydrate alumina extracted per litre of Pregnant Liquor (PGL). This metric is influenced by various factors, including temperature profiles, seed charge, residence time, liquor caustic concentration, alumina supersaturation and seed surface area (SSA) [1].

At Renukoot Alumina Refinery, achieving continuous improvement in precipitation productivity from 73 to 78 g/L involved implementing various process strategies combined with enhanced operational practices. This boost in productivity translated into significant reductions in steam and power consumption for the refinery. However, several instances of lower yield coupled with deviations in quality parameters for calcined alumina, such as occluded Na_2O, < 45 µm fractions and attrition index, were observed.

In the absence of predictive-based control, consistently sustaining the precipitation productivity at this level remained a challenge [2]. Consequently, it became imperative to establish a set of process conditions that not only met but surpassed the targeted productivity level of 77 g/L while ensuring consistent alumina quality.

This paper discusses the stepwise development of predictive and intelligence-driven precipitation control for enhanced alumina productivity with consistent quality using a machine learning approach.

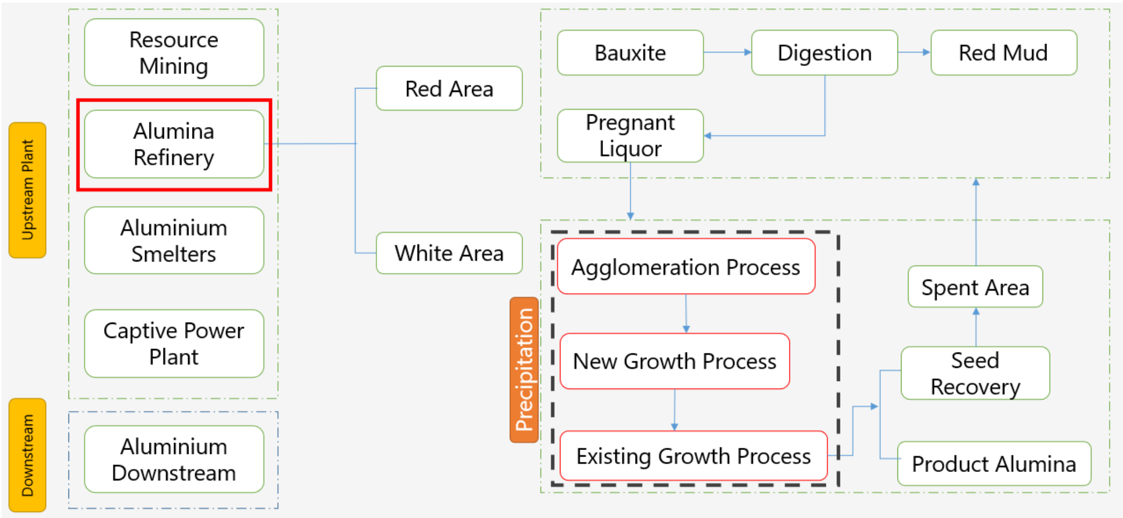

Figure 1. Precipitation process at Renukoot Alumina Refinery.

The precipitation process in alumina refining can be divided into three main process segments namely as shown in Figure 1.

i. Agglomeration
ii. New growth
iii. Existing growth

To ensure a short residence duration, ranging from about 6 to approximately 12 hours, the agglomeration phase of the precipitation process is ideally conducted in a short chain of tanks. These are homogenous tanks, and the slurry is released into a series of larger tanks where the growth occurs. The new growth segment consists of seed addition followed by interstage coolers for optimizing liquor temperature profile. The mid-precipitated slurry is directed from new growth to existing growth segment where sufficient residence time is provided. Lowering soda levels and achieving higher productivity requires optimizing and balancing a number of process variables. Each segment has distinct characteristics and challenges that need to be addressed to optimize the overall precipitation productivity.

2. Technical Details

2.1 Machine Learning and Operations Research Optimization

Machine learning (ML) is a discipline of artificial intelligence (AI) that provides machines with the ability to automatically learn from data and past experiences while identifying patterns to make predictions with minimal human intervention. ML algorithms use computation methods to learn directly from data instead of relying on any predetermined equation that may serve as a model [3].

Operations Research (OR) is an analytical method of problem-solving and decision-making useful in the management of organizations [4]. In operations research, problems are broken down into basic components and then solved in defined steps by mathematical analysis.

The ML and OR approach is shown in Figure 2. The process of operations research used in this presented case be broadly broken down into the following steps:

1. Identifying a problem that needs to be solved.
2. Constructing a model around the problem that resembles the real world and variables.
3. Creating a constraint around the model to run in operating condition.
4. Testing each solution on the model and analysing its success.
5. Implementing the solution to the actual problem.

Given the limitations and challenges of improving productivity with given constraints, ML and OR methodologies are used together to create a data-driven constraint optimization model.

