AI-POWERED DIGITAL TWINS AND CARBON FOOTPRINT MONITORING

STAMICARBON

NEXTCHEM
MAIRE Sustainable Technology Solutions

Conference name	Stamicarbon Symposium
Conference date	May 18-21, 2026
Author(s)	Martino Trabuio and Ali El Sibai
Classification	PUBLIC

SUMMARY

1. ABSTRACT 4
2. INTRODUCTION: DIGITAL TWINS AND AI IN THE CHEMICAL INDUSTRY 4
3. CARBON FOOTPRINT MONITOR 5
   3.1 THE RELEVANCE OF EMISSIONS MONITORING IN A SUSTAINABILITY-DRIVEN ERA........5
   3.2 SOFTWARE ARCHITECTURE AND UNDERLYING MODELS 5
      3.2.1 Software architecture 5
      3.2.2 Life Cycle Assessment (LCA) and emissions monitoring 6
      3.2.3 Leveraging the NX STAMI™ Digital suite..6
   3.3 VISUALIZING ENVIRONMENTAL PERFORMANCE..6
      3.3.1 Real-time emission tracking 6
      3.3.2 Historical trends and reporting 7
      3.3.3 Source attribution and root cause analysis..7
      3.3.4 Energy and utility monitoring..8
4. SOFT N/C Meter 8
   4.1 BACKGROUND: HARDWARE N/C METER 8
   4.2 SOFT N/C METER: MOTIVATION..9
   4.3 SOFT N/C METER: MODELING APPROACHES 9
      4.3.1 A Stamicarbon advantage: Combining plant and synthetic data for reliable ML-based Soft N/C Meter performance:..10
      4.3.2 ML-Based Soft N/C Meter for a real case study 10
      4.3.3 Model Validation 12
      4.3.4 Deployment Phase 13
5. AI-POWERED EVENT DETECTION 13
   5.1 BACKGROUND: ANOMALY DETECTION IN THE PROCESS INDUSTRY..13
   5.2 AI EVENT DETECTION: MODELING METHODOLOGIES 13
   5.3 MODEL DEVELOPMENT AND RESULTS..14
      5.3.1 Data gathering and model deployment 14
      5.3.2 Unsupervised Machine Learning: detecting deviations from normal operation 15
      5.3.2.1 Case Study 1: Rectifying Heater (E302) fouling..15
      5.3.2.2 Case study 2: CO_2 Slippage 16
      5.3.3 Supervised Machine Learning: targeted anomaly detection 17
      5.3.4 Insights for deployment in real plant operations 18

1. ABSTRACT

This paper presents how AI-powered digital twins can accelerate operational excellence and decarbonization in the fertilizer and chemical industries, with a focus on urea plants. Building on Stamicarbon's NX STAMI™ Digital Process Monitor – a high-fidelity digital twin based on first-principles thermodynamic and kinetic models – the paper describes the integration of advanced analytics and machine learning to extract additional value from plant DCS data. Three complementary solutions are detailed, including:

- **Carbon Footprint Monitor:** Applies real-time data acquisition, Life Cycle Assessment models, and standardized emission factors to calculate and visualize greenhouse-gas emissions per product and by source.
- **Soft N/C Meter:** Estimates the reactor outlet N/C ratio using machine learning models trained on selected process variables and enhanced via hybrid datasets that combine historical plant data with synthetic data generated by Technology Training Simulators.
- **AI Event Detection:** Leverages multivariate statistical process monitoring and supervised / unsupervised learning to identify early deviations from normal operation and support root-cause analysis.

Together, these tools enable real-time sustainability reporting, improved process efficiency, and proactive reliability monitoring, providing a scalable pathway for smarter and more resilient plant operations.

2. INTRODUCTION: DIGITAL TWINS AND AI IN THE CHEMICAL INDUSTRY

The chemical industry is undergoing a transformative shift driven by the convergence of digital technologies and advanced analytics. At the forefront of this evolution is the concept of the digital twin, a virtual replica of the physical plant that mirrors real-time operations through continuous data integration. In the chemical sector, digital twins are revolutionizing how plants are designed, operated, and maintained by enabling predictive insights, process optimization, and enhanced decision-making.

The NX STAMI™ Digital Process Monitor by Stamicarbon, the nitrogen technology licensor of NEXTCHEM (MAIRE Group), is a prime example of how digital twin technology is being applied in the chemical industry to enhance operational performance through real-time insights. In its initial deployment, the Process Monitor was exclusively based on first-principles modeling, utilizing rigorously validated thermodynamic and kinetic models developed by proprietary process technology to ensure high fidelity and predictive accuracy.

While the initial deployment of the NX STAMI™ Digital Process Monitor in Urea plants significantly enhanced operational efficiency and reduced steam consumption, it was soon recognized that integrating advanced AI technologies could play a pivotal role in realizing the full potential of this digital twin, unlock deeper optimization opportunities, and deliver additional value to the client.

By applying machine learning algorithms to vast streams of sensor data, AI can detect anomalies, predict equipment failures, and recommend optimal process adjustments in real time. This powerful synergy between digital twins and AI not only drives continuous improvement in efficiency and safety but also supports sustainability objectives by minimizing energy consumption and reducing waste.

As the fertilizer industry faces increasing pressure to innovate and decarbonize, the integration of digital twins and AI offers a compelling pathway to smarter, more agile, and resilient operations. Stamicarbon has developed several digital tools to support this transformation, including the Soft N/C Meter, AI Event Detection, and the Carbon Footprint Monitor.

3. CARBON FOOTPRINT MONITOR

3.1 THE RELEVANCE OF EMISSIONS MONITORING IN A SUSTAINABILITY-DRIVEN ERA

Over the past decades, global emissions of CO_2 and other greenhouse gases have steadily increased. The predicted (and in many cases already observable) negative effects on the environment have fueled a growing and persistent concern for sustainability across society. Governments, investors, and consumers alike are demanding greater transparency and accountability regarding the carbon footprint of companies around the globe. Consequently, organizations are under increasing pressure to report (and mitigate) the connected carbon footprint associated with their activity. Therefore, in today's industrial landscape, particularly in the manufacturing and chemical sector, sustainability is a strategic imperative.

Historically, carbon footprint has been assessed in a retrospective way by means of monthly or even quarterly data collection, periodical reports, and manual calculations prone to error. This approach offers limited value for fast operational decision making and is not guaranteed to comply with future stricter regulations.

