Agglomeration Process Optimization

Jefferson Klister[1], Cleto Azevedo[2], Taynara Valentim[3], Pedro Costa[4], Alok Batra[5], Bhanu Prakash[6] and Narayana Khadri[7]

1. System Analyst
2. Process Consultant
3. Process Engineer
4. System Specialist

Hydro Bauxite & Alumina, Belém, Brazil

5. President & CTO
6. Technical Lead
7. Data Scientist

Atomiton, Santa Clara, USA

Corresponding author: jefferson.klister@hydro.com
https://doi.org/10.71659/icsoba2024-aa017

Abstract

This paper presents a pioneering approach to optimizing the agglomeration process at Hydro's Alunorte alumina refinery through the implementation of a digital twin. This advanced system leverages first principles and data science techniques to provide real-time forecasts and recommendations for critical quality variables, enabling a shift from reactive to proactive process management. The core advancements of this project include the development of precise predictive models, a what-if analysis tool for scenario simulation, and a recommendation tool for operational adjustments. The anticipated benefits of this technology include enhanced product quality, reduced variability, and increased operational efficiency, setting new benchmarks in the alumina refining industry.

Keywords: Agglomeration, Throughput increase, Digital Twin, Predictive models.

1. Introduction

The alumina agglomeration process in the context of hydrate precipitation in the alumina industry is a fundamental step aimed at optimizing the quality and efficiency of the final product. During alumina production, hydrate precipitation involves the formation of alumina hydrate crystals from a sodium aluminate solution. At this stage, agglomeration plays a crucial role by transforming fine alumina hydrate particles into larger, more manageable aggregates. Agglomeration occurs through a series of strategic steps, beginning with the preparation of fine hydrate particles. This preparation includes the classification of the seed, ensuring that the particles are within the appropriate size range for agglomeration. This is followed by a mixing and homogenization phase, where the particles are evenly dispersed, facilitating the formation of consistent aggregates. In the agglomerators, the agglomeration of hydrate particles occurs under carefully controlled conditions. Temperature is a critical factor, as it influences the solubility of the hydrate and the rate of crystal growth. Maintaining the temperature within an optimal range ensures efficient agglomeration, promoting the formation of robust aggregates.

The seed charge, or the amount of fine particles introduced into the agglomerator, is also fundamental to the process. An adequate seed charge ensures that there are enough nucleation points for the formation of aggregates of the desired size. The balance between the amount of seed and the sodium aluminate solution must be maintained to optimize the efficiency of agglomeration.

Residence time, or the duration that the particles remain in the agglomerator, is another essential factor. This time must be long enough to allow the fine particles to fully agglomerate, but not so long as to cause excessive growth of the aggregates. Precise control of the residence time ensures the uniformity of particle size in the final product.

Agglomeration plays a pivotal role in improving the physical handling characteristics of alumina, thus facilitating subsequent processing stages and minimizing dust-related issues. However, the complexity of this process arises from several critical and interdependent factors that require meticulous control to ensure uniform and consistent results. These factors include the control of nucleation – the formation of initial particle clusters; the seeding strategy, which involves introducing "seeds" to promote uniform particle growth; and the control of crystal growth to achieve the desired aggregate size and structure. Additionally, the direct manipulation of the agglomeration process – coalescing smaller particles into larger ones – and ensuring reproducibility and consistency across batches present significant operational challenges. To address these complexities, the refinery has integrated advanced data science techniques and predictive tools. These technologies leverage first principles and empirical data to enhance yield, improve quality, and reduce process variability. The adoption of such innovative approaches not only optimizes the agglomeration process but also sets new benchmarks in operational efficiency and product quality in the alumina refining industry.

The importance of agglomeration in the quality control of alumina cannot be overstated. This process is essential for improving the physical properties of the product, such as density and mechanical strength, as well as increasing reactivity and ease of handling. Producing alumina with particles of adequate size and shape is vital for subsequent processing stages, including calcination and electrolytic reduction, where the efficiency and uniformity of the material directly influence the productivity and quality of the aluminium produced. Therefore, precise control of the agglomeration process during hydrate precipitation is crucial to ensure that the final alumina meets the stringent quality standards required by the industry. This control encompasses everything from the proper classification of the seeds to the optimization of temperature conditions, seed charge, and residence time in the agglomerators, ensuring a high-quality and high-performance final product. Below is being represented the simplified Bayer process and the specific precipitation area with the agglomeration process (Figures 1 and 2).