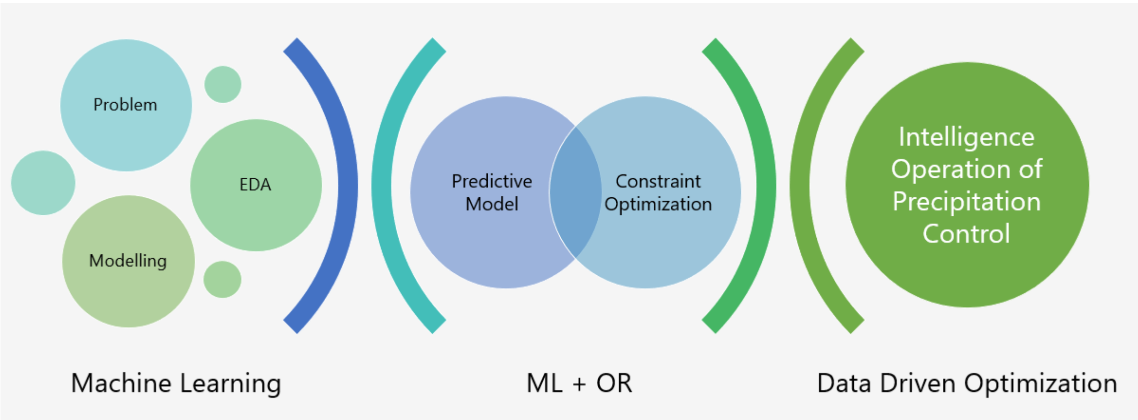

Figure 2. ML and OR approach.

2.2 Model Development

To develop a data-driven machine learning based optimization model for the precipitation process, the entire model development has been divided into three process areas: agglomeration, new growth and existing growth. Three local models were developed and then integrated. Data comprising DCS process data and chemical lab data were considered for the period FY22 and FY23. Data was stratified and sporadic variations were excluded.

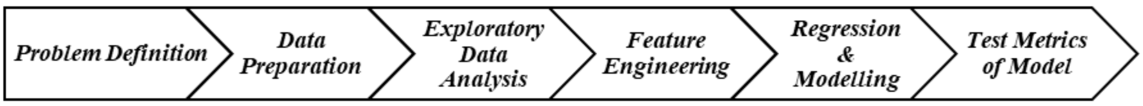

Figure 3. Steps followed in machine learning model development.

Figure 4. Parameters involved in model development of alumina yield.

As shown in Figure 3, after dividing the process, a local optimization model was run to improve the productivity for each process with help of ML and open-source technology, involving exploratory data analysis, feature engineering and modelling. The parameters involved in the overall development of the optimization model for the precipitation circuit are shown in Figure 4.

The stepwise strategy is shown in Figure 5, where individual ML model was developed for productivity, which were then integrated with the constraint-based model for soda and discharge temperature of new growth and existing growth circuits. A single-combined holistic model was then developed for optimization of the precipitation circuit.

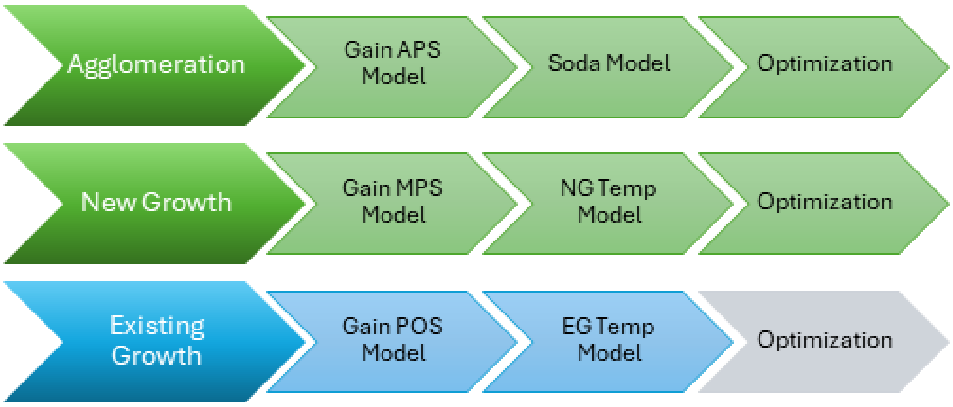

Figure 5. Optimization strategy for precipitation.