In this context, Stamicarbon has leveraged the internal capabilities and experience with digital solutions to develop the Carbon Footprint Monitor, which is designed to provide real-time monitoring of the carbon footprint of industrial operations based on Distributed Control System (DCS) data and without the need for manual calculations. Moreover, such data can be aggregated and visualized for customized periods of time (weekly, monthly, quarterly, yearly…).

This digital solution seamlessly integrates with existing plant infrastructure and enables engineers, operators, and managers to have real-time access to detailed carbon footprint information and make informed decisions.

3.2 SOFTWARE ARCHITECTURE AND UNDERLYING MODELS

Stamicarbon's Carbon Footprint Monitor is a digital solution designed to automatically calculate and provide real-time information on greenhouse gas emissions to stakeholders. This digital solution has been developed with an architecture that guarantees its scalability, interoperability, and precision, leveraging the existing Process Monitor infrastructure.

3.2.1 Software architecture

The Carbon Footprint Monitor is built with a modular architecture that enables real-time data collection, processing and visualization. The system is composed of the following key layers:

- **Data acquisition:** The data acquisition layer is part of the Process Monitor and interfaces with the plant DCS to collect real-time data of relevant measured variables in the process. These variables typically include mass flow rates, temperatures, pressures, and steam utilization, amongst others. The plant connectivity is done following strict cybersecurity standards that enable the connection to be made in a secured way.
- **Processing and calculation**: This layer is responsible for applying standardized formulas and Life Cycle Assessment (LCA) models to transform DCS data into relevant environmental metrics that provide detailed insights into the plant's greenhouse gas emissions.
- **Visualization layer:** This layer is responsible for making the information available for operators, engineers, and managers, in a complete but easily understandable way. This is achieved by a collection of dashboards customized to the needs of the specific plant.

This modular approach ensures the flexibility and scalability of the system while allowing the solution to be fully integrated within other Stamicarbon digital solutions such as the NX STAMI™ Digital Process Monitor.

3.2.2 Life Cycle Assessment (LCA) and emissions monitoring

While the software architecture enables the deployment and usability of the Carbon Footprint Monitor, the heart of this digital solution lies in its implementation and use of Life Cycle Assessment (LCA) methodologies. These calculations are developed in collaboration with subject matter experts and are customized for each specific plant and chemical process. They are designed to calculate and quantify the environmental impact by calculating emissions based on a plant's real-time DCS data, and on values of a first-principle steady-state model of the plant. The main components of these models are:

- **Mass and energy balances:** The system continuously evaluates the input and output flows of materials and energy, ensuring accurate tracking of resource usage and emissions.
- **Steam and utility monitoring:** By incorporating values produced by the steady-state model and other utility-specific metrics, the system can attribute emissions to specific equipment or process steps.
- **Emission factors:** Calculations are based on internationally recognized databases such as Ecoinvent, ensuring consistency and comparability with international standards.
- **Customizable formulas:** Users can adapt or extend the default models to reflect site-specific conditions, product variations, or regulatory requirements.

The LCA models are designed to evolve. As new data becomes available or as regulatory frameworks change, the system can be updated to reflect the latest methodologies and benchmarks.

3.2.3 Leveraging the NX STAMI™ Digital suite

The Carbon Footprint Monitor is not a standalone application - it is deeply integrated into the NX Stami™ Digital suite. This integration offers several strategic advantages:

- **Unified Data Ecosystem:** By sharing a common data infrastructure with other systems, the Carbon Footprint Monitor ensures consistency and reduces data silos.
- **Scalability:** New plants or production lines can be onboard with minimal configuration.
- **Cybersecurity and Compliance:** The platform, compliant with the highest information security standards required by ISO27001, provides enterprise-grade security, access control, and auditability, ensuring that data is protected and trustworthy.
- **Cross-Tool Synergies:** Insights from the Carbon Footprint Monitor can be correlated with data from the Process Monitor or AI event detection tools, enabling root cause analysis and continuous improvement.

This synergy between the Carbon Footprint Monitor and the broader digital ecosystem transforms emissions monitoring from a compliance task into a strategic capability - one that supports real-time decision-making, operational efficiency, and long-term sustainability goals.

3.3 VISUALIZING ENVIRONMENTAL PERFORMANCE

The Carbon Footprint Monitor is not only a powerful engine for real-time data collection and emissions calculation. It is also a user-centric system for environmental insights. The calculated data can be visualized in a user-friendly way via a series of dashboards designed to translate complex data into actionable insights. The dashboards and dashboard groups are customized to each plant's needs and can be modified to accommodate for changes in the organization or in the regulatory framework.

3.3.1 Real-time emission tracking

At the core of the dashboard experience is the real-time carbon footprint tracker. This feature consists of a live view of greenhouse gas emissions calculated per unit of product (e.g. kg CO_2/kg urea). By continuously updating this value based on process data from the DCS, the tracker allows operators and engineers to detect anomalies and inefficiencies, to immediately determine the impact of changes in plant operation, and to compare current performance with historical data. Figure 2.1 shows the main gauge indicators which include the Urea product carbon footprint and the CO_2 equivalent emissions of the Urea process. Figure 2.2 shows the different contributors to the plant equivalent emissions.

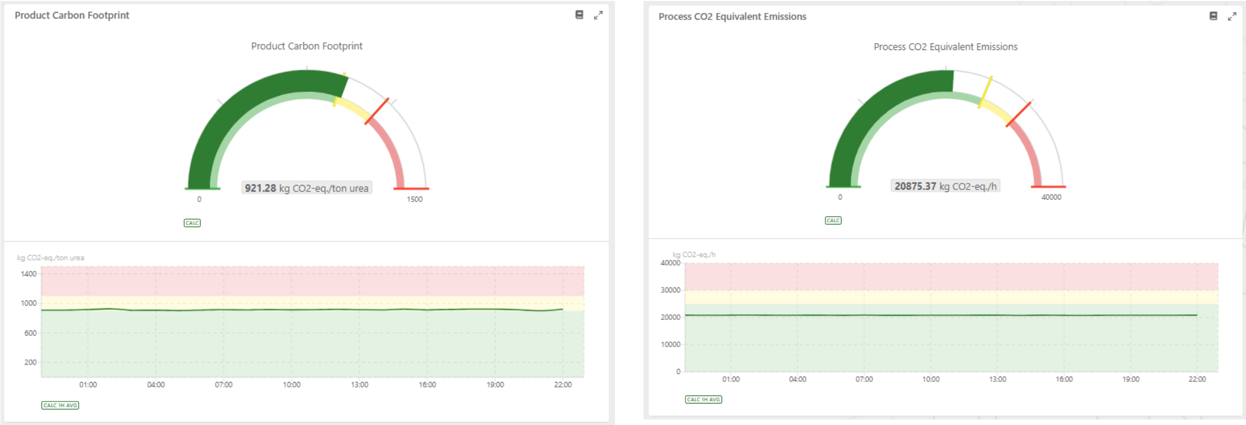

Figure 2.1: *Indicators for product carbon footprint and Urea melt process CO_2 equivalent emissions.*