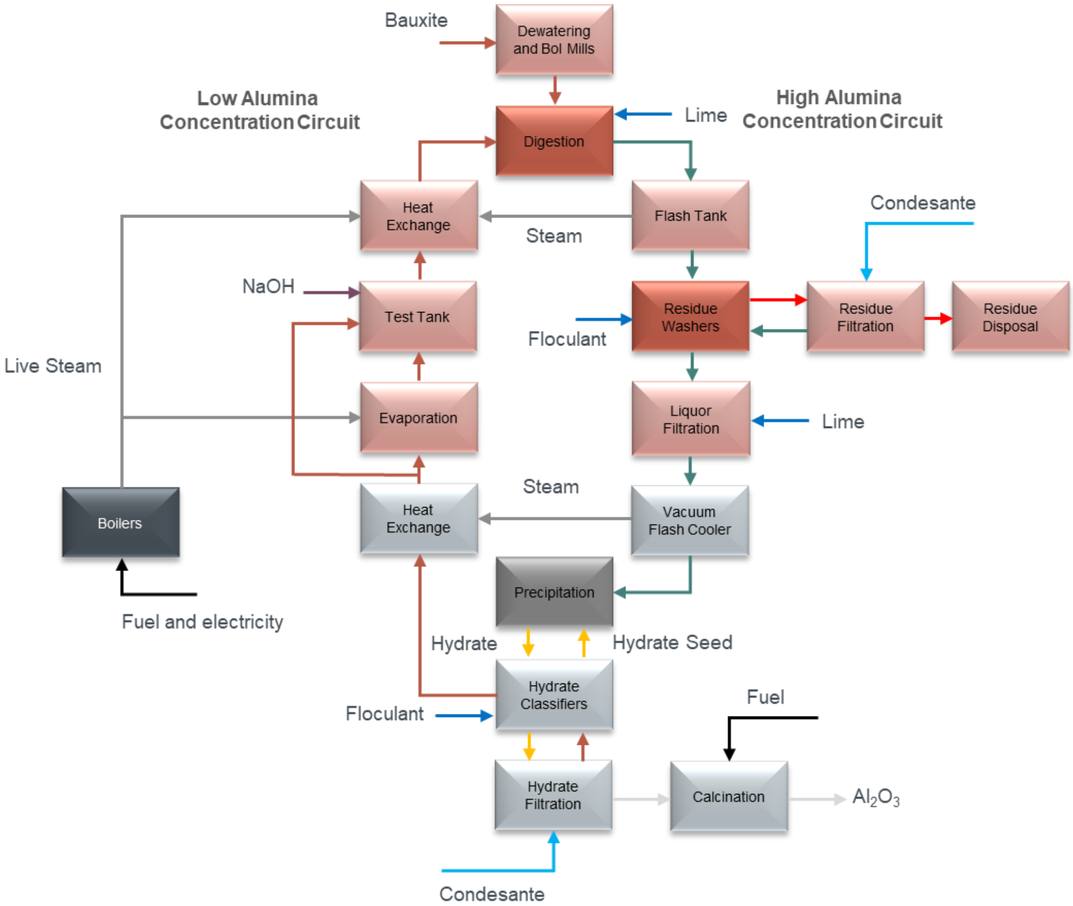

Figure 1. Simplified Alumina Bayer Process.

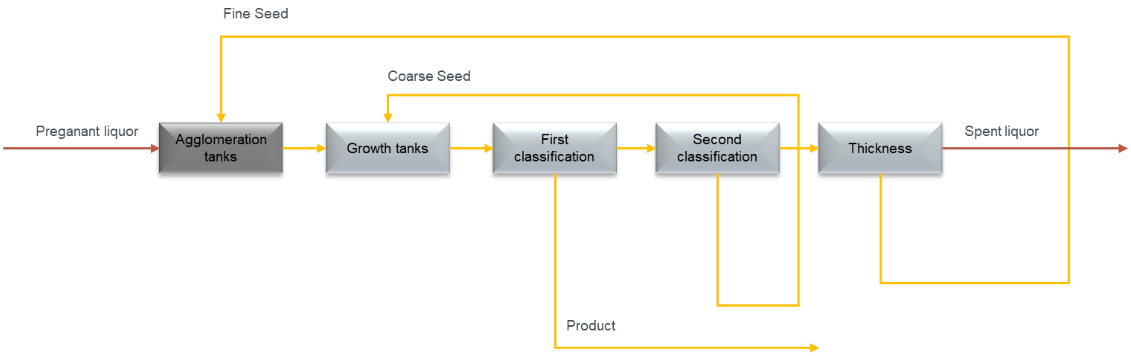

Figure 2. Simplified Precipitation process.

2. Technology Used in the Agglomeration Process Optimization

Optimizing the agglomeration process at the alumina refinery incorporates a blend of first principles and advanced data science techniques to improve yield and process efficiency. The methodology encompasses several critical technologies:

2.1 Agglomeration Model

The model is a fusion of classical engineering principles with modern data science techniques. This innovative approach leverages predictive analytics and advanced modelling to not only streamline but also revolutionize the agglomeration process. The model is fundamentally based on first principles, particularly focusing on mass balance mechanisms, which are essential for tracking and predicting the behaviour of key variables from the input to the output stages of agglomeration, as Figure 3 shows.