3. Results and Discussion

3.1 Regression Model

3.1.1 Agglomeration Model

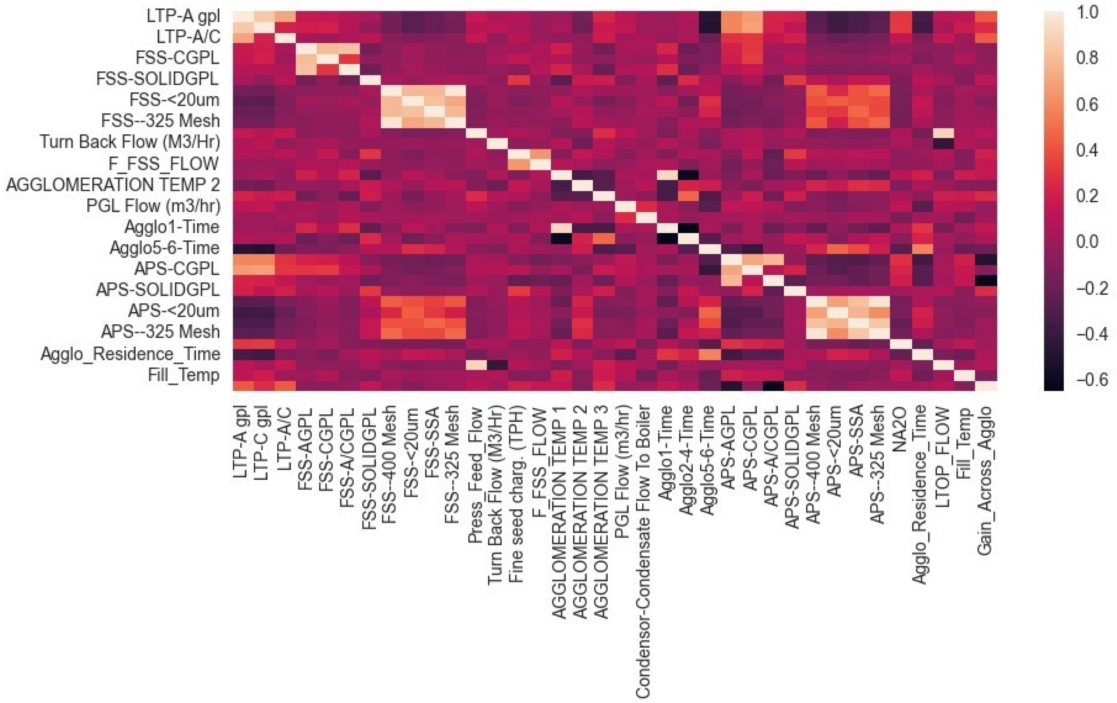

Figure 6. Heat map of variables for agglomeration productivity.

The process data is visualized using heat map to understand the important parameters affecting the agglomeration process, as shown in Figure 6. The agglomeration process is the initial phase of precipitation, where the chemistry of the process is defined with the help of supersaturation, equilibrium alumina, and input feed stream from the red area. The highly correlated parameters such as filling temperature, agglomeration time, and parameters related to supersaturation (namely PGL A/C ratio) are shown in Figure 7. Of all these parameters, agglomeration time appears to be the most influential.

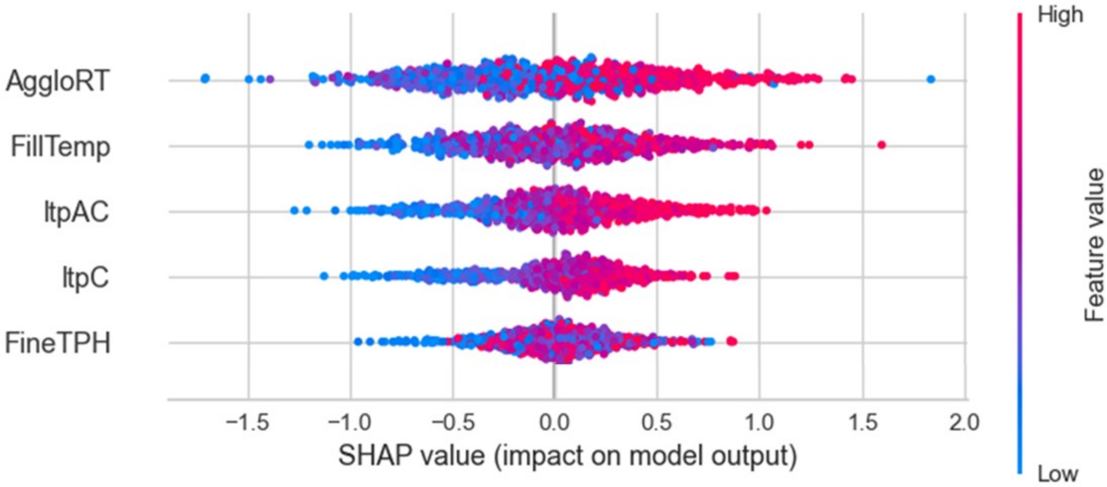

Figure 7. SHAP values for model input on agglomeration productivity.

A regression-based model was created to predict productivity across the agglomeration process. The ordinary least square regression (OLS) model is a statistical method used to estimate the relationship between a dependent variable (target variable) and independent variable. This is a fundamental technique used in linear regression analysis. The five parameters listed in Figure 7 are used for prediction of productivity. From Figure 8, it can be observed that the R^2 value for testing is 0.98. The model also has a high F-statistic value 1.165e+04 and low P value < 0.05, indicating that the model is statistically significant. The MAE and RMSE for the predicted values are 1.2742 and 1.6780 respectively.

```
                             OLS Regression Results
==============================================================================
Dep. Variable:      Gain_Across_Agglo   R-squared (uncentered):          0.980
Model:                            OLS   Adj. R-squared (uncentered):     0.980
Method:                 Least Squares   F-statistic:                 1.165e+04
Date:                Mon, 11 Sep 2023   Prob (F-statistic):               0.00
Time:                        14:49:38   Log-Likelihood:                -3264.1
No. Observations:                1184   AIC:                             6538.
Df Residuals:                    1179   BIC:                             6563.
Df Model:                           5
Covariance Type:            nonrobust
===============================================================================
                          coef    std err          t      P>|t|     [0.025     0.975]
-------------------------------------------------------------------------------
LTP-A/C                78.3489     10.031      7.811      0.000     58.668     98.030
LTP-C gpl               0.1821      0.027      6.734      0.000      0.129      0.235
Fine seed charg. (TPH)  0.0212      0.006      3.369      0.001      0.009      0.034
Agglo_Residence_Time    0.4586      0.145      3.159      0.002      0.174      0.743
Fill_Temp              -0.9501      0.095    -10.022      0.000     -1.136     -0.764
==============================================================================
Omnibus:                       12.432   Durbin-Watson:                   1.024
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               11.946
Skew:                           0.213   Prob(JB):                      0.00255
Kurtosis:                       2.753   Cond. No.                     2.33e+04
==============================================================================
```

Figure 8. Regression model output results for agglomeration productivity.

