



## OPEN ACCESS

## EDITED BY

Andrea Melis,  
University of Bologna, Italy

## REVIEWED BY

Alessandro Vannini,  
University of Bologna, Italy  
Orhan Korhan,  
Eastern Mediterranean University, Cyprus

## \*CORRESPONDENCE

Rui Pinto,  
✉ rpinto@fe.up.pt

RECEIVED 02 August 2024

ACCEPTED 06 September 2024

PUBLISHED 25 September 2024

## CITATION

Belim M, Meireles T, Gonçalves G and Pinto R (2024) Forecasting models analysis for predictive maintenance.  
*Front. Manuf. Technol.* 4:1475078.  
doi: 10.3389/fmtec.2024.1475078

## COPYRIGHT

© 2024 Belim, Meireles, Gonçalves and Pinto. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Forecasting models analysis for predictive maintenance

Marco Belim<sup>1</sup>, Tiago Meireles<sup>2</sup>, Gil Gonçalves<sup>3</sup> and Rui Pinto<sup>3\*</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Faculty of Engineering of the University of Porto, Porto, Portugal, <sup>2</sup>BA Glass, Avintes, Portugal, <sup>3</sup>SYSTEC-ARISE, Faculty of Engineering of the University of Porto, Porto, Portugal

**Introduction:** This study explores the shift toward predictive maintenance through real-time data analytics to minimize machine downtime and improve machinery insights in industrial environments. Predictive maintenance aims to enable proactive interventions by predicting failures, enhancing operational efficiency.

**Methods:** The research was conducted in three stages. First, *BA Glass* equipment was sensorized using *OPC Router* and *PowerStudio SCADA* to facilitate real-time data extraction. A predictive maintenance algorithm was then developed in Python to analyze sensor data, predict failures, and trigger alarms. Finally, various forecasting models, including Linear and Polynomial Regression, Simple and Double Exponential Smoothing, ARIMA, and Prophet, were evaluated using a combination of blocked cross-validation and rolling window methodologies. The algorithm calculated performance metrics such as MSE, RMSE, and MAE for different parameter configurations and training sizes.

**Results:** A comparative analysis between wired and wireless sensors concluded that wireless sensors, although more expensive, were more practical and interchangeable in the factory setting. The results from the evaluation of prediction models showed that the Double Exponential Smoothing (DES) model with an additive damped trend and linear models performed best for datasets with daily seasonality and gradual oscillations. For datasets with stable trends and higher frequency oscillations, ARIMA and Prophet models proved to be more accurate.

**Discussion:** These findings suggest that the choice of sensors and prediction models can significantly impact the effectiveness of predictive maintenance systems. Wireless sensors offer long-term benefits in terms of flexibility and practicality, while the DES model and ARIMA/Prophet models are optimal depending on the dataset characteristics. This research highlights the value of real-time data analytics and predictive models in industrial environments for reducing downtime and improving decision-making.

## KEYWORDS

glass packaging, predictive maintenance, SCADA, OPC, time series analysis, forecasting

## 1 Introduction

The Smart Industry is profoundly transformed by technological innovations such as the Industrial Internet of Things (IIoT), Cloud and Edge Computing, Cyber-Physical Production Systems (CPPS), Artificial Intelligence (AI), and Big Data analytics. At the core of this evolution, maintenance emerges as a critical function ensuring the reliability,

efficiency, and safety of industrial processes and equipment throughout their lifecycle EN (2017); Silvestri et al. (2020). Recent concepts like Industry 5.0 Lu et al. (2022); Huang et al. (2022); Müller et al. (2020); Xu et al. (2021) and Maintenance 5.0 Cortés-Leal et al. (2022); Psarommatis et al. (2023) bring significant benefits to the Smart Industry, enabling predictive maintenance through real-time sensor data and advanced analytics to minimize downtime and promote sustainability. Workforce training programs further enhance creativity and innovation, addressing skill gaps and ensuring workforce relevance Pinto et al. (2024); Pinto et al. (2023).

The adoption of circular economy principles, such as refurbishing and recycling equipment, reduces waste and conserves resources, aligning with global sustainability objectives. Robust data management solutions facilitate comprehensive data collection and analysis, optimizing maintenance operations and decision-making within manufacturing and industrial settings. Maintenance encompasses a spectrum of activities aimed at preserving equipment condition and performance. This includes corrective maintenance, responding to unexpected failures through planned or unplanned interventions, with planned corrective maintenance offering strategic efficiency and safety advantages. Preventive maintenance focuses on scheduled tasks to prevent equipment failures and degradation, optimizing reliability, lifespan extension, and safety Basri et al. (2017).

Integrating diverse maintenance strategies within the Smart Industry framework Márquez and Papaelias (2020) underscores the sector's adaptability and optimization potential. Leveraging advanced analytics and real-time data monitoring enables organizations to proactively manage maintenance operations, minimize downtime, and enhance overall operational efficiency Zonta et al. (2020). Predictive maintenance, rooted in data analysis and real-time asset monitoring, plays a pivotal role in minimizing maintenance frequency and preventing unplanned outages within industrial operations. It utilizes historical and real-time data to proactively identify operational anomalies and potential equipment defects Zonta et al. (2020). This approach differs from traditional condition-based maintenance by predicting when failures will occur, enabling planned corrective interventions before equipment degradation. Predictive maintenance leverages technology such as IoT and AI, integrating techniques like vibration analysis, oil analysis, thermal imaging, and equipment observation to pinpoint areas needing attention.

Integration within the Smart Industry fosters cohesive processes and data flow, facilitating horizontal connections among machines, supply chains, and organizational functions, as well as vertical connections for data-driven decision-making Rüßmann et al. (2015); Laaper et al. (2018). Cloud and Edge Computing further enables real-time data processing and predictive maintenance, with scalable infrastructure and optimized operational responsiveness Dillon et al. (2010); Cao et al. (2020). Big Data analytics, characterized by processing vast datasets at high speeds and diverse types, plays a pivotal role in supporting equipment monitoring, predictive maintenance, and data-driven decision-making Plunkett et al. (2013); Addo-Tenkorang and Helo (2016). By harnessing extensive data resources on machinery performance and maintenance, organizations can tailor their processes for maximum efficiency and operational effectiveness Addo-Tenkorang and Helo (2016); Zonta et al. (2020).

BA Glass Alves (2012), a prominent international glass container manufacturer with a century-old legacy, exemplifies this transformative approach. Present in seven countries, producing over 11 billion containers annually across 12 plants with approximately 4,000 employees, BA Glass serves more than 60 countries worldwide Alves (2023). This work delves into BA Glass's initiative at their Avintes plant, where predictive maintenance strategies intersect with Industry 4.0 technologies to enhance electrical maintenance practices and ensure operational excellence.

The glass packaging industry, serving as a key supplier to food and beverage sectors, emphasizes efficiency amidst continuous high-volume production. To stay competitive, industries like BA Glass prioritize operational efficiency, reducing production losses through enhanced maintenance management and improved equipment reliability Testa et al. (2017); Pombal et al. (2019). This study aims to identify optimal sensors, integrate them with Supervisory Control and Data Acquisition (SCADA) systems using tools such as PowerStudio SCADA Circutor (2013); Wasaya et al. (2021); Ponte et al. (2024), and develop precise prediction models using Machine Learning (ML) techniques. Through these efforts, BA Glass aims to minimize downtime, enhance operational efficiency, and empower decision-makers with insights for proactive maintenance.

This work addresses two primary research questions:

*RQ1 What are the most suitable types of sensors for the diverse machinery within the BA Glass Avintes factory floor?*

*RQ2 What are the most accurate prediction models for predicting the state of components in the Water Cooling System of the AV5 furnace?*

Regarding RQ1, the main goal is to identify and evaluate the most appropriate sensors for monitoring the diverse machinery used in the factory, ensuring they meet operational needs and environmental conditions. This involves assessing various sensors for compatibility, durability, and data transmission capabilities. The selection process includes a comparison between wired and wireless sensors, evaluating factors such as installation complexity, cost, and data accuracy.

As for RQ2, the main goal is to determine the most effective predictive models for forecasting the condition and potential failures of components in the AV5 furnace's water cooling system. This entails developing and testing various ML models, including Linear and Polynomial Regression Nunno (2014), Exponential Smoothing Hu et al. (2013); Bucay-Valdiviezo et al. (2023), AutoRegressive Integrated Moving Average (ARIMA) Mo (2023), and Prophet models Riady (2023); Caicedo-Castro (2023). The evaluation framework employs blocked cross-validation and rolling window methodologies to simulate real-time conditions and assess model performance using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) Plevris et al. (2022).

By addressing these research questions, the study aims to enhance BA Glass's predictive maintenance capabilities, reducing downtime, improving operational efficiency, and supporting a proactive maintenance strategy.

The paper is organized into five more sections. Section 2 briefly presents the BA Glass company, as well as the context in which the

study arose, namely the glass packaging production process and the maintenance processes. [Section 3](#) describes in detail the research methodology, proposed solution and testing methodology, organized into three topics: I) Sensorization and Data Extraction; II) Predictive Maintenance Algorithm; and III) Forecasting Models. [Section 4](#) summarizes the validation/testing results and analyzes how the proposed solution can address the digital transformation of maintenance processes in *BA Glass*. [Section 5](#) covers the key findings of the study, including the potential shortcomings and limitations on their interpretations, and their integration into the current understanding of the problem. Finally, [Section 6](#) concludes the paper, stating final remarks about the study performed while further discussing the proposed approach.

## 2 BA Glass production and maintenance processes

*BA Glass* is a Portuguese company dedicated to the development, manufacturing, and commercialization of glass containers for the food, beverage, pharmaceutical, and cosmetic industries, with over a century-old history. *BA Glass* is present in 7 countries, producing over 11 billion containers annually. With a total of 12 plants and around 4,000 employees, *BA Glass* distributes glass packaging to more than 60 countries around the world [Alves \(2023\)](#).