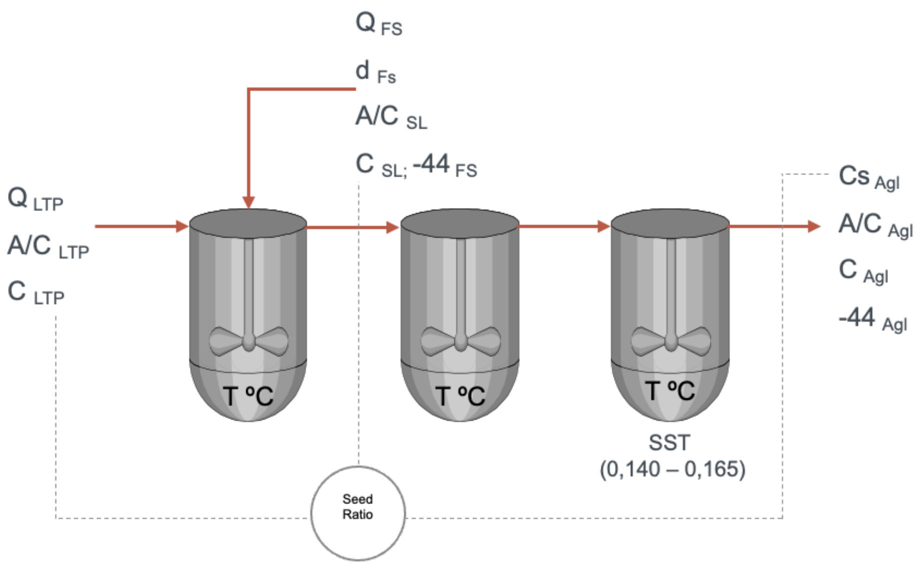

Figure 3. Agglomeration process variables.

At the input stage, which is feeding the first tank on Figure 3, variables (Table 1) such as the liquor flow rate, alumina concentration, and caustic concentration are closely monitored. The precise control of these variables is critical as they directly influence the chemistry and kinetics of the agglomeration reactions, setting the stage for the entire process.

During the seeding stage, the management of seed characteristics becomes paramount. Parameters like specific surface area, particle size, solid concentration, seed density, and seed flow rate are finely controlled. These parameters are vital as they determine the growth dynamics of the agglomerates, ultimately influencing the quality of the output.

The output stage involves a thorough analysis of the 'pregnant liquor', assessing factors such as alumina and caustic concentrations, as well as solid concentration. This analysis is crucial as it provides insights into the efficiency of the agglomeration process and the quality of the final product.

Table 1. Description of key variables on agglomeration process.

Source	Measure	Description
Green Liquor (LTP)	Q_{LTP}	Liquor to precipitator flow → Controls the volume and rate at which liquor is introduced into the agglomeration process.
	A/C_{LTP}	Alumina/Caustic ratio of LTP → Essential for maintaining the chemical balance necessary for effective agglomeration.
	C_{LTP}	Caustic concentration of LTP → Influences the precipitation and stability of alumina particles.
Fine Seed (FS)	Q_{FS}	Fine seed flow → Determines the amount of seed material entering the process, affecting the growth dynamics of the alumina particles.
	d_{FS}	Solid density of fine seed → Critical for ensuring the proper sedimentation and incorporation of seeds into the alumina slurry.

Source	Measure	Description
	A/C_{SL}	Alumina/Caustic ratio of spent liquor
	C_{SL}	Caustic concentration of spent liquor
	-44_{FS}	-44 microns of fine seed
Pregnant Liquor (Agl)	Cs_{Agl}	Solids concentration at 3rd agglomeration tank → Indicates the concentration of solids in the output, a direct measure of process efficiency and product quality.
	A/C_{Agl}	Alumina/Caustic concentration at 3rd agglomeration tank
	C_{Agl}	Caustic concentration at 3rd agglomeration tank
	-44_{Agl}	-44 microns at 3rd agglomeration tank → Tracks the size distribution of alumina particles in the pregnant liquor, a key quality attribute

The model uses the mass balance equation to track changes in alumina mass across the process, incorporating volume and concentration changes in the tanks, for a time from t1 to t2:

$$\Delta AI(mass) = (Vt2 * Ct2 - Vt1 * Ct1) = Inflow + Seeding + Production - Outflow \quad (1)$$

Where:

V	Liquor volume in the tank, L
C	Is the concentration of AI, %
Inflow	Liquor Al_2O_3 concentration * liquor flowrate, m^3/h
Seeding	Seed Al_2O_3 concentration * seed flowrate, m^3/h
Outflow	Agglo Al_2O_3 concentration * (sum of flowrates of liquor and seed), m^3/h

The residence time — the duration that the liquor remains in a tank — is calculated by dividing the tank volume by the flow rate. This measure is crucial for predicting reaction kinetics and ensuring effective agglomeration. The system operates under steady-state conditions, with tanks maintained full to simplify dynamics while accommodating input variability. This setup necessitates precise seeding control to produce consistently high-quality pregnant liquor. The model's accuracy is validated against laboratory results at four to eight-hour intervals, confirming its operational efficacy.