3.1.2 New Growth Model

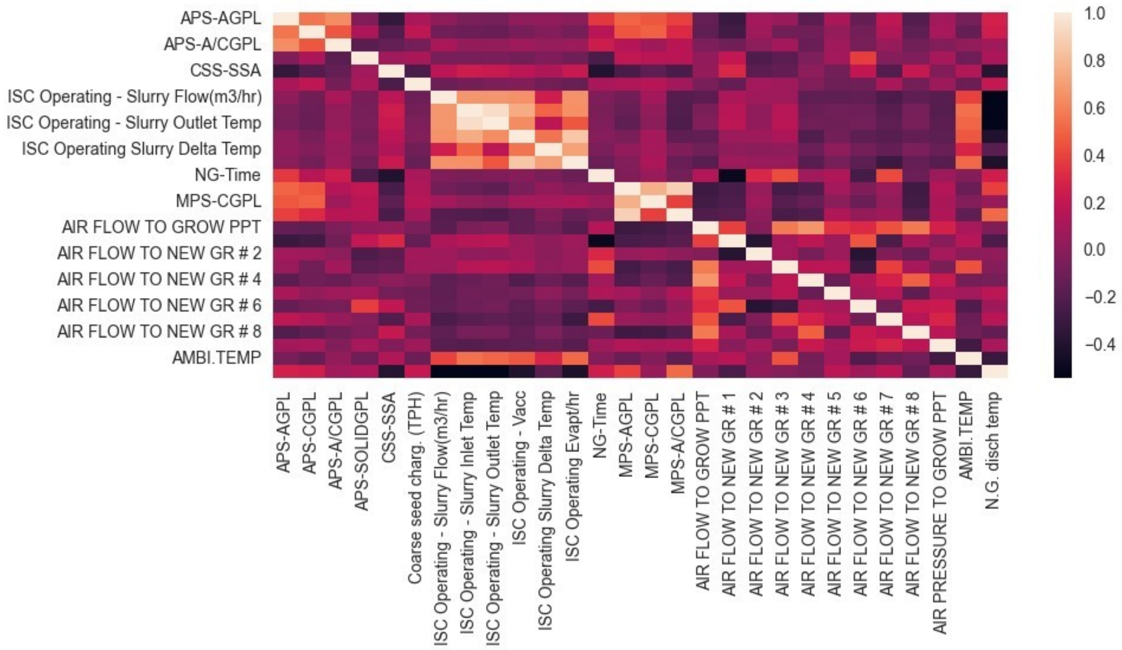

Figure 9. Heat map of variables for new growth discharge temperature.

The new growth process is the secondary part of the precipitation circuit. Maintaining productivity throughout the circuit highly depends on the chemistry of coarse seed charging, seed surface area, interstage cooling, and ambient temperature. Figure 9 shows the heat map of all independent parameters affecting the new growth productivity. Figure 10 shows the shortlisted significant parameters which were considered for the purpose of model building. Of all the parameters, agglomerated precipitation slurry (APS) A/C ratio appears to be the most influential parameter in comparison to other variables.

Figure 11 shows the results obtained from the OLS regression model. The R^2 value for the test value is 0.995. The model has a high F-statistic value 3.657+04 and low P value ~0. The model follows the trend with MAE 1.4712 and RMSE 1.6353.

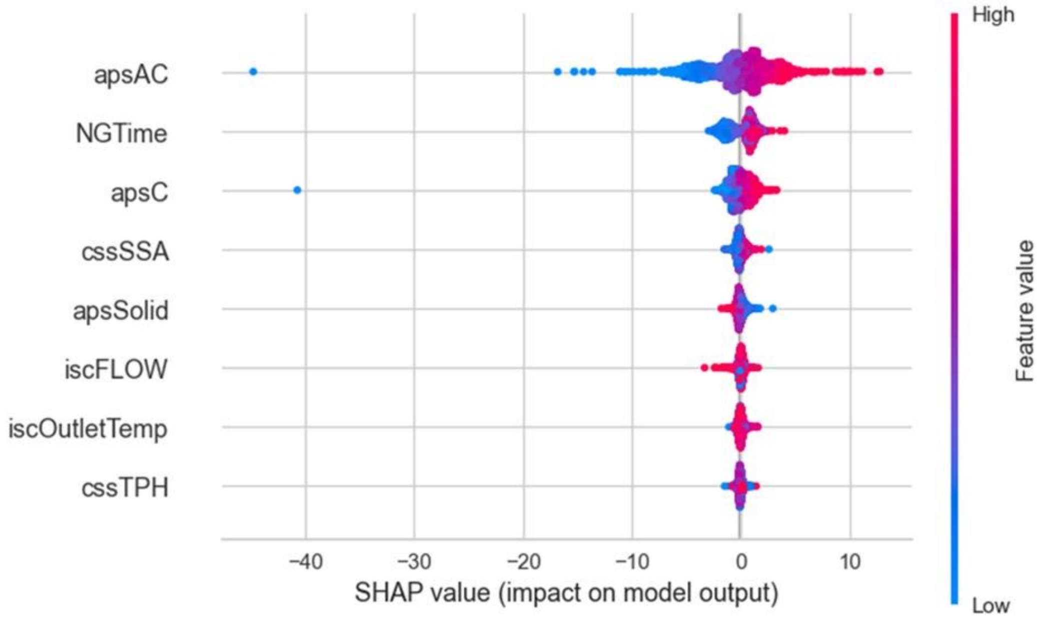

Figure 10. SHAP values for model input on new growth discharge temperature.