Incorporated in 1912, initially focused solely on the commercialization of glass bottles. Subsequently, in 1930, the company initiated industrial activities with semi-automatic technology. In 1947 the company introduced automated technology, with the installation of the automatic feeder mechanism and molding machine in bottle manufacturing. By 1965, the firm was able to expand its weekly production to approximately 350 thousand bottles, a capacity six times greater than that of 1947. In 1969, a new industrial unit began operations in Avintes, Vila Nova de Gaia, with two regenerative furnaces (with heat recovery), pioneering the use of this technology over traditional raw material fusion methods.

1971 came with the installation of the first automatic Individual Section (IS) machine, leading to a substantial increase in production capacity. By 1979, the company relied on five IS machines, one of which was computerized. In 1983, a new regenerative furnace was installed. A new technology was also developed in the area of computerized moulding and automated product quality control in all production lines. 1988 led with the construction of a new furnace that increased the capacity by approximately 40%.

From the 1990s onwards, the company embarked on an expansion phase, characterized by several acquisitions and constructions. The group is currently organized in three different sections:

- The Iberian Division, composed of the plants of Avintes, Marinha Grande, and Venda Nova, all located in Portugal, and León and Villafranca de los Barros, in Spain
- The Central Europe Division, formed by the plants of Jedlice and Sieraków, both located in Poland and Gardelegen, in Germany
- The Southeastern Europe Division, consisting of the plants of Athens, is located in Greece, Bucharest, in Romania, and Plovdiv and Sofia, both in Bulgaria

### 2.1 Production process

Glass is a 100% recyclable material, having the ability to be reused multiple times without quality and characteristic loss. The typical composition of a glass package mainly includes silicon dioxide, sodium oxide, and calcium oxide, with variations of other chemical compounds based on the desired glass colour. These raw materials are stored in silos, measured, and mixed according to specific recipes and then vitrified in furnaces through a fusion process.

At the refractory furnaces, the material is melted at temperatures ranging from 1,500 to 1,600°C. The molten glass then flows through a large container under the influence of gravity, undergoing a tuning stage to ensure uniform thermal homogeneity throughout the melted mass, a critical condition for achieving a high-quality product. The furnace temperature can be real-time controlled using control panels.

The furnace is entirely constructed from refractory material and utilizes natural gas as a fuel source. It is where the raw materials are melted to produce glass, which is then directed through channels or feeders to reach the moulding machines.

The Avintes plant is comprised of three furnaces, designated AV2, AV4, and AV5, each with its own set of production lines. Furnace AV2 accommodates three production lines, as well as furnace AV4. Furnace AV5 has four production lines. It's in these ten production lines that the glass forming is done.

The glass packaging manufacturing process begins with obtaining a molten glass gob, which is then cut and directed by gravity through channels to an automatic moulding machine. As the bottles emerge from the moulds, they are at a temperature of around 600°C, and they undergo a swift cooling process, causing internal stresses to develop. Following a brief period of stabilization on a dead plate, grippers move the bottles onto a conveyor belt that carries them to the subsequent stages of processing.

To release tensions caused by temperature fluctuations as moulds open and close, a heat treatment known as annealing is performed to homogenize the material. This involves coating the packages with tin or titanium chlorides, providing a protective layer on the glass surface to prevent defects caused by surface friction and increasing resistance to mechanical shock. Also, thermal annealing in an annealing lehr, i.e., raising the glass to a higher annealing temperature and then gradual cooling at a medium-low speed, as well as cold-end coating, applied at room temperature.

After completing the annealing process, the pieces are introduced into automatic inspection machines equipped with various defect detection mechanisms to identify and separate faulty pieces. Finally, the packaging is conveyed to the end of the production line, where they are stacked in tiers to form a pallet.

### 2.2 BA Glass maintenance processes

In any high-production facility, dedicated maintenance teams play a crucial role in ensuring machinery efficiency and, consequently, sustained profitability. Within the fast-paced glass packaging industry where *BA Glass* operates, disruptions to furnaces or control systems can have significant adverse consequences. To uphold optimal operational conditions and enhance machinery

upkeep, the presence of diverse and responsive maintenance departments is essential to quickly address and resolve any operational challenges.

The efficient operation of glass container manufacturing at *BA Glass Avintes* relies on multiple maintenance departments, being the Electrical Maintenance and Instrumentation Department (DEI) the most important. Their responsibilities encompass the upkeep, servicing, and calibration of electrical and instrumentation equipment used in production facilities. The electrical maintenance component's role is to preserve the functionality of all electrical equipment, covering power distribution systems, motors, switchboards, transformers, and various other electrical components.

To ensure the proper functioning of the electrical infrastructure, they regularly conduct testing, maintenance, and troubleshooting. Regarding the instrumentation component, their tasks include calibrating and maintaining machinery and control systems in the factory. This responsibility extends to instruments dedicated to process control and monitoring, such as temperature sensors, pressure gauges, flow meters, and level sensors.

Furthermore, the DEI is committed to continuous improvement, working on energy efficiency projects, implementing new technologies to increase efficiency, and making recommendations for upgrading electrical systems and sensors. Collaborating with other departments, they seek opportunities to enhance operations and implement best practices, contributing to an overall improvement in performance.

The DEI at *BA Glass Avintes* employs a comprehensive approach to maintenance, encompassing both reactive and proactive strategies. In response to unforeseen challenges such as equipment failures or breakdowns, the team reacts promptly by performing troubleshooting and repairs to restore operational status. In emergencies, such as power outages or severe equipment failures, the team mobilizes quickly to implement urgent corrective measures, minimizing downtime.

On the proactive side, preventive maintenance takes centre stage. The team actively works to identify and address potential issues before they escalate, conducting regular inspections, tests, and maintenance tasks to ensure equipment remains in optimal condition. This preventive approach aligns closely with the systematic preventive methodology, involving the performance and adherence to maintenance tasks at predetermined intervals, often without the need for equipment sensorization.

However, the current reliance on manual monitoring, where personnel physically measure machine parameters, has revealed critical challenges due to the absence of real-time sensors, resulting in a lack of timely identification of machine anomalies, potentially leading to operational disruptions. This inadequacy poses a significant challenge, as machines necessitate sensorization, alarms require systematic development, and data from already sensor-equipped machines demand comprehensive analysis.

Ultimately, faced with the need to minimize machine downtime and gain a comprehensive insight into the current state of the plant's machinery, the Avintes plant recognized the need to implement predictive maintenance strategies. Predictive maintenance, relies on two main components: I) real-time health information, which requires the collection of timely data; II) the development of a

consumption prediction tool utilizing ML regression models, by leveraging the data collected.

### 3 Materials and methods

We followed a clearly defined workflow to implement predictive maintenance at *BA Glass*. This process began with identifying sensors tailored to the factory's requirements and integrating them into the plant's SCADA system. Next, we established a communication link between the SCADA system and the *OPC Router inray Industriesoftware GmbH (2024)* for real-time data extraction. A statistical model was then developed to predict the degradation status of critical components. Finally, we implemented an algorithm to analyze the statistical models and generate alarms within the *PowerStudio SCADA* system. By proactively addressing potential machine issues, the organization aims to significantly reduce downtime and associated operational costs. The integration of real-time monitoring and predictive analytics is designed to provide decision-makers with valuable insights, thereby facilitating the shift from reactive to proactive maintenance strategies.

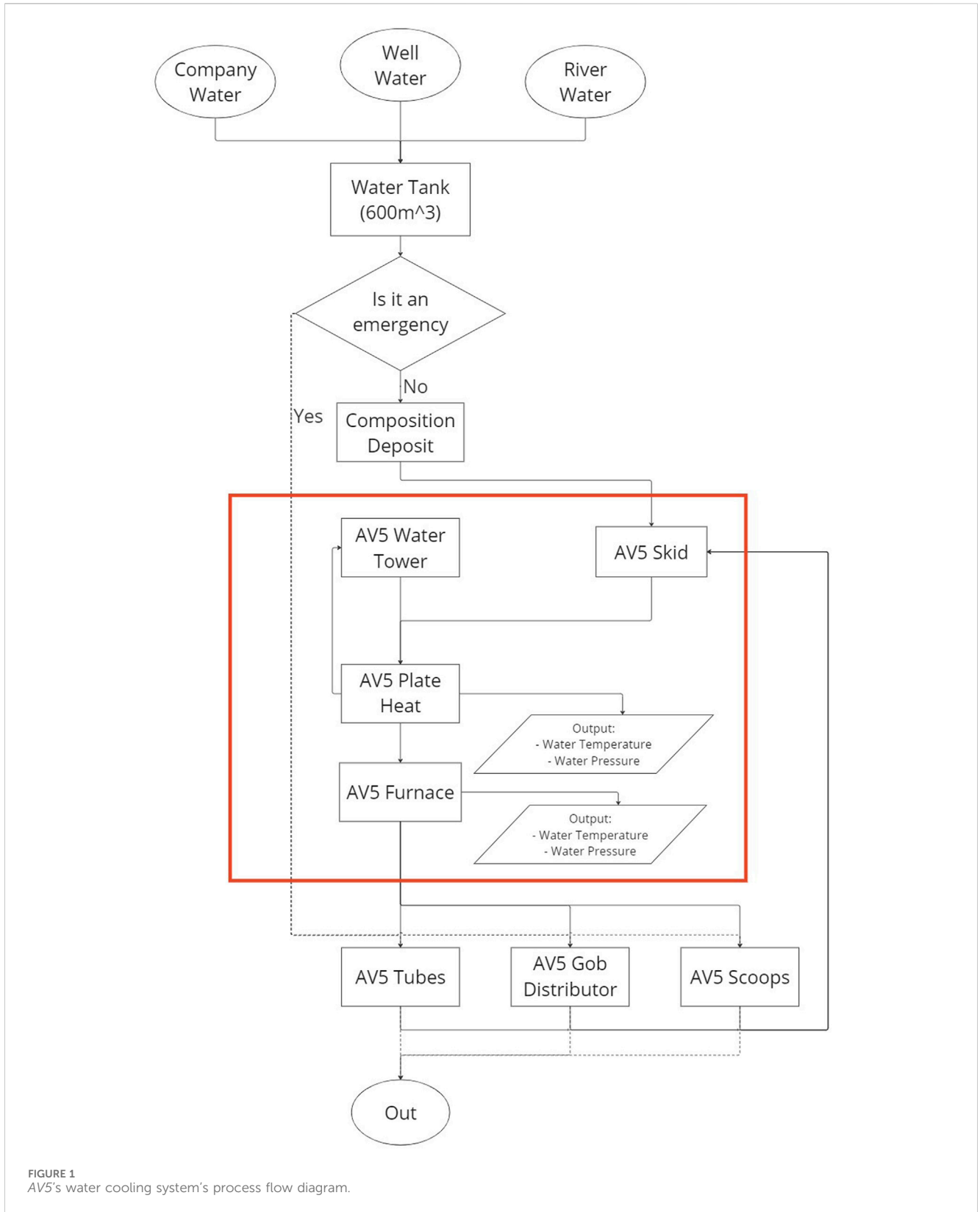
The successful implementation of predictive maintenance in industrial settings relies heavily on acquiring substantial datasets. This initiative began with the exploration of suitable sensors and predictive maintenance for the fans in the AV5 production line machines. We know that in the initial stage, the introduction of sensors would provide limited data, posing challenges for data extraction and effectively applying existing models of predictive maintenance. Thus, the focus was on the Water Cooling System sensors of the AV5 furnace, where a significant volume of data was already available.