2.2 Data Processing and Model Integration

The data processing framework designed for the agglomeration process is robust, capable of managing, cleaning, and analysing vast amounts of data generated during the agglomeration process. This framework is crucial for supporting the predictive models. Continuous data collection from agglomeration tanks, seeding operations, and laboratory analyses ensures a rich dataset. Initial data cleaning addresses discrepancies such as missing values and outliers, aligning and consolidating data to ensure accuracy and usability.

To manage the sparsity of data and maintain computational efficiency, data is down sampled to a fifteen-minute interval. Techniques such as linear interpolation and more sophisticated methods like spline regression are used for missing data imputation.

2.3 Process Twin via Machine Learning

The process twin uses advanced machine learning algorithms to simulate and predict key process outcomes. This technological facet is integral for dynamic process management. Specific models were developed for predicting Solids Concentration (CSagglo), Agglomeration Index (AI), and Supersaturation (SST). These models are crucial for real-time process control and optimization, as they update its prediction on hourly basis. Allowing closer monitoring on the process response over the last process adjustments. Also, extensive training on 2 years of historical data ensures that the models are robust, with validation processes in place to minimize errors and optimize performance metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

2.4 Analytics and Prediction

As a segment of the technology framework, operational decision support leverages outputs from the machine learning models to facilitate proactive decision-making in the agglomeration process. Thus, continuous predictions of key parameters such as solid concentration and particle size distribution allow for real-time adjustments. An automated feedback loop integrates these predictions with operational controls to dynamically adjust process variables as required. An intuitive dashboard displays real-time data and predictions, seen in Figure 4, aiding operators in identifying trends, making informed decisions, and optimizing operational parameters on 2 operational lines of the refinery.

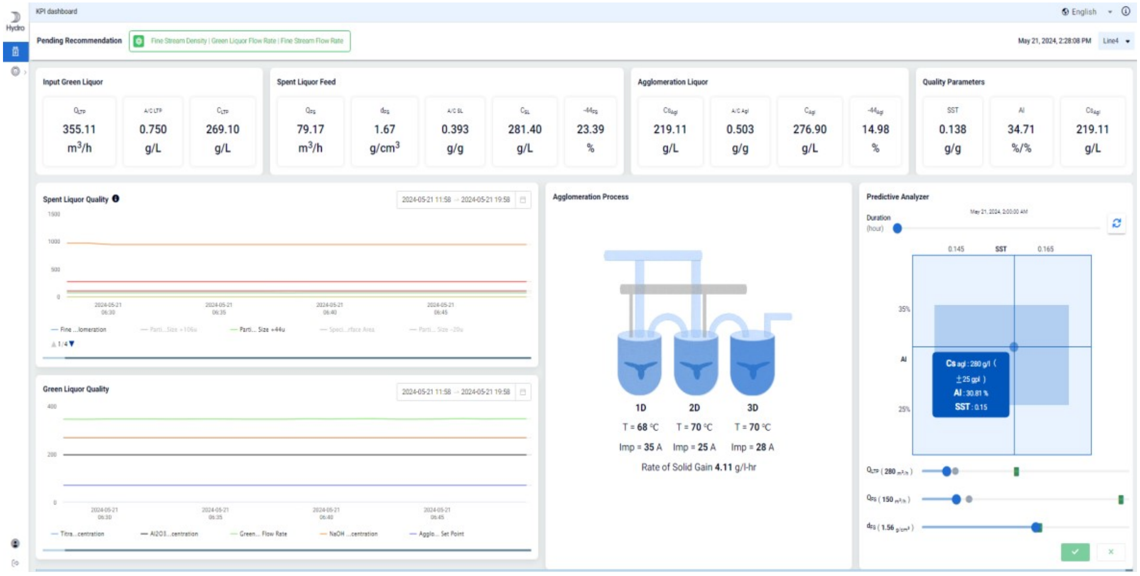

Figure 4. Agglomeration process dashboard.

3. Process Simulator

The Process Simulator at Norsk Hydro's Alunorte refinery, seen in Figure 5, represents a cornerstone of technological advancement in the agglomeration process optimization. This sophisticated tool models the dynamic interactions of various process variables in real time, utilizing both historical and current data to forecast future process conditions and outcomes. Its predictive capability is pivotal for enabling a proactive approach to process management, significantly enhancing decision-making and strategic planning capabilities.

Figure 5. Agglomeration process simulator feature.

As functionalities of the process simulator, it integrates comprehensive data streams from multiple sources, including input green liquor quality, spent liquor feed, and agglomeration liquor. This integration allows for meticulous process behaviour analysis of flow rates, alumina, caustic concentrations, and particle size distributions from these inputs, ensuring a holistic view of the process dynamics.