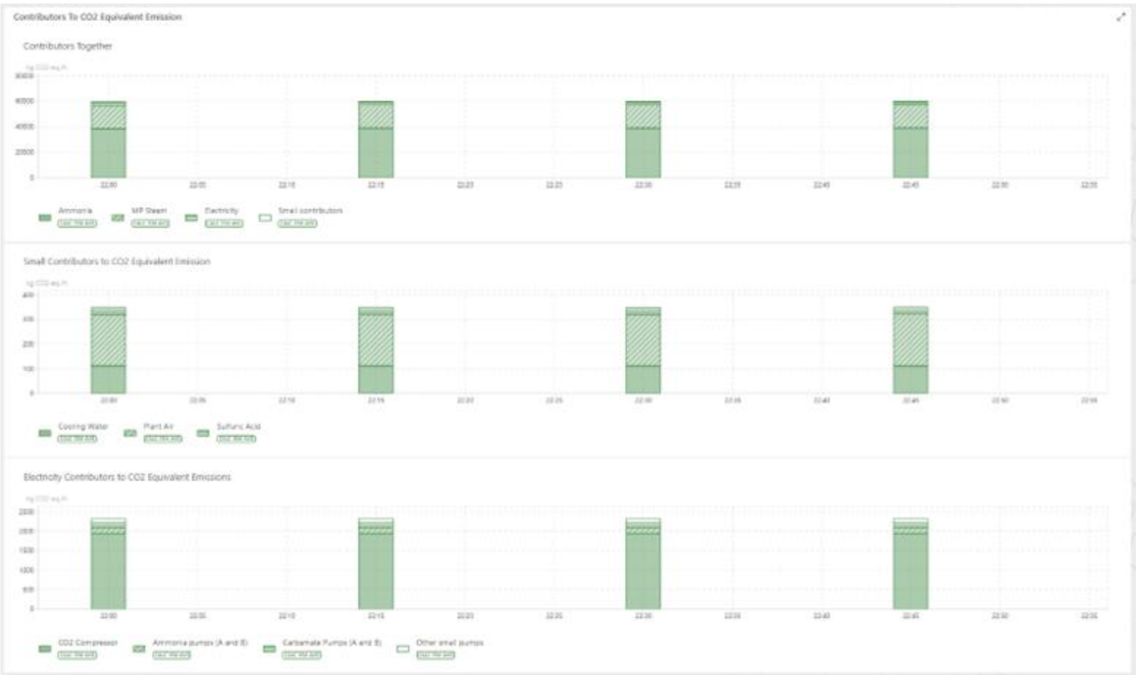

Figure 2.2: *Indicators of the contributors to the CO_2 equivalent emissions.*

3.3.2 Historical trends and reporting

The dashboard functionality also includes trend analysis capabilities for a customized period of time. These views allow the user to navigate historical data of the plant regarding greenhouse gases emissions by aggregating emissions data over time. Consequently, users can easily identify long-term patterns and seasonal variations, evaluate the effectiveness of process improvements, and generate reports for internal stakeholders or external auditors.

3.3.3 Source attribution and root cause analysis

A powerful feature of the Carbon Footprint Monitor is its ability to break down plant emissions by source and to visualize them in user friendly dashboards as shown in Figure 2.3. In this way, dashboards can attribute emissions to specific:

- Equipment such as reactors, heat exchangers and compressors.

- Plant sections.
- Operational parameters such as steam consumption and electricity usage.

Data presented with this granularity provides very valuable information and supports root cause analysis and continuous improvement. As an example, if a spike in emissions is detected, users can quickly trace it back to a specific equipment or process condition.

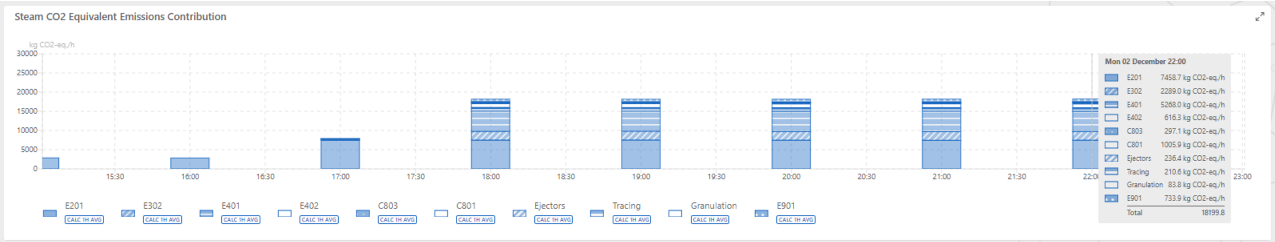

Figure 2.3: *Indicators of the steam consumer and their contribution to the CO_2 equivalent emissions.*

3.3.4 Energy and utility monitoring

In addition to emissions data, the dashboards provide insights into energy consumption and utility usage. This includes:

- Steam and electricity usage per unit of product.
- Energy intensity metrics.
- Correlation between energy use and emissions.

By integrating energy and emissions data, the Carbon Footprint Monitor supports a holistic view of environmental performance and helps identify opportunities for both carbon and cost savings.

4. SOFT N/C Meter

4.1 BACKGROUND: HARDWARE N/C METER

A hardware N/C meter is a specialized device employed in urea production facilities to determine the Nitrogen-to-Carbon (N/C) molar ratio in the liquid phase at the reactor outlet. This ratio is a critical parameter influencing the efficiency of conversion to urea. The optimal N/C ratio corresponds to the azeotropic point, facilitating maximum conversion and minimizing energy consumption. Departures from the ideal ratio result in reduced efficiency and increased operational expenses.

Stamicarbon's N/C Meter continuously analyzes the reactor's liquid effluent to provide real-time readings of the N/C ratio. It provides:

- Continuous analysis of reactor effluent.
- Quick results without lab sampling.
- Increased safety by eliminating manual sampling.
- Cost savings through reduced maintenance and lab expenses.