The process flow diagram for the AV5 furnace Water Cooling System highlights sensorized components, particularly focusing on the water path surrounding the heat exchanger, as represented in [Figure 1](#). This area includes the water flow from the tower and skid, looping back to the tower, water discharge into the furnace, and return to the skid. Critical sensors for monitoring this system are temperature and pressure sensors on both the cold and hot sides of the heat exchanger.

Several common issues can affect the AV5 Water Cooling System, including heat exchanger clogging, tower cooling failure, and pump failures on both the cold and hot sides. These issues can be analyzed based on pressure and temperature readings from the sensors. [Table 1](#) provides a comprehensive matrix detailing potential failures in the system. It is important to note that some results labelled as "abnormal" are theoretically possible based on sensor conditions but have not occurred in practice.

#### 3.1 Sensorization and data extraction

For the sensorization process, we selected sensors based on the specific operational needs and environmental conditions of *BA Glass*. Both wired and wireless sensors were evaluated for their effectiveness and practicality. We integrated these sensors with the SCADA system using *PowerStudio SCADA*, ensuring seamless data acquisition and monitoring.



The data extraction phase involved establishing a communication link between the SCADA system and the *OPC Router*. This link was crucial for the real-time transmission of sensor data. The *OPC Router* was configured to collect and

transmit data from various sensors to a centralized database, where it was processed and stored for subsequent analysis.

Efforts focused on identifying sensors that met the specific requisites of *BA Glass* and the scope of this study. By carefully

TABLE 1 Heat exchanger monitoring matrix.

Cold side temperature	Cold side pressure	Hot side temperature	Hot side pressure	Failure description
High	Low	High	Low	Heat exchanger clogging
High	Low	High	Normal	Heat exchanger clogging
High	Low	Normal	Low	Heat exchanger clogging
High	Low	Normal	Normal	Abnormal
High	Normal	High	Low	Cooling tower ventilator failure
High	Normal	High	Normal	Cooling tower ventilator failure
High	Normal	Normal	Low	Abnormal
High	Normal	Normal	Normal	Cooling tower ventilator failure
Normal	Low	High	Low	Heat exchanger clogging
Normal	Low	High	Normal	Cool side pump failure
Normal	Low	Normal	Low	Heat exchanger clogging
Normal	Low	Normal	Normal	Cool side pump failure
Normal	Normal	High	Low	Hot side pump failure
Normal	Normal	High	Normal	Abnormal
Normal	Normal	Normal	Low	Hot side pump failure
Normal	Normal	Normal	Normal	No failure



FIGURE 2 To the left, the Sure Cross® QM30VT2 Vibration and Temperature Sensor. To the right, the WISE-2410 LoRaWAN Wireless Condition Monitoring Sensor.

balancing the technical requirements of BA Glass, cost considerations, and the need to integrate with existing systems, the decision to focus on the Sure Cross® QM30VT2 Vibration and Temperature Sensor and the WISE-2410 LoRaWAN Wireless Condition Monitoring Sensor, as represented in Figure 2.

- Compatibility with Company Requirements: The primary criterion for sensor selection was compatibility with BA Glass’s operational needs. The sensors had to meet the

highest dust protection certification (IP6X), which is critical in a glass manufacturing environment where dust and particulate matter can interfere with sensor performance. Additionally, the sensors needed to be capable of transmitting data at five-minute intervals to provide timely and actionable information for predictive maintenance.

- Integration with Existing Systems: Another key factor was the sensors’ ability to integrate seamlessly with BA Glass’s existing infrastructure, specifically the SCADA system. The selected

sensors were required to support communication protocols such as RS-485, RS-232, or Modbus TCP, which are standard in industrial environments and necessary for effective data exchange between the sensors and the monitoring system.

- **Cost Considerations:** Acquisition cost was also a significant factor in the selection process. The study was constrained by budgetary considerations, which necessitated focusing on cost-effective solutions that would still meet the technical requirements of the factory.
- **Operational Flexibility:** The *Sure Cross*<sup>®</sup> *QM30VT2* was chosen for its direct Modbus compatibility, making it a straightforward, cost-effective option for wired sensor integration. On the other hand, the *WISE-2410* was selected for its wireless, battery-operated design, offering greater flexibility in sensor placement without the need for extensive wiring, though it introduced additional complexity with the requirement of a LoRaWAN gateway.
- **Project Scope and Time Constraints:** The study was conducted within a specific timeframe and scope, which limited the number of sensors that could be thoroughly analyzed. The selection of these two sensors was based on their alignment with the project's immediate needs and the feasibility of their implementation within the study's duration.

The *Sure Cross*<sup>®</sup> *QM30VT2* *Vibration and Temperature Sensor* was chosen primarily for its ability to function as a Modbus Slave and directly connect to a Modbus-485 network. Its Modbus compatibility made it a reliable and cost-effective option for integration into the existing infrastructure. Alternatively, the *WISE-2410* *LoRaWAN Wireless Condition Monitoring Sensor* offered the advantage of a wireless, battery-operated design, providing flexibility in placement. However, its reliance on the *WISE-6610* *LoRaWAN Gateway* for connectivity introduced additional cost and complexity. While the *WISE-2410* has a prolonged battery life of 2 years, maintaining the desired data update frequency of every 5 min may require careful consideration and potential trade-offs affecting battery longevity.

These two sensors cater to different preferences and operational scenarios within the *BA Glass* facility. The *QM30VT2* emphasizes simplicity, direct connectivity, and compatibility with existing systems, while the *WISE-2410* introduces the benefits of wireless operation but necessitates additional infrastructure considerations.

In sum, the methodology for implementing predictive maintenance involved extracting and managing sensor data from the *AV5* furnace water system, emphasizing real-time data acquisition and model-based alarm generation. An *OPC Router* was integrated as an intermediary to access and manipulate data from the SCADA system. Initially, historical sensor data was extracted via the SCADA GUI, generating a CSV file. For ongoing real-time data acquisition, the *OPC Router* facilitated continuous updates of sensor readings from the SCADA system to a CSV file every 5 min, ensuring timely data availability for analytics and model generation.

## 3.2 Predictive maintenance algorithm

The core of our predictive maintenance strategy was the development of a robust consumption prediction algorithm. This algorithm was designed to analyze real-time data from the sensors and apply statistical models to predict the degradation status of critical components. The algorithm utilized techniques such as vibration analysis, oil analysis, thermal imaging, and direct equipment observation to identify potential failures.