Another powerful feature of the process simulator is its ability to allow operators to simulate different operational scenarios. By adjusting input parameters such as flow rates and seed density, operators can explore their effects on critical outputs like solid concentration, supersaturation levels (SST), and agglomeration index (AI). This feature not only facilitates a deeper understanding of process sensitivities but also aids in identifying potential optimizations. Through advanced algorithms, the simulator predicts the performance of the agglomeration process under varied conditions. These predictions, synthesized from a blend of real-time data and historical trends, are optimized for accuracy and relevance. The ability to forecast process outcomes enhances proactive process adjustments and strategic planning, thereby improving overall operational efficiency. Thus, the simulator possesses various usages on process optimization such as optimal parameter identification, where the simulator plays a crucial role in identifying the most effective operational parameters. By simulating various settings and their impacts, it pinpoints configurations that maximize efficiency and product quality. This capability supports continuous improvement initiatives and helps in achieving optimal operational conditions.

Also serves as a vital tool in yield management. By providing early warnings based on simulated outcomes, it allows for pre-emptive measures to mitigate potential issues before they materialize. Operator can take one or more of the following actions

1) Change input inflow liquor flow rate
2) Change Seed flow rate
3) Change fine seed density for the incoming seed stream

This proactive approach significantly reduces the likelihood of out of range quality of the outflow liquor and enhances the reliability of the agglomeration process.

Furthermore, it can be an educational tool, acting as an invaluable educational platform for training new operators. It enables them to understand the complex dynamics of the agglomeration process in a risk-free environment. This educational aspect is particularly beneficial for enhancing operator understanding and capability, allowing them to experiment with process changes and see the potential effects without the risks associated with real-world trials. The process simulator is closely integrated with the refinery's digital twin technology, forming a feedback loop that continuously updates and refines process models based on real-time data and outcomes. This integration ensures that the digital twin remains accurate and effective in simulating and predicting process behaviours, further enhancing its utility in operational decision-making.

4. Recommendation tool

The recommendation tool (Figure 6) is an advanced predictive system designed to optimize the alumina agglomeration process. Leveraging the insights derived from the process simulator, this tool utilizes machine learning algorithms to analyse process data in real time and provide precise operational adjustments. It enables actionable recommendations that can be directly implemented into the production workflow, thus enhancing both efficiency and product quality.

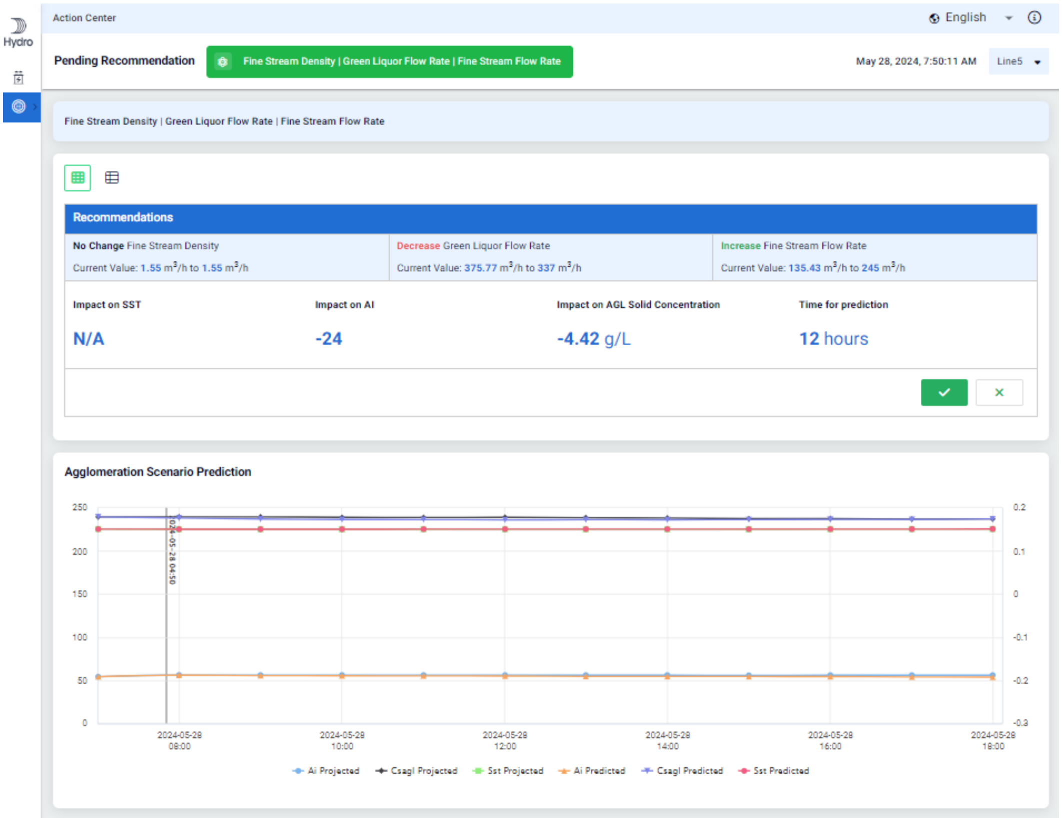

Figure 6. Agglomeration recommendation dashboard.