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:               gain_mps   R-squared (uncentered):                  0.995
Model:                            OLS   Adj. R-squared (uncentered):             0.995
Method:                 Least Squares   F-statistic:                         3.657e+04
Date:                Tue, 17 Oct 2023   Prob (F-statistic):                       0.00
Time:                        11:30:54   Log-Likelihood:                        -2357.5
No. Observations:                1032   AIC:                                     4727.
Df Residuals:                    1026   BIC:                                     4757.
Df Model:                           6
Covariance Type:            nonrobust
=====================================================================================================
                                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------
APS-AGPL                            0.6061      0.028     21.361      0.000       0.550       0.662
APS-CGPL                           -0.1815      0.017    -10.743      0.000      -0.215      -0.148
CSS-SSA                             0.0072      0.001      8.574      0.000       0.006       0.009
ISC Operating - Slurry Outlet Temp -0.2043      0.042     -4.866      0.000      -0.287      -0.122
NG-Time                             0.4465      0.039     11.552      0.000       0.371       0.522
AMBI.TEMP                          -0.0647      0.018     -3.666      0.000      -0.099      -0.030
==============================================================================
Omnibus:                        1.445   Durbin-Watson:                   1.143
Prob(Omnibus):                  0.486   Jarque-Bera (JB):                1.489
Skew:                           0.057   Prob(JB):                        0.475
Kurtosis:                       2.854   Cond. No.                         524.
==============================================================================
```

Figure 11. Regression model output results for new growth discharge temperature.

3.1.3 Existing Growth Model

The existing growth process is the final stage of precipitation where productivity is defined by residence time through the circuit and the discharge temperature from the new growth process. A total of 16 parameters from the existing growth circuit were analysed using a heat map as shown in Figure 12. Of these, 4 parameters namely MPS-AGPL, MPS-CGPL, EG-Time were considered in the model. The prediction of existing growth productivity is shown in Figure 13. The MAE and RMSE of the model are 1.1288 and 1.5989 respectively.

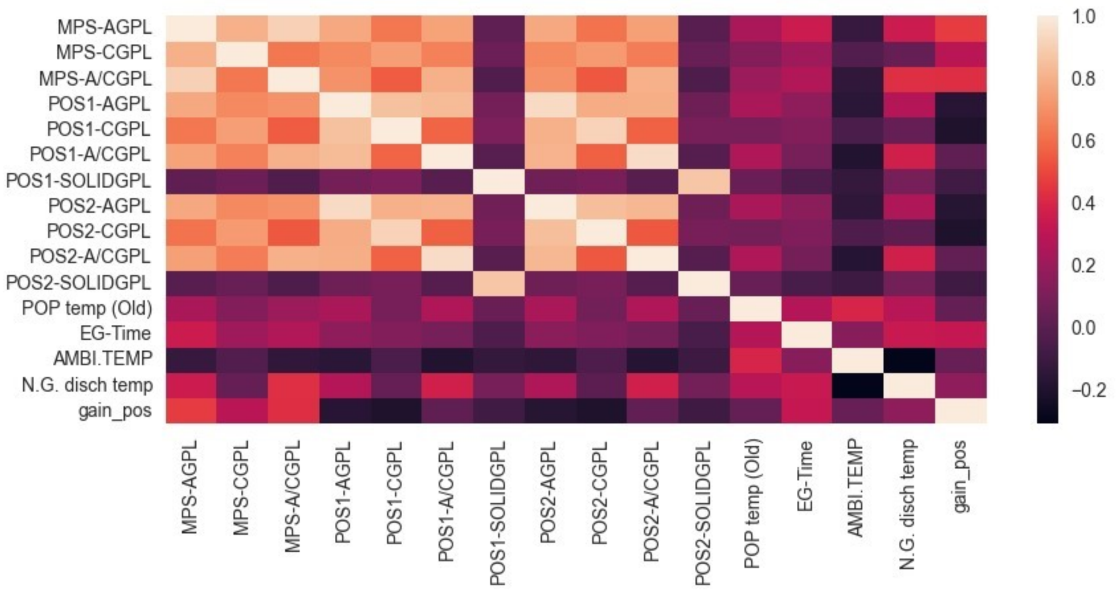

Figure 12. Heat map of variables for Existing growth discharge temperature.