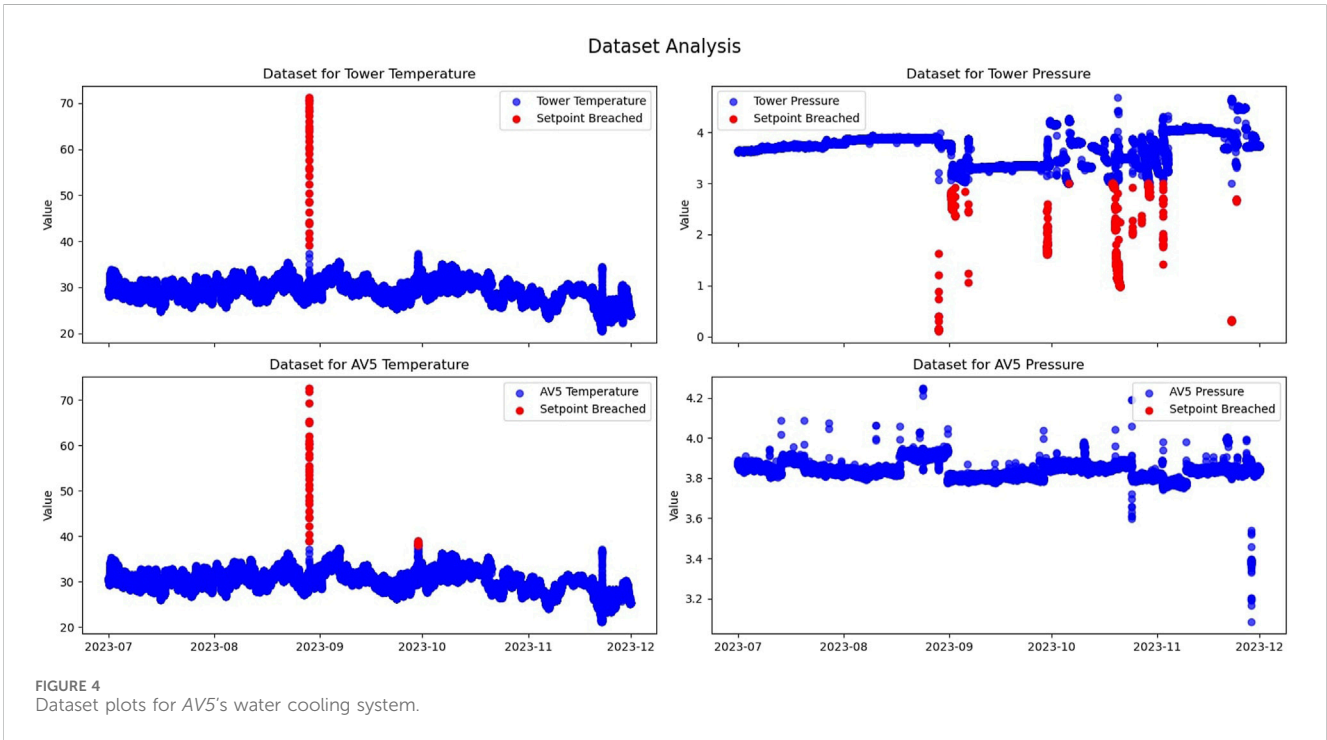
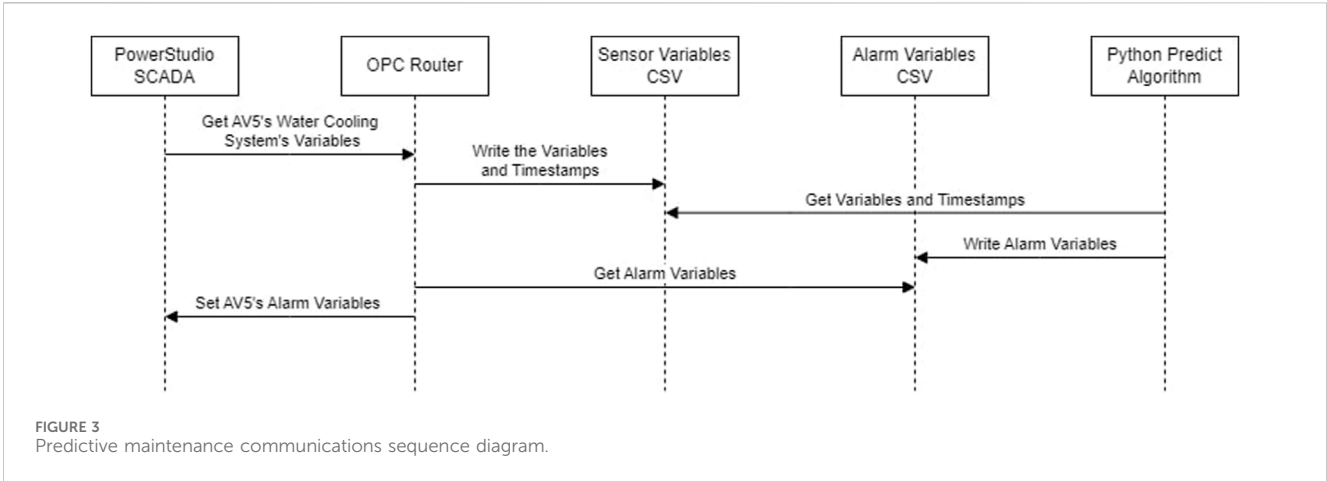
To ensure the algorithm's effectiveness, we incorporated features that enabled it to learn from historical data and improve its predictive accuracy over time. The algorithm generated alarms and notifications, which were integrated back into the *PowerStudio* SCADA system, providing real-time alerts to maintenance personnel. The primary goal of these alerts was to offer actionable insights that enabled maintenance teams to intervene before a failure occurred. By predicting a component's degradation or imminent failure, the alerts allowed operators to make informed maintenance decisions proactively, rather than reacting to unexpected breakdowns. This enables workers to make informed and timely decisions supported by the prediction algorithm.

The usability of these alarms is not merely a secondary feature but is central to the success of predictive maintenance. Alarms serve as the critical interface between the predictive algorithm and the maintenance personnel, translating complex data and predictions into actionable insights. Without these alarms, the output of the algorithm remains inaccessible to workers, rendering the predictive maintenance system ineffective. Thus, the functionality and reliability of the alarm system are fundamental to achieving the primary goal of predictive maintenance—preventing unplanned downtime and enhancing operational efficiency by enabling proactive interventions.

To enable predictive maintenance and alarm functionality, an algorithmic framework was developed using Python. This involved reading and filtering sensor data from a CSV file, applying prediction models to forecast setpoint breaches, and activating alarms based on predefined thresholds and a fault matrix. The algorithm was designed to incorporate a 15-min delay mechanism to discern genuine threats from transient fluctuations, enhancing system reliability. Additionally, alarm activation and deactivation were contingent on three consecutive predictions above or below thresholds, ensuring sustained improvements in system conditions.

The algorithm's adaptability was underscored by allowing users to adjust alarm thresholds based on operational requirements. The implementation of the algorithm involved establishing connections between Python, the *OPC Router*, and the *PowerStudio* SCADA system, ensuring seamless transmission of alarm variables, as represented in [Figure 3](#).

This comprehensive methodology enabled meticulous data analysis to identify patterns, correlations, and outliers, forming the foundation for robust predictive maintenance models and future evaluative efforts.



### 3.3 Forecasting models

A comprehensive dataset extracted from *PowerStudio SCADA* and *OPC Router* was utilized to develop and test robust predictive maintenance models. The dataset, represented in [Figure 4](#) encompasses pressure and temperature readings from the hot and cold sides of the heat exchanger, along with corresponding timestamps spanning approximately 44,000 entries from July to November 2023.

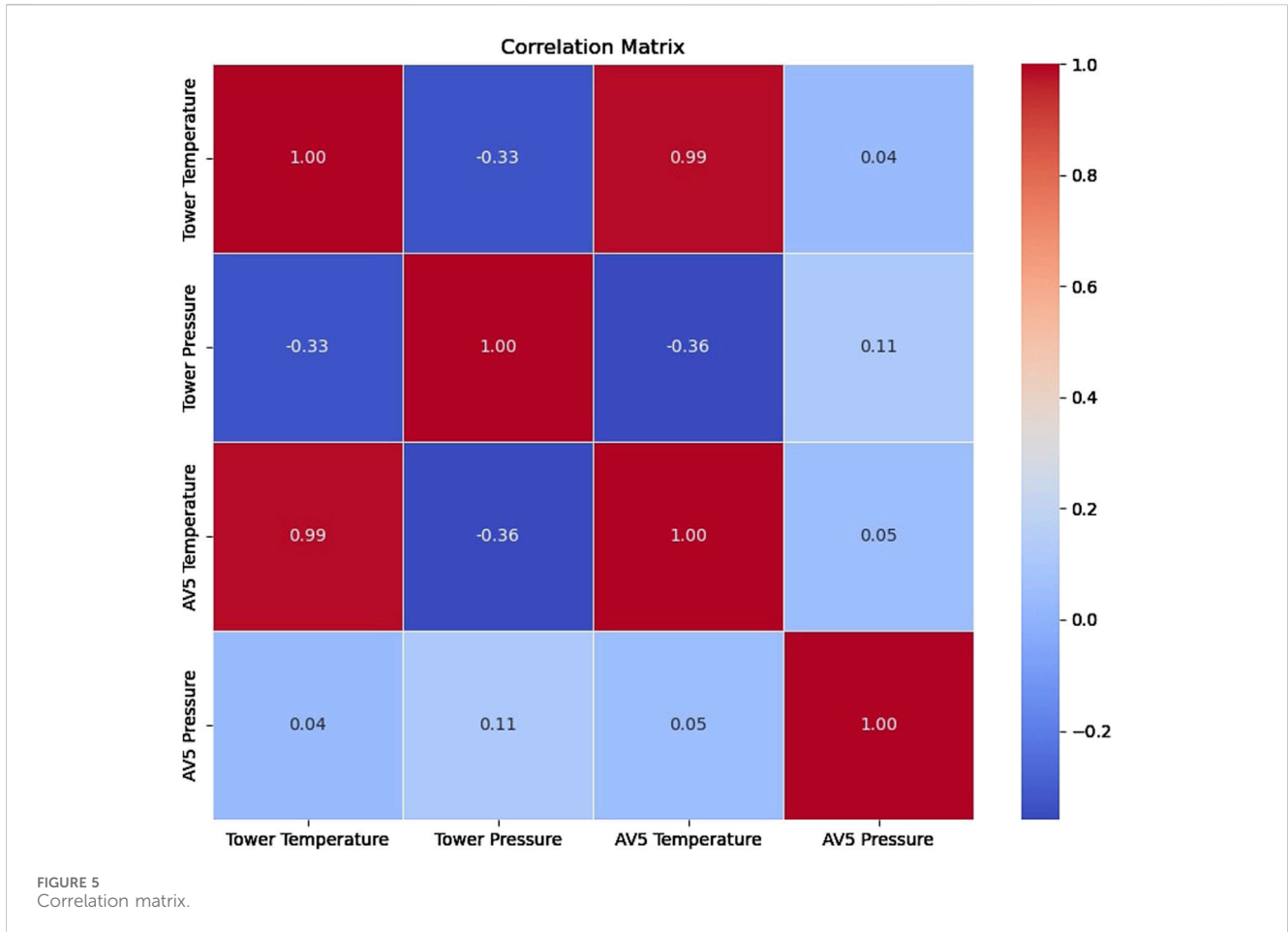
Visualizing this dataset revealed distinct patterns and insights. Temperature fluctuations exhibited synchronous behaviours on both sides of the heat exchanger due to the thermal exchange process within the system. Conversely, pressure readings on the tower side (cold) often exceeded setpoints, indicating potential

issues with tower cooling, while the furnace-side pressure maintained stability with occasional minor deviations.