The backbone of the recommendation tool is its ability to continuously analyse both real-time and historical process data. By predicting the outcomes of potential adjustments, the tool automatically suggests optimal changes to operational parameters such as green liquor flow rate, fine stream density, and fine stream flow rate. These recommendations aim to optimize the process conditions and enhance overall performance.

For each recommended modification, the tool provides a detailed analysis of the expected impacts on key quality parameters, including solid concentration of pregnant liquor, supersaturation levels, and the agglomeration index. This feature allows operators to assess the potential benefits and understand the trade-offs involved with each recommended action. It facilitates data-driven decision-making and ensures that all adjustments are made with a clear understanding of their likely outcomes.

The recommendation tool features a user-friendly interactive dashboard that displays real-time data and model predictions in an intuitive format. Equipped with visual aids such as graphs and charts, the dashboard enables operators to easily adjust input parameters and visualize the effects of these adjustments on predicted outcomes.

5. Agglomeration Optimization

The integration of digital twin technology represents a transformative approach to managing the alumina agglomeration process. By leveraging real-time data and predictive analytics, this technology provides a robust framework for pre-emptively adjusting process parameters to optimize the quality of the output, particularly the pregnant liquor. This section delves into the nuances of the variables, process data, and predictive results that highlight significant opportunities for process optimization.

The digital twin system utilizes a first-principle material balance model to dynamically simulate the agglomeration process. By integrating real-time measurements with historical data, the system can forecast essential outputs, enabling proactive adjustments to the process. The model training includes derived features such as the Concentration Ratios, which calculate the ratio of Al_2O_3 to NaOH concentrations—a critical factor in predicting tank conditions. Additionally, the Solid Concentration Gain feature aids in anticipating changes in the solid concentration, providing valuable insights into the quality evolution of the agglomerated product.

The system's real-time predictive capabilities facilitate the optimization of process parameters, thus ensuring the consistent quality of the pregnant liquor. It generates predictions for vital parameters like solid concentration, particle size increase, and chemical supersaturation. The real-time predictive feedback includes hourly gain predictions, where the system computes the hourly gain for solid concentration, agglomeration index, and supersaturation. These predictions are used to adjust the process dynamically, offering a solution to the limitations imposed by delayed laboratory results, which traditionally hinder timely process modifications.

5.1 Data Challenges

Data are sourced from various origins, such as real-time sensor readings and laboratory test results. These data points are gathered at irregular intervals, lacking a consistent timeline. Consequently, the input and output variables are not continuously available, resulting in sparse datasets. This irregularity poses several significant challenges, including:

1. **Intermittent data collection**: Lab results, crucial for monitoring output variables, are obtained only a maximum of two times per day. This sparse data collection complicates any continuous process analysis and real-time predictive modelling.
2. **Lag in output variable availability**: Output variables critical for process adjustments are only available coinciding with the infrequent lab results, limiting real-time optimization capabilities.
3. **Irregular input monitoring**: Essential input variables that significantly impact the process outcomes are not recorded at regular intervals, presenting significant hurdles in modelling and prediction accuracy.
4. **Delayed impact observation**: The effects of inputs like green liquor and fine seed on the agglomeration process are only observable after several hours as the materials progress through the system to tank 3, introducing a delayed response challenge in real-time monitoring.

5. **Complex dependencies**: Output variables depend on the integrated history of state changes over several hours rather than being directly correlated with real-time or specific past input values.
6. **Time-lagged effect analysis**: Any operational adjustments can only be evaluated after a delay, as the outputs reflect the accumulative effect of past inputs and operations over time.
7. **Uncertain future inputs**: Accurate predictions for future hours are challenging as input variables like fine seed and green liquor, which significantly impact outputs, can vary unpredictably.

5.2 Approach to solutions

Through 2 years of past data variables seen in Table 2, the challenges outlined above were addressed. It has been devised a comprehensive approach involving several key strategies:

1. **Modelling residence time**: a simplified plug flow model was implemented to track the residence time of materials in the three-tank system. This model helps determine the historical input data required at the time of output prediction, acknowledging the delayed effect of inputs on outputs.
2. **Data imputation**: To bridge the data gaps, linear interpolation is utilized between available data points, assuming minimal variability between consecutive lab results. This approach, while simplistic, provides a continuous dataset for more stable model training.
3. **Feature engineering and data preparation**:
   - **Time-averaged inputs**: Averages of input variables are computed over the calculated residence times to better represent their effect on the final output.
   - **Utilization of latest lab results**: Incorporating the most recent lab results allows the model to account for the current state of the process, which is critical for predicting immediate future states.
   - **Derived data metrics**: Additional features are derived using first principles and mass balance equations to more comprehensively represent the process dynamics at both input and output stages.
   - **Data sampling and aggregation**: The dataset is standardized to hourly intervals. For granular data, the values are averaged; for missing data, linear interpolation is employed.
   - **Machine learning model training**: Separate models are developed for each key output parameter— solids concentration, supersaturation, agglomeration index— using engineered features. The models are trained to minimize Mean Absolute Percentage Error (MAPE), and the best-performing hyperparameters are selected for each model.
   - **Operational Integration**: The models are integrated into the operational framework, allowing for real-time prediction based on the last available input and lab data. This integration supports dynamic process management by predicting how current conditions will influence near-future outputs.