```
                              OLS Regression Results
=======================================================================================
Dep. Variable:               gain_pos   R-squared (uncentered):                  0.971
Model:                            OLS   Adj. R-squared (uncentered):             0.971
Method:                 Least Squares   F-statistic:                         1.691e+04
Date:                Tue, 31 Oct 2023   Prob (F-statistic):                       0.00
Time:                        08:59:18   Log-Likelihood:                        -4165.2
No. Observations:                2026   AIC:                                     8338.
Df Residuals:                    2022   BIC:                                     8361.
Df Model:                           4
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
MPS-AGPL            0.3284      0.019     16.883      0.000       0.290       0.367
MPS-CGPL           -0.0599      0.008     -7.430      0.000      -0.076      -0.044
EG-Time             0.1139      0.011     10.141      0.000       0.092       0.136
N.G. disch temp    -0.1341      0.020     -6.647      0.000      -0.174      -0.095
==============================================================================
Omnibus:                        4.815   Durbin-Watson:                   1.349
Prob(Omnibus):                  0.090   Jarque-Bera (JB):                4.285
Skew:                           0.051   Prob(JB):                        0.117
Kurtosis:                       2.799   Cond. No.                         139.
==============================================================================
```

Figure 13. Regression model output results for Existing growth discharge temperature.

3.2 Constraints-Based Optimization Model

The constraints-based optimization model (CBM) is a computationally low-cost approach which is used to analyse complex systems. The CBM mathematical model is developed on known constraints and relationships between the independent variables. In the present case, the output obtained from ML model is used in CBM model for optimizing the precipitation circuit. Integrating ML model with CBM has advantages such as reduced bias in ML model during auto training, hence improving the accuracy of the model [5].

3.2.1 Agglomeration Model

The predicted agglomeration productivity is now fed to a local optimization model with given constraints to optimize the process. Constraint model for alumina soda has been developed and integrated with agglomeration productivity to deliver quality with respect to soda. Soda is an important factor in product alumina, derived from equilibrium alumina, PGL caustic, and temperature across the agglomeration process. Figure 14 shows the output of regression model with the MAE and RMSE for the predicted values of 0.0176 and 0.0246 respectively. The model's accuracy is high enough to meet statistical requirements.

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:                   NA2O   R-squared (uncentered):                   0.992
Model:                            OLS   Adj. R-squared (uncentered):              0.992
Method:                 Least Squares   F-statistic:                          2.962e+05
Date:                Mon, 11 Sep 2023   Prob (F-statistic):                        0.00
Time:                        09:11:41   Log-Likelihood:                          5319.3
No. Observations:                2327   AIC:                                 -1.064e+04
Df Residuals:                    2326   BIC:                                 -1.063e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Eq.Alumina     0.0035   6.39e-06    544.262      0.000       0.003       0.003
==============================================================================
Omnibus:                      813.953   Durbin-Watson:                   0.504
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            19185.459
Skew:                          -1.098   Prob(JB):                         0.00
Kurtosis:                      16.894   Cond. No.                         1.00
==============================================================================
```

Figure 14. Regression model output results for soda model.

3.2.2 New Growth Model

The quality of the new growth process is defined by the discharge of slurry, which is highly dependent on interstage cooler input flow, outlet temperature, and surrounding ambient temperature. Constraint model for new growth exit temperature has been developed and integrated with new growth productivity model to safeguard the process during winter when occasional vanadium contamination in alumina is observed. The comparison of predicted with actual values of new growth discharge time is shown in Figure 16. The MAE and RMSE for the predicted values are 0.8719 and 1.1480 respectively.