A critical challenge involved addressing abnormal pressure drops to 0 bar, which was determined to be intentional during maintenance activities or benign operational behaviours. To maintain the focus on predictive model integrity, these outlier points were excluded from the analysis to ensure an accurate forecast of potential system faults rather than predicting known anomalies.

Exploring variable relationships through a correlation matrix, represented in [Figure 5](#), highlighted strong correlations between temperatures and moderate inverse relationships between tower pressure and temperature variables. This observation guided the development of similar univariate prediction models focusing on





individual variables' historical trends, emphasizing simplicity and interoperability.

Initial model evaluations involved segmenting the dataset into distinct phases to visually assess model performance before quantitative evaluation. This iterative process informed the selection of an 8-h prediction window, equating to 100 measurements, to forecast system component failures with practical operational relevance and proactive maintenance planning. By integrating data visualization, statistical analysis, and model testing, this methodology established a robust framework for developing reliable predictive maintenance solutions.

The choice of forecasting models was guided by prior research, which demonstrated their effectiveness in similar predictive contexts. In this case, the study focuses on the analysis of Linear and Polynomial Regression, Simple and Double Exponential Smoothing, ARIMA, and Prophet models because these models represent a range of well-established and widely used forecasting techniques that are suitable for different types of time series data typically encountered in industrial settings. Each model has distinct characteristics and strengths, making them valuable for different predictive maintenance scenarios:

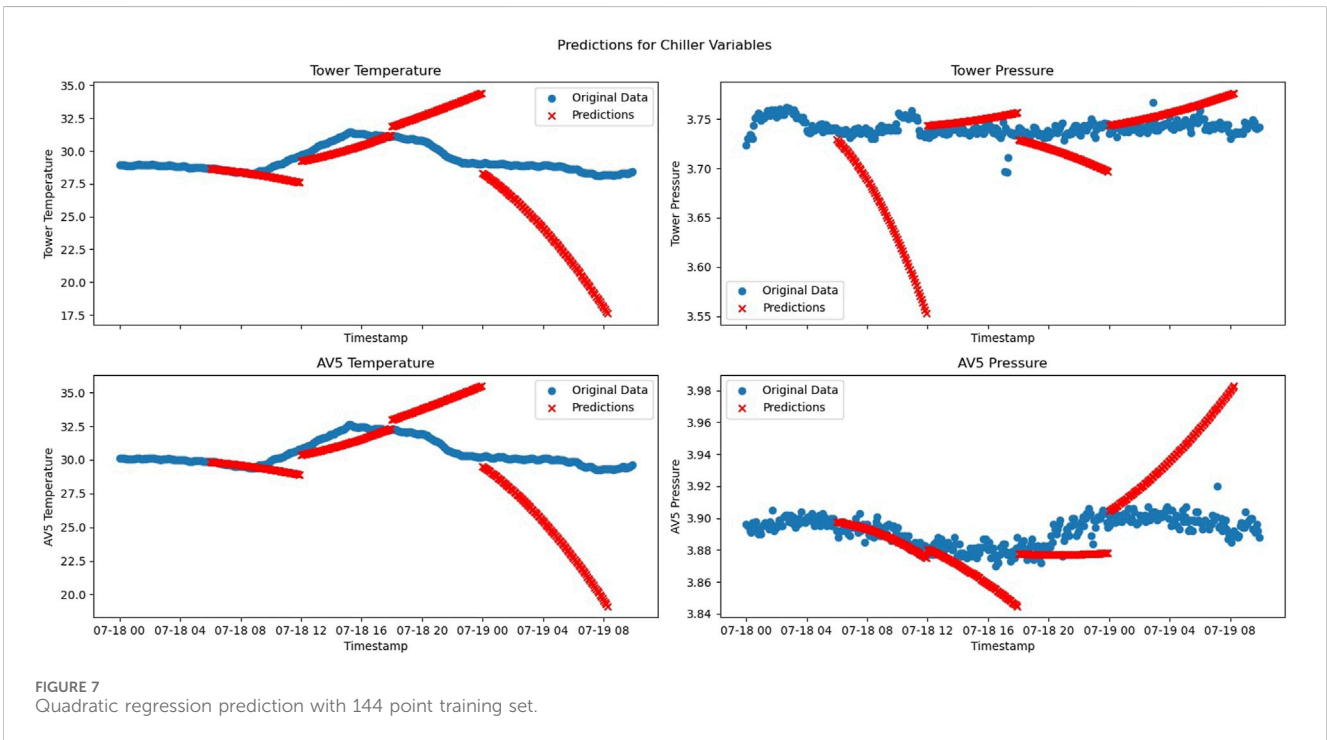
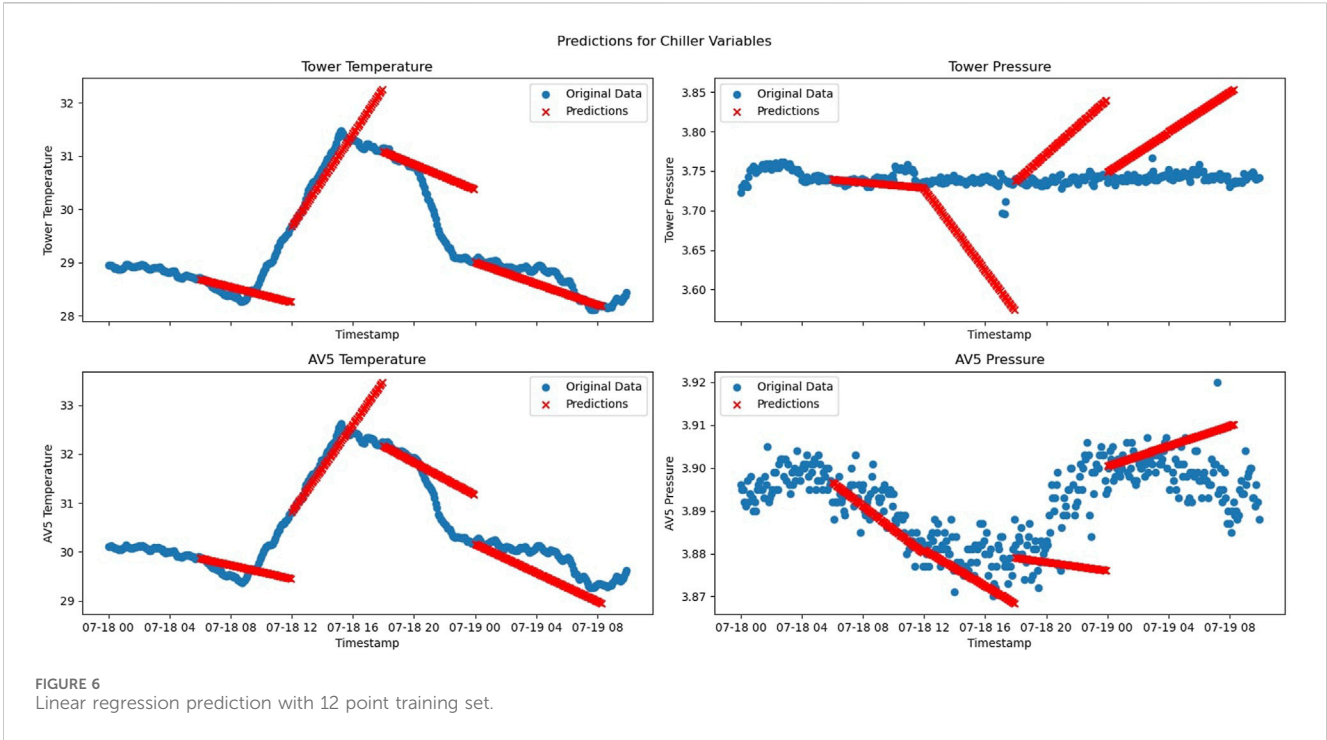
- **Linear and Polynomial Regression:** These models are fundamental tools in predictive modelling, useful for identifying and extrapolating linear and non-linear trends in data. They provide a straightforward approach to

forecasting, making them ideal for situations where the relationship between variables is relatively simple and stable.

- **Simple and Double Exponential Smoothing:** Exponential smoothing models are particularly effective for forecasting data with trends and seasonal patterns. The Double Exponential Smoothing model, with an additive damped trend, is designed to handle data with trends that either strengthen or weaken over time, which is common in maintenance scenarios where equipment wear and tear may not follow a consistent pattern.
- **ARIMA:** ARIMA is a powerful model for analyzing time series data that exhibit non-stationarity and autocorrelation. It is particularly suited for datasets where past values influence future values, making it ideal for predicting equipment behaviour over time.
- **Prophet:** Prophet is a model developed by Facebook designed to handle time series data with complex seasonality and missing data points. It is highly flexible, making it suitable for industrial datasets that may have irregular intervals and require robust handling of outliers and seasonal effects.

### 3.3.1 Linear and polynomial regression

Time series forecasting involves predicting future values based on historical observations arranged in a chronological sequence. Two regression models commonly used in this context are Simple Regression and Polynomial Regression Nunno (2014).



Simple Regression entails fitting a straight line to the historical data, assuming a linear relationship between time and observed values. The regression equation takes the form represented in Equation 1.

$$Y = mX + b \tag{1}$$

$Y$  is the predicted value,  $X$  is the time variable,  $m$  is the slope, and  $b$  is the intercept. While straightforward, this method may overlook complex patterns in the data.

Polynomial Regression, on the other hand, offers greater flexibility by fitting a polynomial equation to the data. The general form is represented in Equation 2.

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_nX^n \quad (2)$$

$n$  is the degree of the polynomial, and  $b_0, b_1, \dots, b_n$  are coefficients. This model can capture more intricate patterns but requires careful tuning to avoid overfitting [Nunno \(2014\)](#).

The research delved into Linear and Polynomial models, exploring their performance across varying training dataset sizes to forecast future temperature and pressure trends, with training sizes ranging from 12 measurements (equivalent to 1 h) to 288 measurements (24 h). These are represented in [Figures 6, 7](#).