Benefits of accurate and automated measurement include:

- **Improved yield:** Reduces recycling and increases urea output.
- **Faster startup:** Speeds up stabilization after load changes.
- **Energy savings:** Up to 2% reduction in specific steam consumption and 60% reduction in ammonia losses.
- **Environmental benefits:** Better control over ammonia emissions.
- **Enhanced safety:** Eliminates manual sampling and reduces lab dependency.

This is especially valuable during plant startup, when feed measurements are unstable and quick N/C readings help stabilize operations.

4.2 SOFT N/C METER: MOTIVATION

Currently, the Urea industry is undergoing a profound transformation driven by digitalization, with soft sensors and digital twins emerging as pivotal technologies in this evolution.

Soft sensors, which are virtual instruments that estimate process variables using mathematical models and real-time data, offer a compelling alternative to traditional hardware sensors. They enable continuous monitoring of critical parameters that are either difficult or costly to measure directly. Unlike hardware sensors, soft sensors require no physical installation or maintenance. Their deployment is faster and more scalable. They can replace faulty or missing instruments and reduce reliance on lab analysis and manual interventions. Furthermore, their integration with digital twins, which are data-driven replicas of physical plants, unlocks new dimensions of operational intelligence and process optimization.

With these considerations in mind, the concept of a software-based alternative to the hardware N/C Meter emerged with the following key objectives:

- Provide clients with an expanded suite of digital solutions that enhances plant performance.
- Support AI integration and digital transformation initiatives.
- Enable flexible, scalable deployment through a subscription-based model with minimal upfront investment.
- Accelerate implementation and validation with shorter lead times.
- Ensure adaptability to various plant designs and operational conditions.

4.3 SOFT N/C METER: MODELING APPROACHES

The Soft N/C Meter can be developed using either first-principles models or data-driven models, each offering distinct advantages.

First-principles models rely on fundamental physical, chemical, or thermodynamic laws to describe the behavior of the Urea process. These models are built using equations derived from conservation laws (mass, energy, momentum) and known system parameters, allowing for the estimation of unmeasured variables based on measurable inputs. As a urea technology licensor, Stamicarbon has superior capabilities in developing soft N/C meters by applying first-principles models built on its proprietary, proven, and highly accurate thermodynamic and kinetic models.

In contrast, data-driven models, such as those based on machine learning (ML) techniques, utilize historical operational data to identify patterns and correlations among process variables without explicit reliance on the underlying physical or chemical relationships. These models are effective in managing intricate, nonlinear processes, and prove particularly valuable when the development of first-principles models is challenging or not feasible. By leveraging large datasets from plant operations, data-driven models can deliver accurate predictions and real-time insights, complementing or enhancing traditional modeling approaches.

It is important to note that large datasets are a fundamental prerequisite for developing reliable machine learning models in industrial and process engineering applications. Machine learning algorithms rely on extensive data to capture the inherent variability and complexity of plant operations of Urea processes, which involve nonlinear behaviors and dynamic interactions among numerous process variables. Insufficient data can lead to overfitting, poor generalization, and inaccurate predictions. Conversely, large datasets enable robust feature extraction, improve model accuracy, and reduce the impact of noise and anomalies. For the Soft N/C Meter application, this means collecting high-resolution data over extended time periods and with high variability.

4.3.1 A Stamicarbon advantage: Combining plant and synthetic data for reliable ML-based Soft N/C Meter performance:

Urea plants generally operate under steady-state or normal conditions to maintain safety and efficiency. Furthermore, urea plants may lack hardware that measures the N/C values at the reactor outlet. As a result, historical data collected from these plants is limited and at best often lacks coverage of a wide range of operating conditions. This creates a data scarcity problem, restricting the ability of ML models to generalize across different operating conditions. Consequently, ML models trained solely on stable plant data may perform poorly when faced with unseen or abnormal scenarios.

To overcome these limitations, Stamicarbon can leverage synthetic data produced by its proprietary Technology Training Simulators (TTS). These simulators use highly accurate first-principles models to mimic process behavior across a variety of conditions - including varying loads, synthesis pressures, and reactor water content - as well as other scenarios that cannot be easily replicated in an operational plant without incurring risk. This approach also eliminates the significant expenses associated with conducting experiments in the actual plant environment.

A hybrid approach that combines synthetic data with real plant data can significantly improve the robustness and accuracy of ML-based soft sensors. Real plant data ensures that models capture true process dynamics, noise patterns, and operational constraints, while synthetic data introduces variability across different operating conditions, including rare or extreme scenarios. This combination reduces overfitting to historical data and enhances predictive performance under unseen situations, enabling reliable real-time estimation of hard-to-measure variables such as the N/C ratio at reactor outlet.

4.3.2 ML-Based Soft N/C Meter for a real case study

A client expressed the need to monitor the Nitrogen-to-Carbon (N/C) ratio without the implementation of a physical meter. Our team developed an advanced Soft N/C model using machine learning, which analyzes historical process data and N/C measurements from 2021 and 2022. By applying supervised learning - mainly neural networks - and training on key variables like flow, temperature, and pressure collected from DCS tags, this solution can provide accurate, real-time N/C predictions and improve operational efficiency. The next sections detail how this approach works and summarize the results of the model.

The machine learning (ML) modeling process generally consists of four key stages: data collection and preprocessing, model development and training, model validation, and deployment. This structured approach was also applied in the case detailed below.

Data collection and preprocessing

The first stage involved gathering historical operational data from the existing urea production plant. The data was cleaned and organized to ensure consistency and reliability, then divided into subsets for training, test, and validation. Through exploratory data analysis and feature selection, 17 critical process parameters - such as temperatures, pressures, and stream flow rates - were identified as the most relevant inputs (features) for the model. The Nitrogen-to-Carbon (N/C) ratio was defined as the sole output (target variable) to be predicted. Additional preprocessing steps included handling missing values and outliers, normalizing and scaling input features for neural network compatibility, and ensuring diversity in the dataset to improve generalization.

Model development and training

The second stage utilized a neural network architecture due to its ability to capture complex nonlinear relationships between process variables. The dataset was split so that 80% was used for training and 20% for testing. Hyperparameters such as the number of hidden layers, neurons per layer, activation functions, learning rate, and regularization techniques were optimized to achieve a balance between predictive accuracy and computational efficiency.

Figure 3.1 (left) below illustrates the performance of the neural network model during the training phase. The x-axis represents time in minutes, while the y-axis shows the predicted and actual values of the N/C ratio. Two curves are plotted: one for the model's predictions during training (marked with blue crosses) and one for the actual measured values (shown as an orange line).