Table 2. All variables used in the predictive models.

Variable source	Description	Unit
Input	Flow to green liquor	m^3/h
	Al_2O_3 concentration at LTP input	g/L
	NaOH concentration at LTP input	g/L
	Flow to fine seed	m^3/h
	Density of fine seed	g/cm^3

Variable source	Description	Unit
	Particle Size +44 µm of fine seed	%
	Specific Surface Area (SSA) of fine seed	m^2/g
	PS Median of fine seed	µm
	Al_2O_3 concentration of fine seed	g/L
	NaOH concentration of fine seed	g/L
Output	Al_2O_3 concentration from Tank 3	g/L
	NaOH concentration from Tank 3	g/L
	Particle Size +44 µm from Tank 3	%
	SSA from Tank 3	m^2/g
	PS Median from Tank 3	µm
	Solid concentration from Tank 3	g/L
Derived	Alumina to caustic concentration at green liquor and fine seed	%
	Concentration gain from mass balance equation	kg
	SST at input and output	%
	Agglomeration Index	%

5.3 Output and Plots

When using multivariate regression models with values at the time of entry, the results were not promising, as shown in Table 3. These models failed to provide accurate predictions and were not suitable for use in production environments. Given the limitations of standard data science and machine learning techniques, a specialized algorithm was developed to address the unique challenges of the data, achieving better predictive performance and reliability in production scenarios.

Table 3. Initial MAPE for predicted KPIs.

KPI	MAPE
Agglomeration Index	65
Supersaturation	5.3e-05
Solids concentration	5.2

As agglomeration output data from Tank 3, used to calculate solid concentration, agglomeration index, and supersaturation are sparsely available, with a maximum of only two data points recorded per day, it results in only two actual KPI predicted values per day.

A scatter plot (Figure 7) with time on the x-axis and input points on the y-axis highlights the sparse data, showing around 128 data points collected over 80 days.

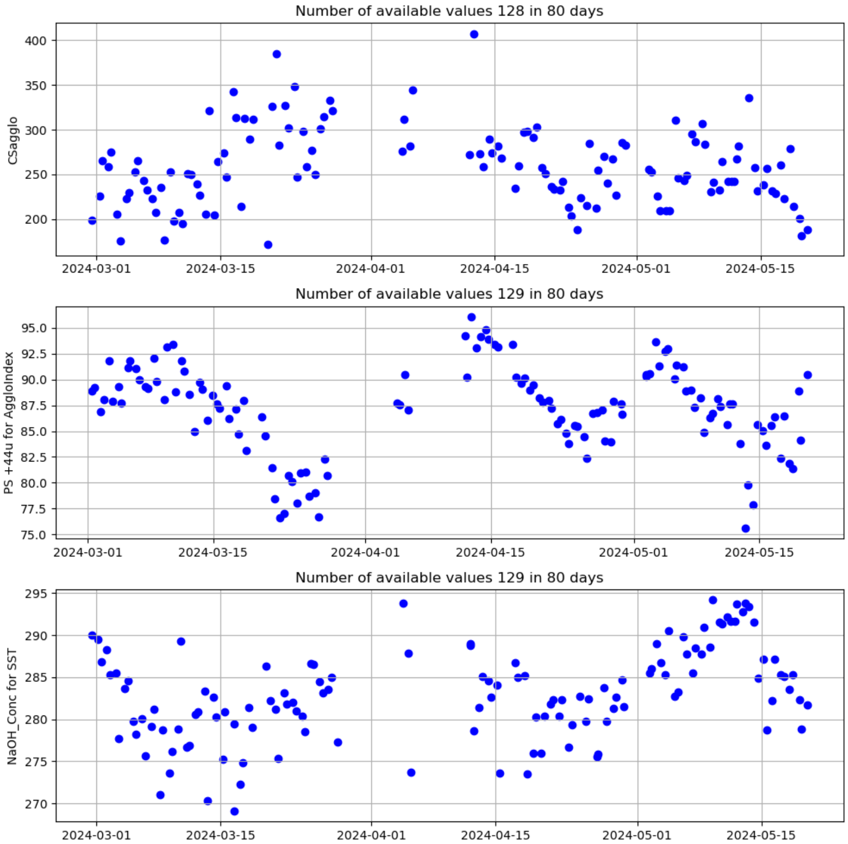

Figure 7. Sparse data points for target variables: solids concentration, agglomeration index and supersaturation, respectively.

5.3.1 Model Training

After successfully managing the data sparsity, a tree-based model was applied to predict the target prediction KPIs; solids concentration, supersaturation and agglomeration index.