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:        N.G. disch temp   R-squared (uncentered):              0.999
Model:                            OLS   Adj. R-squared (uncentered):         0.999
Method:                 Least Squares   F-statistic:                     5.801e+05
Date:                Tue, 17 Oct 2023   Prob (F-statistic):                   0.00
Time:                        10:36:46   Log-Likelihood:                    -3305.0
No. Observations:                1606   AIC:                                 6616.
Df Residuals:                    1603   BIC:                                 6632.
Df Model:                           3
Covariance Type:            nonrobust
=======================================================================================================
                                     coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------
ISC Operating - Slurry Flow(m3/hr)   -0.0023      0.001     -4.039      0.000      -0.003      -0.001
ISC Operating - Slurry Outlet Temp    0.9800      0.008    125.747      0.000       0.965       0.995
AMBI.TEMP                             0.1276      0.010     12.727      0.000       0.108       0.147
==============================================================================
Omnibus:                      135.560   Durbin-Watson:                   0.782
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              238.271
Skew:                           0.590   Prob(JB):                     1.82e-52
Kurtosis:                       4.472   Cond. No.                         174.
==============================================================================
```

Figure 15. Regression model output results for new growth discharge temperature.

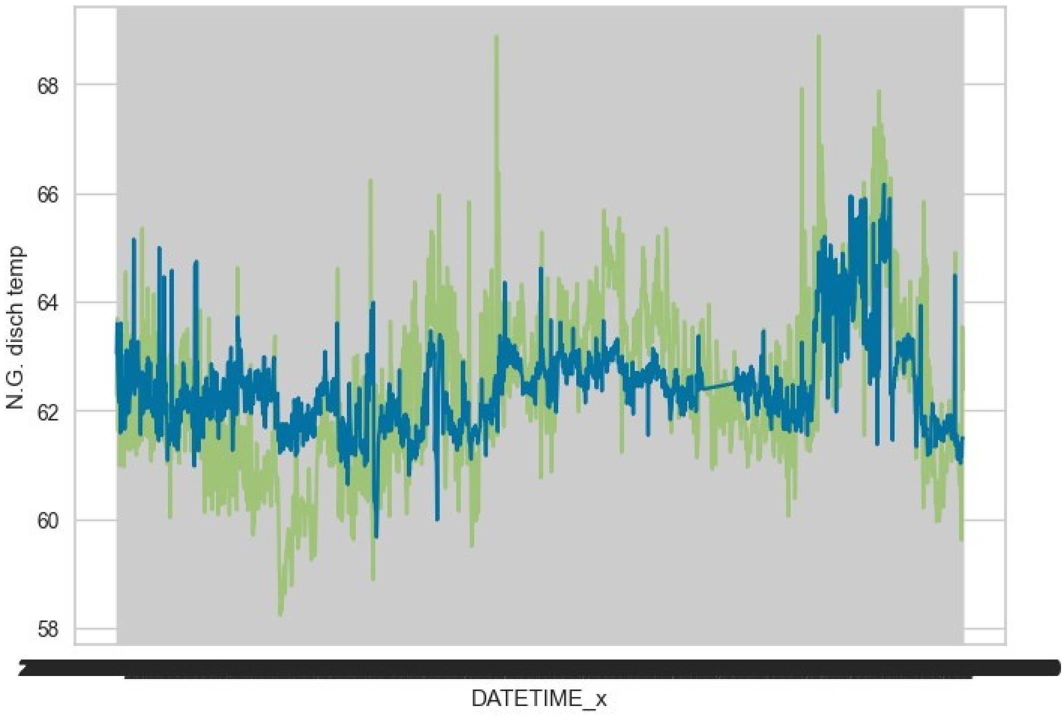

Figure 16. Predicted values of new growth discharge temperature.

3.2.3 Existing Growth Model

The existing growth process is crucial to maintain constraints across the existing growth process to prevent vanadium deposition, which is managed by existing growth residence time, new growth discharge temperature, and ambient temperature. A constraint model for the existing growth exit temperature has been developed and integrated with existing growth productivity model to safeguard process during winter when occasional vanadium contamination in alumina and rare oxalate issues are observed sometimes. The MAE and RMSE for the predicted values are 0.5664 and 0.7365 respectively for the predicted values, as shown in Figure 18.

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:          POP temp (Old)   R-squared (uncentered):                  1.000
Model:                             OLS   Adj. R-squared (uncentered):             1.000
Method:                  Least Squares   F-statistic:                         1.630e+06
Date:                 Tue, 31 Oct 2023   Prob (F-statistic):                       0.00
Time:                         09:32:21   Log-Likelihood:                        -3116.0
No. Observations:                 2026   AIC:                                     6238.
Df Residuals:                     2023   BIC:                                     6255.
Df Model:                            3
Covariance Type:             nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
EG-Time            -0.0329      0.007     -4.983      0.000      -0.046      -0.020
AMBI.TEMP           0.2209      0.004     52.664      0.000       0.213       0.229
N.G. disch temp     0.7953      0.003    289.178      0.000       0.790       0.801
==============================================================================
Omnibus:                       49.154   Durbin-Watson:                   0.962
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               52.458
Skew:                          -0.374   Prob(JB):                     4.06e-12
Kurtosis:                       3.249   Cond. No.                         20.0
==============================================================================
```

Figure 17. Regression model output results for existing growth discharge temperature.