The close alignment between the predicted and actual N/C ratio values across the entire time range demonstrates the model's ability to accurately capture the underlying process dynamics. The model successfully tracks both gradual changes and sharp fluctuations in the N/C ratio, indicating that it has learned the relevant relationships between the input features and the target variable. Overall, the model demonstrates a high level of predictive accuracy.

Quantitatively, the model achieved an R^2 (coefficient of determination) of 0.996, indicating an excellent fit between predicted and actual values. The Mean Absolute Error (MAE) was 0.0052, and the Mean Squared Error (MSE) was 0.00005 during training. These metrics confirm the model's high accuracy and low prediction error.

This strong agreement between the model's predictions and the actual data during training suggests that the neural network is well-calibrated and capable of generalizing to unseen data, provided that similar process conditions are encountered. The results validate the effectiveness of the chosen architecture and the adequacy of the training dataset.

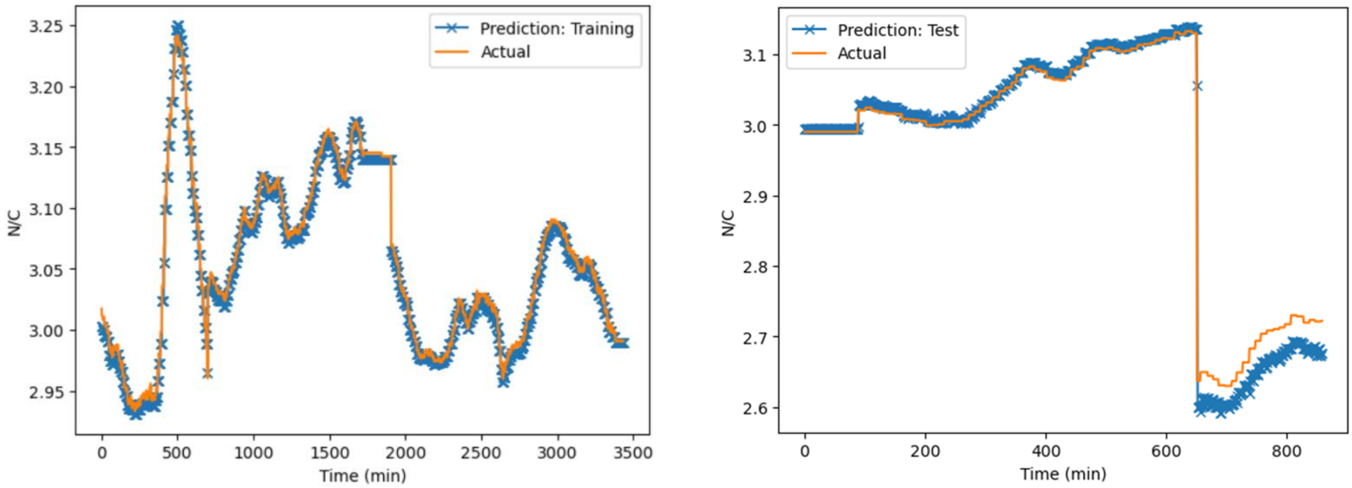

Figure 3.1: *Comparison of model training (left) and test (right) predictions with actual N/C ratio data over time, illustrating the predictive accuracy on both seen and unseen operating conditions*

Figure 3.1 (right) above displays the neural network model's predictive performance on the test dataset. Overall, the model's predictions closely follow the actual N/C ratio values throughout most of the test period, successfully capturing both steady-state and dynamic changes. There is a notable deviation around the 600-minute mark, where the model's prediction diverges from the actual value, likely due to an abrupt process change since such low N/C values are not typical during normal operation. Despite this, the model maintains a high level of accuracy across most of the test range.

Quantitatively, the model achieved an R^2 (coefficient of determination) of 0.98, indicating a strong correlation between predicted and actual values. The Mean Absolute Error (MAE) was 0.022, and the Mean Squared Error (MSE) was 0.0017 on the test dataset.

These metrics confirm that the model generalizes well to unseen data and maintains low prediction error outside the training set.

4.3.3 Model Validation

The validation phase assessed the neural network model's ability to generalize to new, unseen data from the urea production plant. Several validation scenarios were tested, each represented by a figure showing the predicted and actual N/C ratio values over time.

These results confirm that the model maintains high predictive accuracy and low error across diverse validation datasets. The R^2 values are generally above 0.95 (except for Validation 3, which still shows acceptable accuracy) which indicate a strong correlation between predicted and actual N/C ratios, while the low MAE and MSE values demonstrate minimal prediction error (see Table 3.1 on the following page).

The slight variation in performance across different scenarios reflects the inherent complexity and variability of plant operations, but overall, the model proves robust and reliable for practical deployment.