Linear models demonstrated limitations in capturing complex temperature patterns due to their inherent linearity, resulting in deviations particularly evident during peak oscillations in the dataset. These models struggled to adapt to nuanced temperature trends, highlighting the challenge of representing intricate data patterns with simplistic linear functions.

Conversely, linear models exhibited more stability in pressure predictions, especially with larger training sizes. Tower pressure predictions benefitted from consistent interval patterns, facilitating more accurate forecasts despite weaker data correlations. However, as training sizes increased, linear models encountered challenges adapting to evolving temperature trends, leading to deviations from actual data.

Polynomial models, showcased significant discrepancies across all training sizes, failing to capture dataset intricacies accurately. While there were instances of acceptable performance with smaller training sizes for specific variables, overall, polynomial models exhibited exponential divergence from actual data, raising doubts about their reliability for this dataset.

### 3.3.2 Exponential smoothing

Exponential Smoothing [Bucay-Valdiviezo et al. \(2023\)](#), a statistical technique employed in diverse fields such as finance, economics, marketing, and operations, stands out as a powerful tool for time series forecasting. Operating on the principle of assigning exponentially decreasing weights to past observations, it adapts dynamically to underlying patterns or trends, making it a versatile choice in various forecasting scenarios [Hu et al. \(2013\)](#).

This methodology, designed specifically for univariate time series data, employs a weighted linear sum of past observations or lags, with a crucial element being the assignment of exponentially decreasing weights.

In its simplest form, Simple Exponential Smoothing (SES) utilizes the current observation and the previous forecast to make predictions, controlled by the parameter  $\alpha$ . The SES predicts the next observations based on the smoothed average, without considering the trend and seasonality of the data. The choice of  $\alpha$  follows the principle that a smaller  $\alpha$  places more emphasis on past observations, making the smoothing process more persistent, while a larger  $\alpha$  gives more weight to the most recent observation, making the smoothing process more responsive to changes.

Double Exponential Smoothing (DES), known as Holt's trend model, comes into play when data exhibits a linear trend but lacks a seasonal pattern, introducing an additional parameter,  $\beta$ , to control the trend's influence. Moreover, Triple Exponential Smoothing (TES), also known as Holt-Winters Exponential Smoothing, is applied when data showcases both linear trends and seasonal patterns, introducing the new smoothing parameter,  $\gamma$ .

The advantages of Exponential Smoothing lie in its simplicity, computational efficiency suitable for large datasets, and adaptability to changes in the time series. Its effectiveness with limited historical data, thanks to higher weights assigned to recent observations, further enhances its appeal for real-time forecasting [Hu et al. \(2013\)](#).

With a focus on short-term forecasting situations and the acknowledgment of limitations in long-term reliability, Exponential Smoothing stands out as an effective, adaptive, and computationally efficient tool [Bucay-Valdiviezo et al. \(2023\)](#). The choice of method should be made judiciously based on the context and nature of the time series data, recognizing both its strengths and limitations.

In the case of SES, represented in [Figure 8](#), there exists the option to manually select the desired alpha parameter or allow the model to optimize it autonomously. This flexibility in parameter selection adds an adaptive dimension to the SES model, potentially enhancing its performance in capturing evolving trends within the dataset. However, SES consistently yielded flat forecasts across all predictions. This characteristic stems from SES's inherent simplicity, as it does not incorporate trend or seasonal components.

While SES may perform adequately for pressure predictions, where values typically remain within narrow intervals for extended periods, its efficacy diminishes when forecasting temperatures. Furthermore, SES's inability to capture fluctuations hampers its utility in predictive maintenance, as it cannot effectively anticipate whether values will surpass predefined thresholds, limiting its application in preventing potential system issues.

For the DES model, represented in [Figure 9](#), similar parameter optimization was employed, allowing the model to determine optimal values for  $\alpha$  and  $\beta$  autonomously. DES offers additional configurability, including the specification of trend type (additive or multiplicative) and whether it is damped. In this instance, default settings were retained, corresponding to a non-damped additive trend.

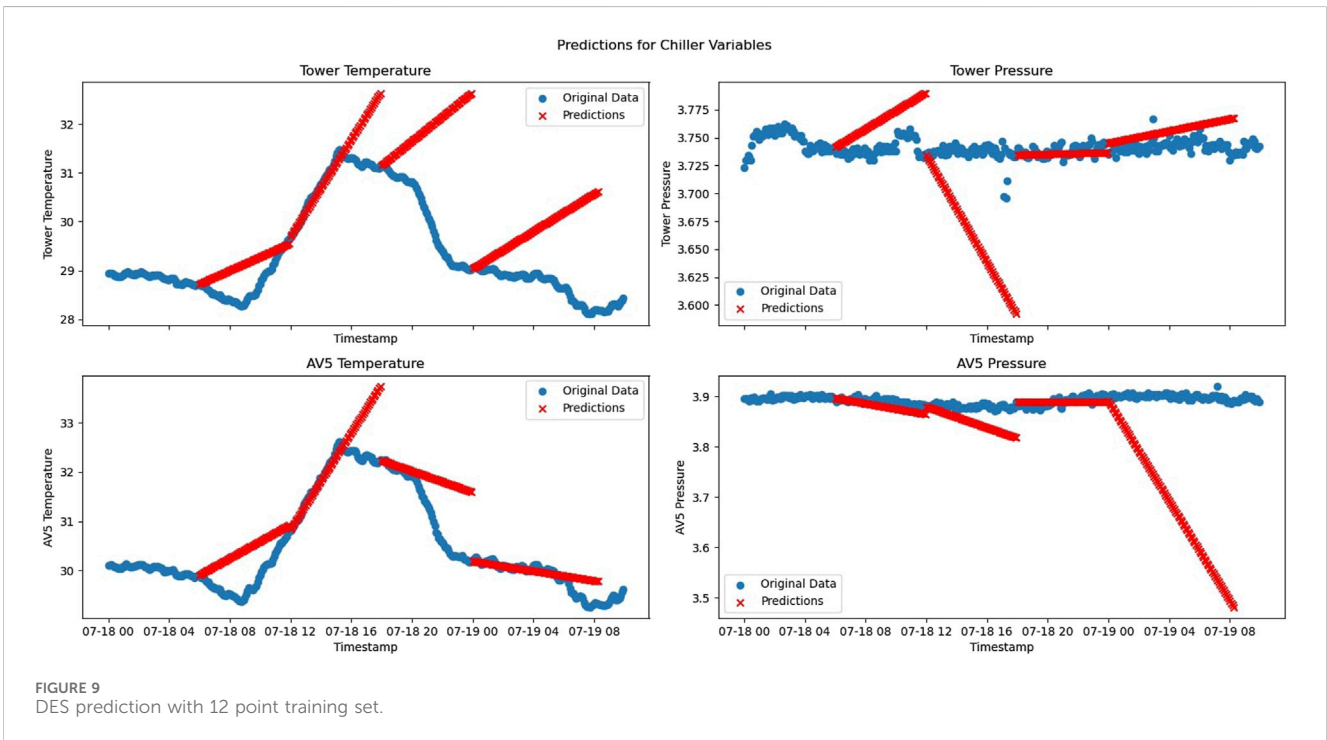
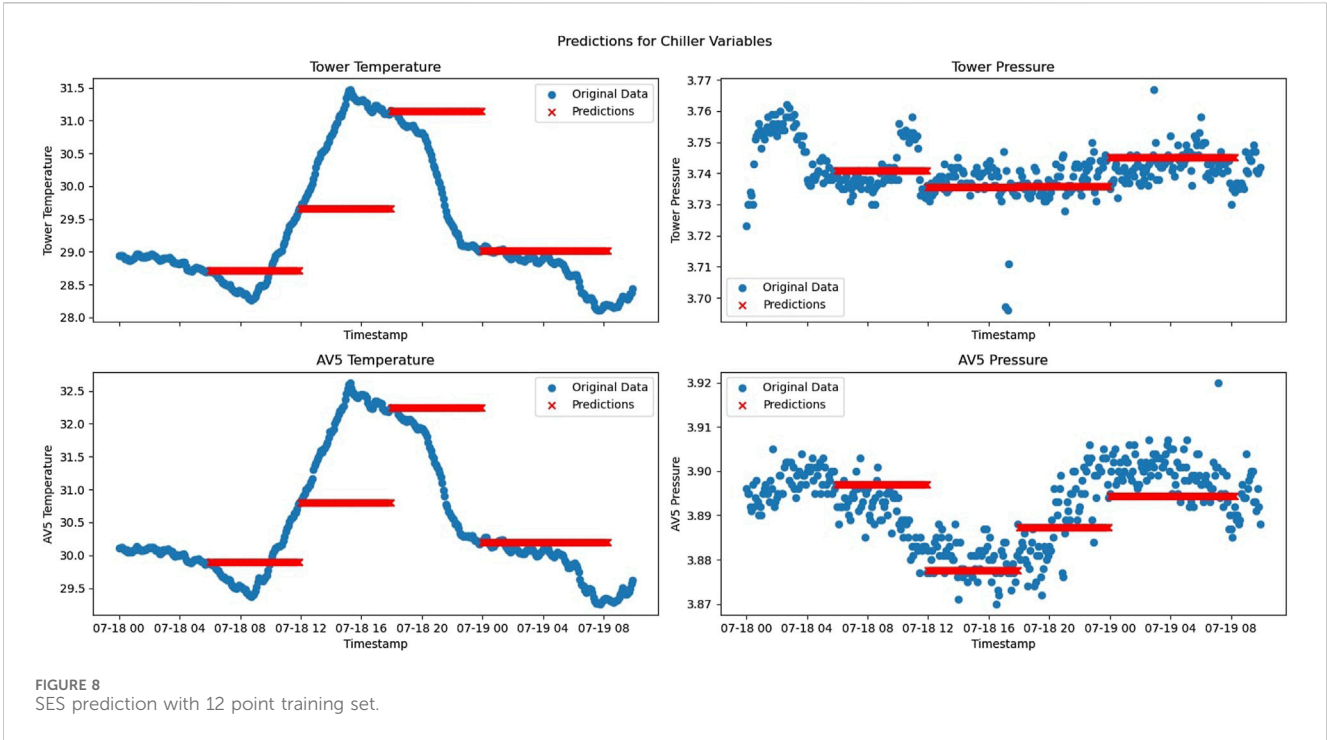
The DES model exhibited performance akin to linear models but notably displayed similarity in predictions with different training data sizes (144 and 288 data points), as represented in [Figure 10](#). This phenomenon can be attributed to the abundance of training data, resulting in very small alpha values that diminish the model's ability to capture underlying patterns effectively. The resemblance in predictions underscores the challenge of balancing sufficient historical data with alpha values to optimize model responsiveness.