Below (Figure 8) you see how the models were predicting against actual values from the process when plotting the data.

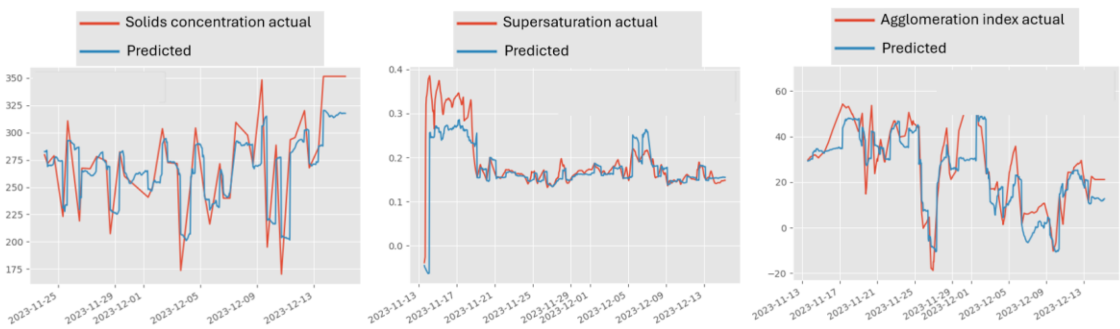

Figure 8. Prediction & Actual values from prediction output of all 3 predicted KPIs.

All results seen in Figure 8, were obtained through the best hyperparameter set for each KPI, as displayed in Table 4.

Table 4. Hyperparameter set and its results for each predicted KPI.

KPI Predicted	Hyperparameters	MAPE
Solids concentration	{'n_estimators': 282, 'max_features': 1.0, 'max_leaves': 12656}	12.26
Supersaturation	{'n_estimators': 130, 'max_features': 0.7766377519728662, 'max_leaves': 6750}	1.11e-05
Agglomeration Index	{'n_estimators': 356, 'max_features': 1.0, 'max_leaves': 6738}	1.86

These models are trained as per lines (specific combination of tanks, sizes, and other parameters. Hence the mechanism can be applied to any line by letting the model being trained for that specific configuration of the line.

6. Conclusion

The integration of digital twin technology into the agglomeration process at Alunorte's refinery represents a significant leap forward in the alumina industry. By employing advanced simulation and predictive analytics, the refinery is ready to shift from a reactive approach to a proactive and predictive management strategy, profoundly transforming process control and optimization.

Thus, the introduction of digital twin technology is poised to bring about significant improvements in the refinery's operations. Enhanced operational efficiency is a primary expected benefit, with the technology's real-time monitoring capabilities anticipated to allow for more effective adjustments in the agglomeration process. This should lead to improved throughput while maintaining quality standards. The predictive aspect of this technology is also projected to make operational processes more time and resource efficient.

In terms of product quality, there is an expectation of marked improvement in the consistency and quality of the alumina produced, thanks to the precise control over process variables that the digital twin technology offers. Additionally, the integration of predictive maintenance is likely to reduce operational downtime, thereby potentially enhancing productivity and the reliability of processes.

The environmental impact of the refinery's operations is also expected to be positively influenced by the digital twin technology. It is anticipated that the technology will facilitate more efficient resource use and waste reduction. The prospect of improved process control leading to decreased raw material consumption and emissions points towards a more sustainable operational model.
Overall, the digital twin technology is expected to be a catalyst for progress in achieving operational efficiency, product quality, and environmental sustainability. These anticipated developments signify a commitment to ongoing improvement and responsible industrial practices.

As per possible future developments, an addition of new KPIs like Mesh -365 and Occluded Soda to the digital twin setup is expected to give a clearer view of the process. These KPIs could provide more details that might help improve the agglomeration process, aiming for better quality and efficiency.

7. References

1. Masoud Bahrami Babaheidari et al., The agglomeration kinetics of aluminum hydroxide in Bayer process. *Powder Technology*, Iran, 12 March 2012, 224, 351-355.
2. D. Ilievski et al. Agglomeration during precipitation II - Mechanism deduction from tracer data. *Aiche*, March 1995, 525 -535.

3. D. Ilievski et al., Agglomeration during precipitation I - tracer crystals for $Al(OH)_3$ precipitation. Aiche, March 1995, 518 -524.
4. Foster Provost et al., Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, *O'Reilly Media*, July 2013.
5. Trevor Hastie et al., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. *Springer*. 2001.
6. Gareth James et al., *An introduction to statistical learning with applications in R*. Springer New York, 2013.
7. Ian Goodfellow et al., Deep Learning. *MIT Press*. November 2016.
8. Hal Varian, Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 2014, 3-28.
9. David Lazer et al., The parable of Google Flu: Traps in big data analysis. *Science, 2014,* 1203-1205.
10. Isabelle Seyssiecq et al., Modeling gibbsite agglomeration in a constant supersaturation crystallizer. *Chemical Engineering Science,2014, 55*, 5565-5578.