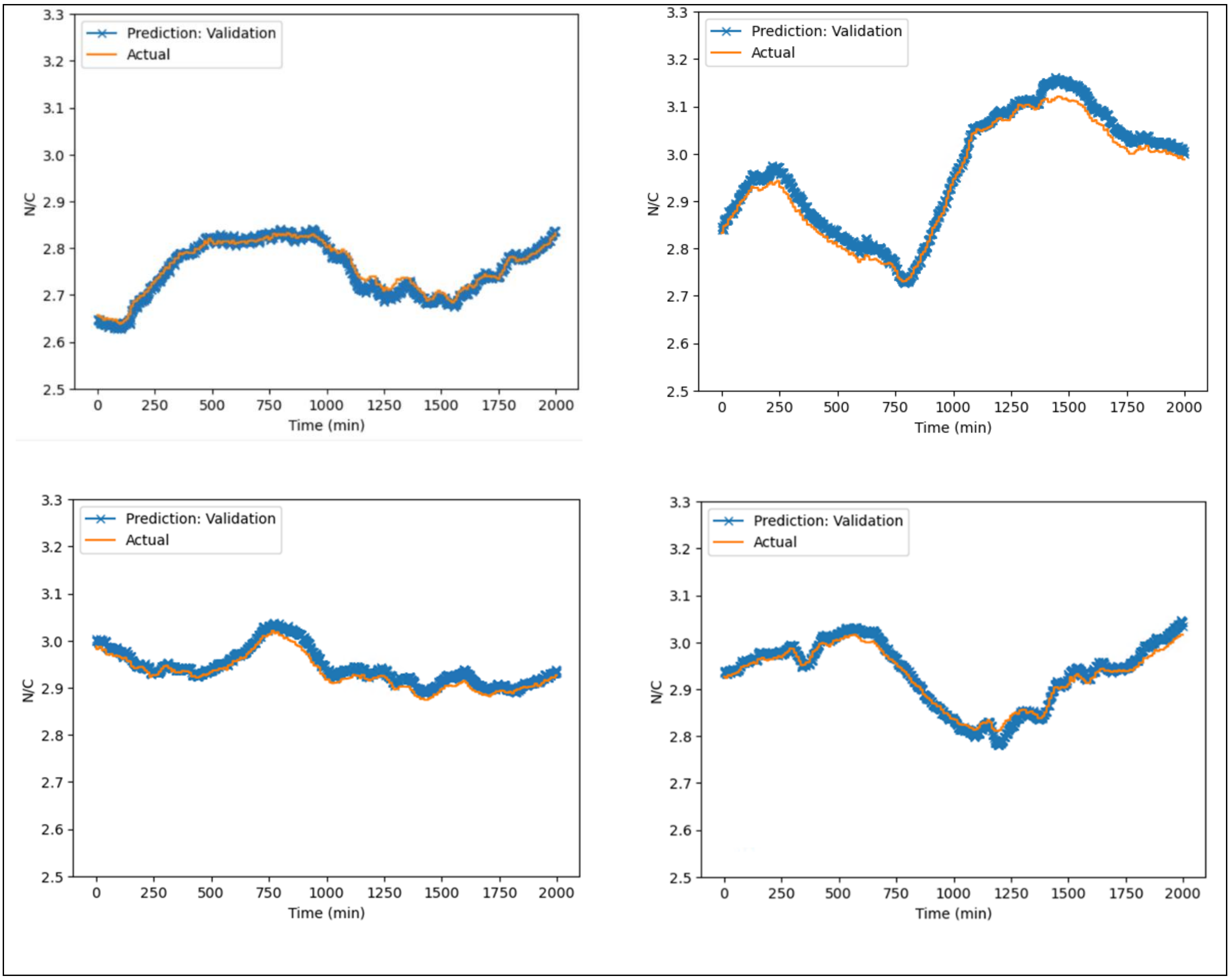

Figure 3.2 *Neural network validation performance shown in four segmented time windows, illustrating the agreement between predicted and measured N/C ratios*

Table 3.1: *Model Performance Summary*

Phase	R^2	MAE	MSE
Training	0.996	0.0052	0.00005
Testing	0.980	0.022	0.0017
Validation 1 (upper left)	0.96	0.013	0.0028
Validation 2 (upper right)	0.97	0.031	0.0012
Validation 3 (lower left)	0.84	0.021	0.00053
Validation 4 (lower right)	0.95	0.018	0.00046

4.3.4 Deployment Phase

Finally, the trained model is integrated into a cloud-based environment for real-time data ingestion and prediction. This integration will enable continuous monitoring of the N/C ratio at the reactor outlet and offer a scalable architecture for future model updates and re-training.

5. AI-POWERED EVENT DETECTION

5.1 BACKGROUND: ANOMALY DETECTION IN THE PROCESS INDUSTRY

Plant operation in chemical industries is inherently complex, requiring rapid identification and resolution of issues to maintain safety, efficiency, and profitability. During normal operation, anomalies may occur without being immediately evident to operators. These anomalies can stem from multiple factors such as instrument or equipment failure, human error, or process upsets. If left undetected, they can lead to cascading consequences:

- **Loss of efficiency:** Deviations from optimal conditions reduce yield, increase energy consumption, and compromise product quality.
- **Profitability impact:** Persistent inefficiencies accumulate into significant financial losses over time.
- **Equipment damage:** Prolonged abnormal conditions accelerate wear and tear, leading to unplanned maintenance or costly failures.
- **Safety risks:** In high-pressure, high-temperature environments such as Urea synthesis, undiagnosed anomalies pose serious hazards to personnel and the environment.

Traditional monitoring systems rely on threshold-based alarms for individual parameters, which often fail to capture subtle multivariate patterns indicative of emerging problems. This limitation motivated the development of **AI Event Detection**; a proactive solution designed to monitor complex process data in real time and detect early signs of abnormal behavior before issues propagate.

While anomaly detection is crucial in every process, it could be more complicated for Urea production. The intrinsic slow dynamics of the Urea process, caused by the liquid residence times in different equipment and by the many recycles from the different sections, create a slow chain of effects that make anomalies hard to identify quickly.

5.2 AI EVENT DETECTION: MODELING METHODOLOGIES

AI Event Detection leverages machine learning and multivariate statistical analysis to detect and categorize process anomalies in complex industrial environments. Unlike traditional systems, this approach analyzes multiple variables simultaneously to uncover hidden correlations.

The utilized models incorporate different multivariate statistical process monitoring methods (MSPM) to analyze the correlation between key process variables. Unlike univariate approaches that monitor individual variables in isolation, MSPM evaluates multiple correlated process variables simultaneously, capturing the underlying relationships and detecting subtle deviations from normal operating conditions.

The most common MSPM techniques include Principal Component Analysis (PCA) and Partial Least Squares (PLS), which reduce dimensionality while preserving critical process information. More complex models utilize kernel functions to map data into a higher-dimensional space or incorporate other types of machine learning methodologies to tackle strong non-linearities in the relationships between the considered variables.

These methods enable the identification of abnormal patterns that may not trigger traditional single-variable alarms, providing early warnings of process upsets, sensor drift, or equipment malfunctions. By leveraging MSPM, the system enhances anomaly detection accuracy and robustness, particularly in environments such as Urea synthesis where dozens of interdependent variables interact dynamically.

These methodologies have been widely adopted across various sectors of the process industry to enhance operational reliability and efficiency. Practical examples of multivariate statistical process monitoring applications include:

- Monitoring in polymerization and petrochemical processes to detect subtle shifts in temperature and pressure profiles that indicate fouling or catalyst deactivation.
- Quality control in pharmaceutical production to maintain compliance with strict quality standards by monitoring multiple critical quality attributes simultaneously, ensuring batch consistency and early detection of deviations.
- Energy companies utilize it in power plants to monitor boiler and turbine performance, identifying anomalies caused by sensor drift or process instability before they escalate into costly failures.

5.3 MODEL DEVELOPMENT AND RESULTS

AI-based event detection in Urea production was explored through two complementary machine learning approaches: unsupervised learning and supervised learning. Both methodologies aim to enhance anomaly detection in the recirculation section of a Urea Melt plant, where slow process dynamics and strong variable interdependencies make early detection challenging.

5.3.1 Data gathering and model deployment

The performance of machine learning models for anomaly detection depends heavily on the diversity and quality of the training dataset. In industrial environments, this requirement often poses a challenge because urea plants typically operate under steady-state conditions to ensure safety and efficiency. As a result, historical plant data rarely covers the full spectrum of operating scenarios, especially abnormal or upset conditions. This limitation restricts the ability of ML models to generalize and accurately detect anomalies under unseen circumstances.