### 3.3.3 ARIMA

ARIMA models combine Autoregressive (AR), Integrated (I), and Moving Average (MA) components to analyze and predict future trends in time series data. Central to the configuration of the ARIMA model are the parameters  $p$ ,  $d$ , and  $q$ , each playing a distinct role in capturing the temporal dependencies and achieving stationarity in the time series [Mo \(2023\)](#).

The autoregressive order  $p$  represents the number of lag observations included in the model. In the AR component, the current observation is modelled as a linear combination of its  $p$  past observations. A higher  $p$  allows the model to capture longer-term dependencies, making it adept at capturing trends and periodic patterns within the time series.

The differencing order  $d$  signifies the number of times the time series is differenced to achieve stationarity. Stationarity is a

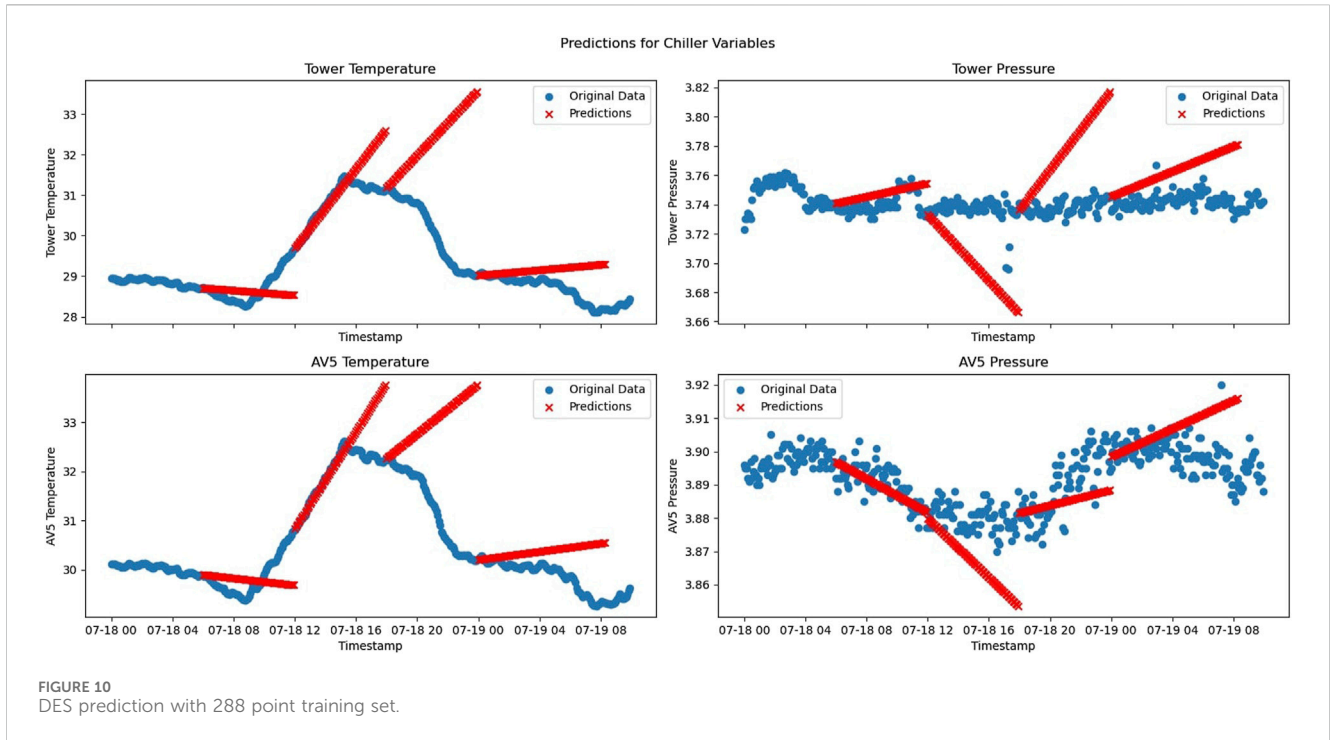


prerequisite for effective time series modelling, and the differencing process eliminates trends or seasonality. A higher  $d$  implies more differencing steps are needed to transform the time series into a stationary form. The determination of  $d$  involves evaluating the number of differencing steps required to achieve stationarity.

The moving average order  $q$  denotes the number of lagged forecast errors included in the model. The MA component

captures the relationship between the current observation and  $q$  past forecast errors. A higher  $q$  enables the model to account for longer-term dependencies in the forecast errors, aiding in capturing any remaining serial correlation in the differenced series.

Careful consideration of  $p$ ,  $d$ , and  $q$  is essential in configuring the ARIMA model to accurately reflect the temporal characteristics



of the time series, ensuring robust forecasting and meaningful insights into future trends.

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) play crucial roles in the thorough analysis of ARIMA models, as they serve as essential tools for identifying the appropriate orders ( $p$ ,  $d$ ,  $q$ ) in the model. The examination of decay or cut-off patterns in these functions enables analysts to discern the temporal dependencies within the time series [Bucay-Valdiviezo et al. \(2023\)](#).

Both ACF and PACF start with a lag of 0, resulting in a correlation of 1 with the time series itself. The 95% confidence interval in ACF and PACF plots helps identify statistically significant correlations. For the AR order, a significant spike at lag  $p$  in the PACF plot, with subsequent lags having insignificant values, suggests an autoregressive order of  $p$ . For the MA order, a significant spike at lag  $q$  in the ACF plot, with subsequent lags having insignificant values, indicates a moving average order of  $q$ . The presence of repeating patterns at regular intervals in the ACF and PACF plots may also suggest the existence of seasonal components.

Still, to perform the ACF and PACF analysis, one needs to assume stationarity, making it necessary first to perform a stationarity check in the dataset. One can be used with the Augmented Dickey-Fuller (ADF) test.

Central to the ADF test are two key components: the  $p$ -value and the null hypothesis ( $H_0$ ). The  $p$ -value is a statistical metric that quantifies the probability of observing the ADF test statistic or a more extreme value under the assumption that the null hypothesis is true. In the context of the ADF test, the null hypothesis posits the presence of a unit root in the time series, indicating non-stationarity [Bucay-Valdiviezo et al. \(2023\)](#). A high  $p$ -value suggests a failure to reject the null hypothesis, implying that the data likely possesses a unit root and is subsequently non-stationary. Conversely, a low

$p$ -value provides evidence to reject the null hypothesis, signalling that the data lacks a unit root and is stationary.

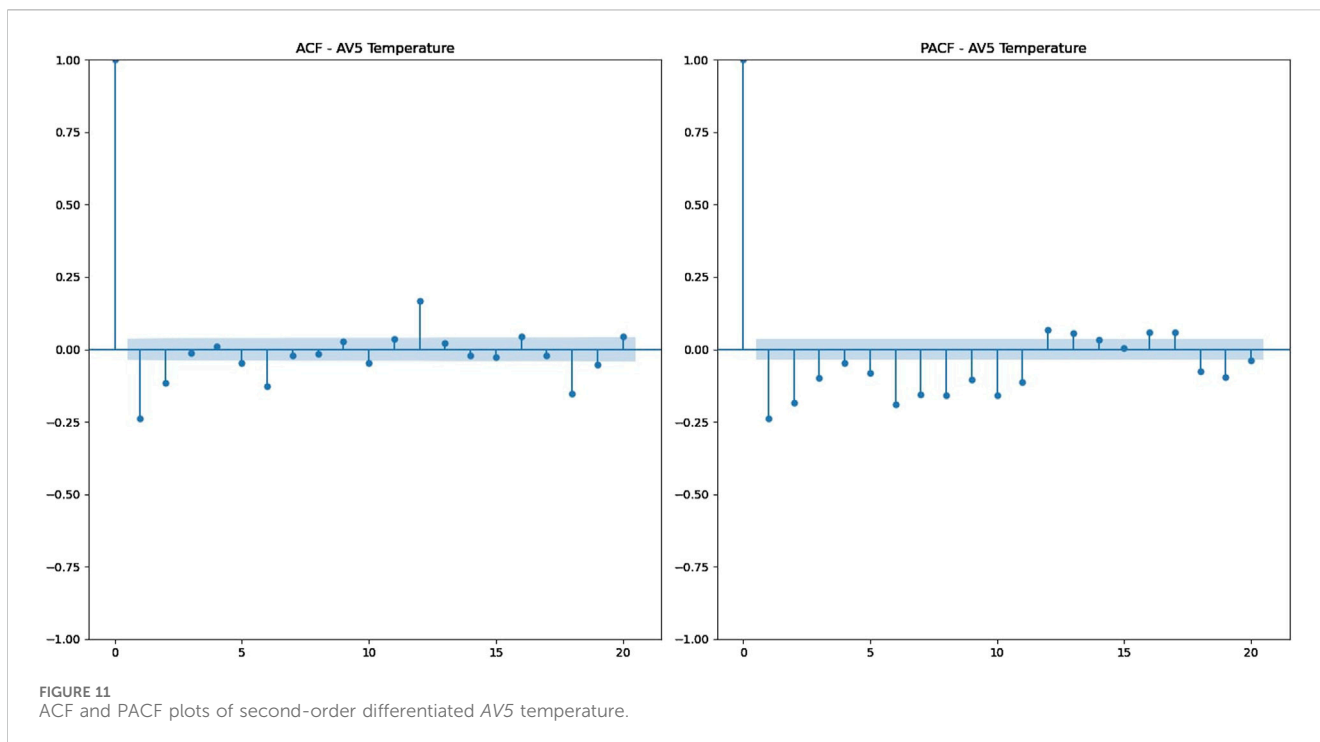
In more practical terms, the interpretation of the  $p$ -value is deeply intertwined with the decision-making process. If the  $p$ -value exceeds the chosen significance level, often set at 0.05, one fails to reject the null hypothesis and the time series is deemed likely non-stationary. On the other hand, if the  $p$ -value is equal to or below the significance level, it provides evidence to reject the null hypothesis, indicating stationarity. This pivotal outcome sets the stage for subsequent analyses, leading to the ACF and PACF examination, as explained before.