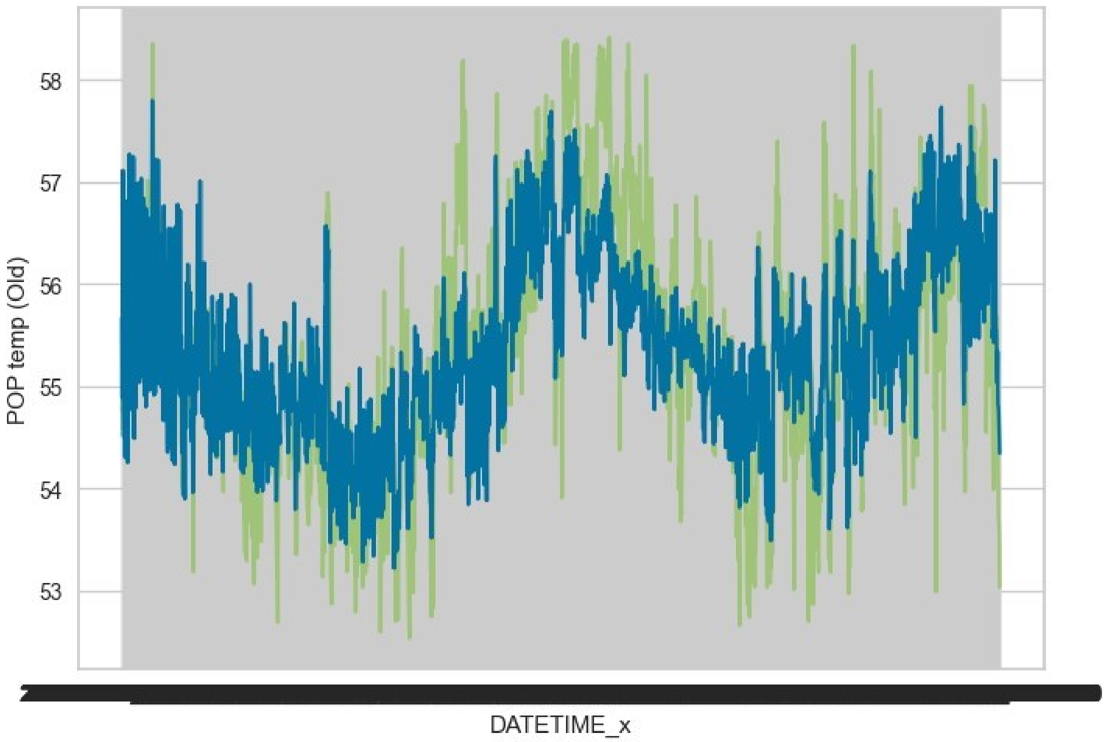

Figure 18. Predicted values of existing growth discharge temperature.

3.3 Integration of Local Models

After creating individual models for productivity and constraints, all models were integrated into a single global model using a Python-based framework, specifically Python Optimization Modeling Objects (PYOMO). This provides a robust foundation for developing and applying models, as the platform is well tested in wide variety of contexts [6]. It also consists of rich set of libraries which can support integrating various models. The three models of agglomeration, new growth and existing growth tank models were integrated to obtain a single optimization model for precipitation.

3.4 Cloud Computing for Model Lifecycle Management (Architecture)

For model lifecycle management and governance, Oracle cloud computing technology is leveraged to deploy the model and sustain the solutions. Figure 19 breaks down the cloud life cycle into four phases. Every stage paves the way for the one after it, therefore following the order is crucial to the success of the process. The risk involved with cloud projects is decreased by using this stepwise strategy. The services used include:

1. **ODI (Oracle Data Integrator)**: To import data from source to the IT-side database.
2. **ADW (Oracle Autonomous Data Warehouse)**: To store data from DCS and laboratory servers.
3. **Oracle Data Science**: To create machine learning models and host them.
4. **Oracle Virtual Machine**: To create and provide solutions to end-users via a full-stack dashboard.

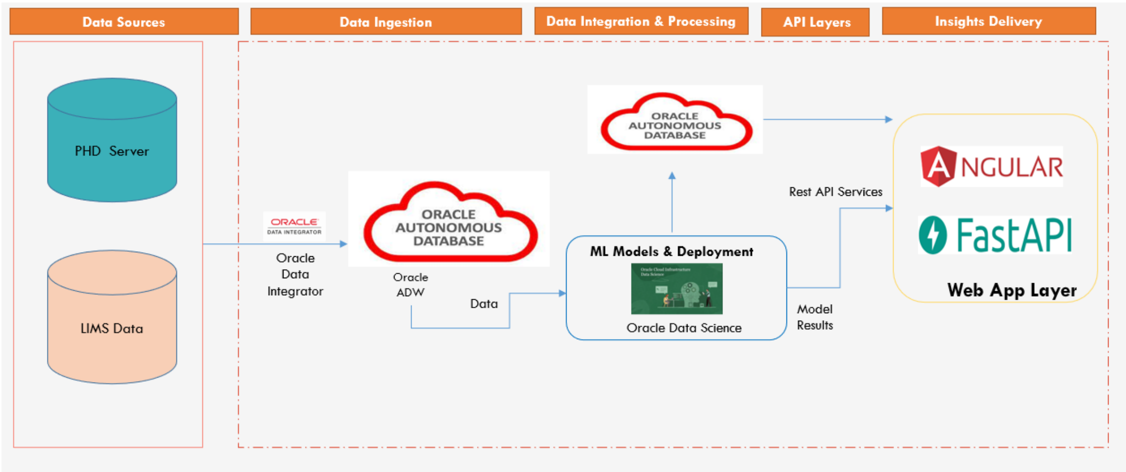

Figure 19. Architecture for cloud computing.

4. Conclusion

The development and deployment of an intelligence-driven precipitation control model using ML has proven to be a significant advancement in optimizing alumina productivity while maintaining consistent quality. By partitioning the precipitation process into agglomeration, new growth, and existing growth segments, and integrating constraints within each segment, our model ensures a balance between productivity and quality. This data-driven approach not only enhances operational excellence but also highlights the potential of AI-driven solutions in alumina refining. The use of cloud computing for model lifecycle management further ensures the robustness and scalability of the solution, paving the way for future innovations in the industry.

5. References

1. Den Hond et al., Alumina yield in the Bayer process past, present and prospects, Essential Readings, *Light Metals*, 2016, Volume 1 Alumina and Bauxite, 528-533.
2. Monique Delaney, Optimization of the Control and Monitoring Strategies for Precipitation Yield and Particle Sizing using Predictive Modelling, *Proceedings of 40th International ICSOBA Conference, Athens,* 2022, pp. 10-14.
3. Emily Sullivan, Understanding from machine learning models, *The British Journal for the Philosophy of Science*, 2022.

4. Michael Carter, Camille C. Price and Ghaith Rabadi, Operations research: a practical introduction, *Chapman and Hall/CRC*, 2018.
5. Prati Rana et al, Recent advances on constraint-based models by integrating machine learning, *Current opinion in biotechnology*, 2020, vol.64, 85-91.
6. William E. Hart et al, *Pyomo – optimization modeling in python*, Springer Textbook, 2017 Edition.