The present study relied exclusively on synthetic data generated by Stamicarbon's proprietary Technology Training Simulator (TTS). The TTS is built on highly accurate first-principles models that replicate the dynamic behavior of the Urea Melt process under a wide range of operating conditions (also see section 3.3.1 where the same methodology is described for Soft N/C Meter). This approach offers several advantages such as controlled variability of the process values, safe replication of abnormal scenarios and cost efficiency. For future deployments on a real plant application, the existing models would need to be adjusted using plant data or plant-specific synthetic data for anomalies for which operations data is not available.

The AI Event Detection models presented in this paper have been created and deployed for the TTS of a Urea 2000+ plant. This solution allows us to work on the models and demonstrate the results of these developments within an unconstrained demo environment. In addition, by leveraging synthetic data from TTS, the models achieved exposure to a broader range of conditions than would be feasible with plant data alone.

During the data gathering, a randomized design of experiments was applied to seven critical process variables, introducing controlled deviations around steady-state values to ensure the robustness of the model developed at varying conditions. Additional scenarios were simulated to represent specific anomalies, such as the rectifying heater (E302) fouling and CO_2 slippage, which were later used for supervised model development. The resulting dataset comprised hundreds of data points across multiple operating regimes, providing a comprehensive foundation for training and testing both unsupervised and supervised models.

For future pilot projects, the models will be adjusted or reworked with the integration of real plant data that would capture the behavior of the specific plant.

5.3.2 Unsupervised Machine Learning: detecting deviations from normal operation

The first approach focuses on identifying deviations from normal operating conditions without requiring prior knowledge of specific events. The model was trained using data representing steady-state and near-optimal conditions, enabling it to recognize when the process behavior diverges from expected patterns. Diagnostic indicators were implemented to quantify the degree of deviation and provide real-time alerts when anomalies occur.

Testing on simulated scenarios demonstrated the model's ability to detect a range of abnormal conditions, including heat exchanger fouling, CO_2 slippage, and flowrate disturbances. Deviations outside the modeled space were flagged early, even when individual process variables remained within normal ranges.

5.3.2.1 Case Study 1: Rectifying Heater (E302) fouling

The fouling of the rectifying heater E302 is a slow phenomenon caused by the deposit of iron oxide on the tube side of the heat exchanger. For this test, the fouling was simulated as a rapid linear decrease of the fouling coefficient over a one-hour period. The results as displayed on the AI Event Detection dashboard deployed on the Process Monitor are shown in Figure 4.1. The displayed diagnostics show two different types of deviations: deviation within expected process behavior and deviation outside of expected behavior.

The first type of deviation, within the expected process behavior, accounts for process dynamics and variables relationships present in the training data. When a deviation of this type is identified, it is easier to point at the tags showing anomalies as the model understands these correlations.

The second type of deviation, outside of expected behavior, accounts for the anomalies not present in the training data. In this case the fouling of E302 was not included in the dataset used to train the model as it is not considered a normal operating condition, so the AI Event Detection model identified a deviation in the process variables as seen in Figure 4.1. The comparison with the heat transfer coefficient of E201 (displayed from 100% to 0% in the same figure) showed a high accuracy in estimating the severity of the deviation and following its linear increase.

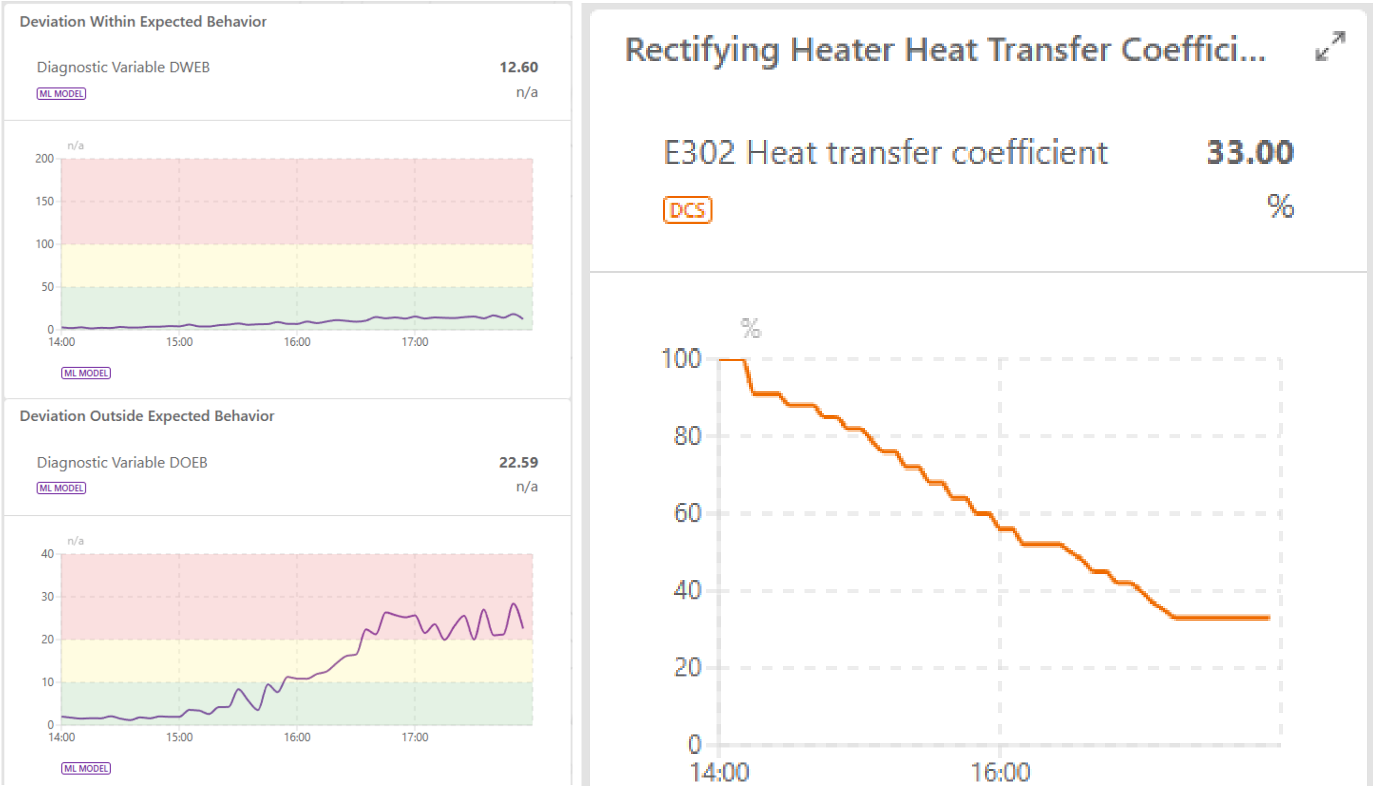

Figure 4.1: *Dashboard and diagnostics values, with indication of the heat transfer coefficient during the E302 fouling scenario.*

5.3.2.2 Case study 2: CO_2 Slippage

The CO_2 slippage is a scenario that can occur when the HP Stripper level is too low and the CO_2 gas exits the stripper together with the liquid solution. This anomaly has been simulated in the TTS where slippage occurs when the HP stripper level drops below 40% and increases exponentially as the level decreases. The model results are presented in Figure 4.2 together with the HP stripper level and the slippage CO_2 flowrate in mol/s.

The first diagnostics variable detects a deviation at the highest CO_2 slippage flowrate; however it is not a high value. This, like the previous example, derives from the slippage not being used in the training of the model. The second diagnostic heavily reacts to the CO_2 slippage and increases to very high values when the slippage flowrate increases and impacts the process. The value then decreases when the flow is decreased and the plant control system can bring the plant back to normal operation.

Figure 4.2: *Dashboard and diagnostics values, including E201 level, during the CO_2 slippage scenario.*

5.3.3 Supervised Machine Learning: targeted anomaly detection

To complement the general-purpose detection offered by unsupervised learning, supervised models can be developed for specific anomalies considered critical for plant reliability. These models are trained on labeled datasets combining normal and abnormal conditions, allowing them to predict the occurrence and severity of targeted events.

A case study focused on fouling in the rectifying heater E302, as already tested for the unsupervised machine learning model. The supervised models successfully tracked fouling progression during simulation, providing accurate and stable predictions as shown in Figure 4.3.

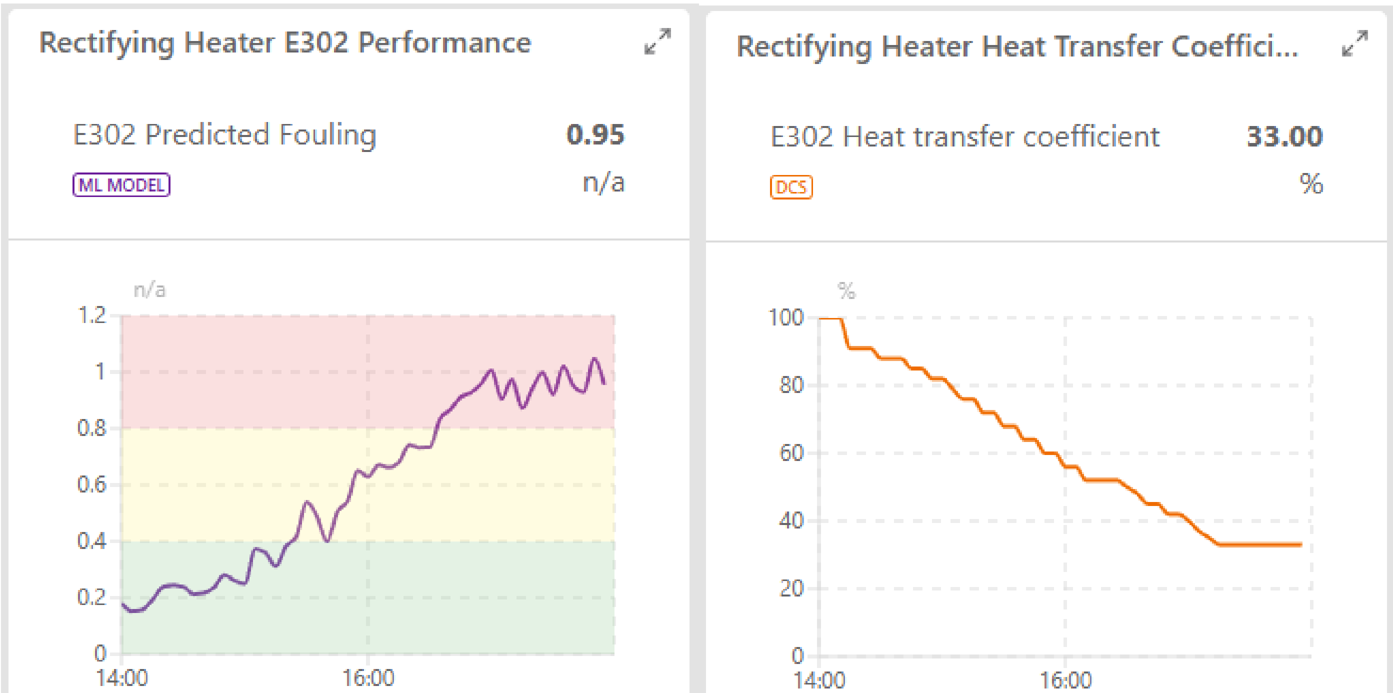

Figure 4.3: *Results for the E302 fouling prediction and E302 heat transfer coefficient.*

This targeted approach enables tailored monitoring strategies for high-risk scenarios, reducing the likelihood of undetected failures and improving decision-making for maintenance planning.

5.3.4 Insights for deployment in real plant operations

Combining unsupervised and supervised learning creates a robust framework for anomaly detection in complex chemical processes. The first approach offers broad coverage for unknown deviations, while the second delivers precision for predefined events.

The developed models are ready to be deployed on a real-plant application to support the plant's operation. The focus on tuning the model for an industrial application will be on:

- Expanding training datasets to include a wide range of realistic and desired operating conditions.
- Enhancing diagnostic capabilities to identify root causes of detected anomalies for the specific plant.
- Developing additional supervised models for specific events based on the results and anomalies detected by the unsupervised machine learning model.

This hybrid strategy strengthens operational reliability and safety, paving the way for scalable AI-based monitoring solutions across urea and other process industries.

Stamicarbon B.V.

REGISTERED OFFICE
Mercator 3, 6135 KW Sittard,
The Netherlands
P.O. Box 53 - 6160 AB Geleen
P +31 46 4237000
F +31 46 4237001

m.trabuio@nextchem.com
Process Engineer

a.elsibai@nextchem.com
Senior Process Engineer

stamicarbon.com

STAMICARBON
NEXTCHEM Sustainable Technology Solutions

NEXTCHEM