In implementing the ARIMA forecasting model, the initial steps involved conducting a stationarity test on the dataset using the ADF test and examining ACF and PACF plots to inform model development [Bucay-Valdiviezo et al. \(2023\)](#).

Given the dataset's potential seasonal component and the desire to accommodate an ARIMA model, a subset of 864 measurements (equivalent to 3 days) was selected to adequately capture seasonal cycles. The ADF test revealed that temperature datasets demonstrated stationarity with  $p$ -values below 0.05, while pressure datasets for the tower and AV5 furnace did not exhibit stationarity, necessitating differencing. However, inspecting ACF and PACF plots for temperatures revealed slow convergence to confidence intervals, suggesting differencing could benefit these datasets as well.

First-order differencing was applied to AV5 pressure and second-order differencing to AV5 temperature, Tower temperature, and Tower pressure, achieving convergence to ACF and PACF confidence intervals and confirming stationarity through ADF tests, as represented in [Figure 11](#). These steps informed the determination of the differencing parameter ( $d$ ) for each time series in the ARIMA model.

Further examination of ACF and PACF plots informed the selection of parameters ( $p$  and  $q$ ) for each model, with all plots



cutting off at lag 2 and displaying geometric decay in PACF patterns. Consequently,  $q$  was set to 2, and  $p$  to 0 for all models.

The ARIMA models generated forecasts characterized by trend lines similar to those from Linear Regression and DES, with more robust performance observed in forecasting pressures, potentially attributed to pressure values maintaining consistency around their mean despite significant fluctuations. This underscores the ARIMA model's ability to capture and predict such patterns effectively.

### 3.3.4 Prophet

The Prophet model [Riady \(2023\)](#); [Caicedo-Castro \(2023\)](#), an innovative forecasting tool, was introduced by Facebook's Core Data Science team in 2018. Developed as an open-source tool, it was originally designed for forecasting daily data but has since been extended to accommodate a diverse range of seasonal patterns. This model encompasses a piecewise linear trend, seasonal variations, holiday effects, and a white noise error term.

The key components of the model delve into the intricacies of its structure. The piecewise linear trend, with automatic changepoint selection and an optional logistic function for an upper trend bound, offers a dynamic representation of temporal evolution. Seasonal components, expressed through Fourier terms with default orders for annual and weekly seasonality, capture recurring patterns. Holiday effects are introduced as simple dummy variables, contributing to a comprehensive modelling approach.

The Prophet model employs a Bayesian approach, enabling the automatic selection of changepoints and enhancing adaptability to diverse datasets for robust forecasting. It serves as a versatile tool for time series forecasting, demonstrating resilience to missing data, trend shifts, and outliers [Sharma et al. \(2022\)](#).

The final model selected for its simplicity of implementation and rapid performance is the Prophet model, leveraging the *prophet* library. This model, designed to handle the seasonal component of

the dataset, benefits from a larger training dataset, utilizing the 864 data points for training, as represented in [Figure 12](#).

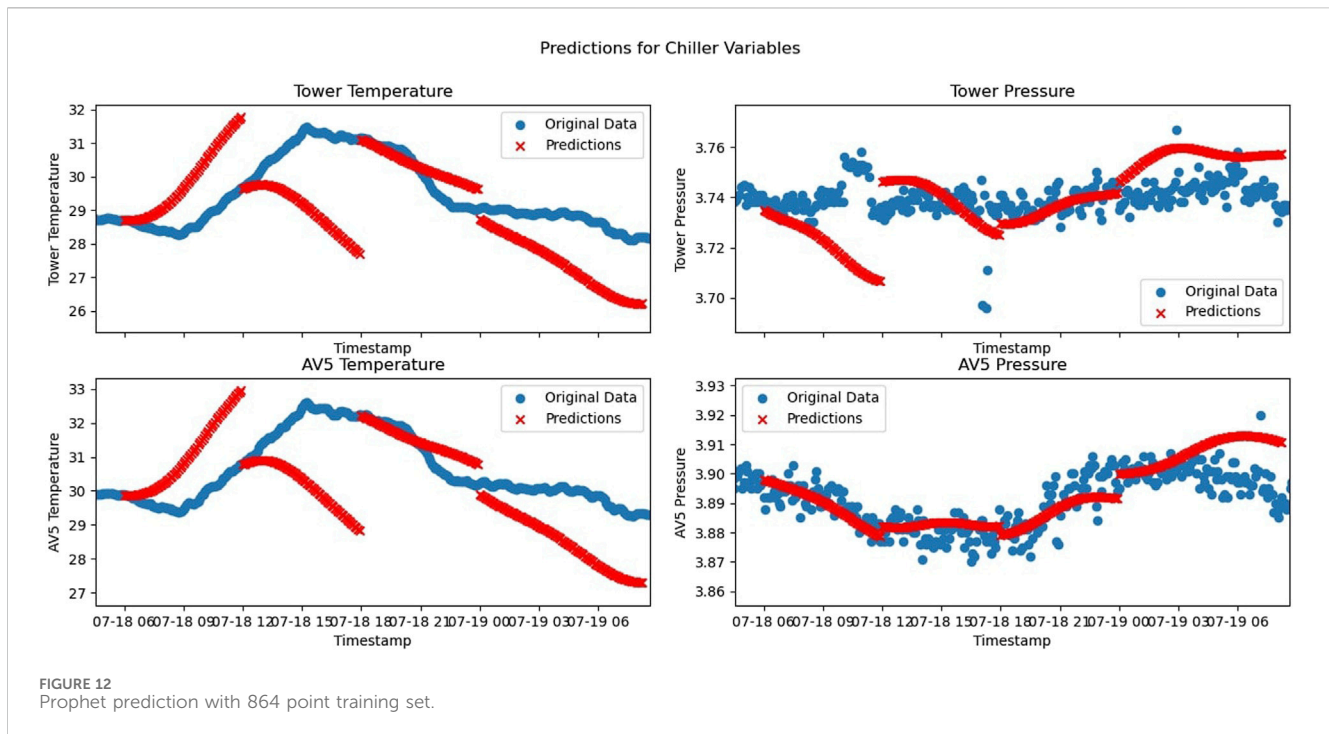
Upon reviewing the documentation, it becomes apparent that several parameters can be customized by the user, including the changepoint scale, seasonality scale, uncertainty samples, and more. For the preliminary model configuration, only the *daily\_seasonality* was set to *true*, while *weekly\_seasonality* and *yearly\_seasonality* were set to *false*. Other parameters were left at their default values.

In the observed graphs, the forecasted lines exhibit behaviour relatively similar to the actual data. However, they tend to curve too early. Further exploration and fine-tuning of the Prophet model, in conjunction with an optimized training size, could potentially enhance its accuracy in capturing the intricate patterns present in the temperature and pressure datasets. This iterative process of refinement is essential for ensuring the model's effectiveness and reliability in predictive maintenance applications.

## 4 Results

For the evaluation of the models, an algorithm was developed to calculate performance metrics. The algorithm developed for model evaluation is designed to assess the performance of various predictive models by systematically exploring different combinations of model parameters and training sizes. The core steps involved in the algorithm's realization are as follows:

1. Model and Parameter Selection: The algorithm begins by allowing the user to pre-select the predictive models of interest (e.g., Linear Regression, Polynomial Regression, Simple Exponential Smoothing, Double Exponential Smoothing, ARIMA, Prophet) and their associated parameters. For each model, the user specifies a range of



parameters to be tested. For instance, in the case of ARIMA, parameters such as  $p$ ,  $d$ ,  $q$  values are selected, while for Exponential Smoothing models, parameters like  $\alpha$ ,  $\beta$ , and trend dampening are chosen.

2. Training Size Specification: The user also selects the range of training sizes to be evaluated. The training size represents the number of data points used to train the model before testing it on the subsequent data points. The algorithm iterates through all possible combinations of the selected training sizes with the model parameters.
3. Blocked Cross-Validation with Rolling Window Strategy: Unlike traditional rolling window methods, where the training set incrementally expands, the algorithm employs a blocked approach. Each training block has a fixed size as defined by the user, and both the training and testing sets move across the time series data in a sequential manner. Specifically, the training block is defined by the current combination of parameters and training size, while the testing block immediately follows the training block. After each evaluation, the blocks are shifted by one data point along the time series, maintaining the training block's fixed size. This method simulates real-time data processing, where models are constantly updated with the latest available data and then used to predict the next set of outcomes.
4. Performance Metric Calculation: For each combination of model parameters and training sizes, the algorithm calculates performance metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics are key indicators of how well the model predicts future data points based on the training it has received. MSE provides a measure of the average squared difference between observed and predicted values, penalizing larger errors more heavily. RMSE is the square root of MSE,

offering a measure in the same units as the data. MAE provides the average magnitude of the errors, giving a more interpretable measure of prediction accuracy.

5. Result Exportation: Once all combinations have been evaluated, the results, including each model's parameters, training size, and corresponding performance metrics, are exported to a CSV file. This comprehensive CSV file allows for easy comparison of models and their performance under different configurations, facilitating the selection of the best-performing model and parameter set for the specific time series data.
6. Optimal Model and Parameter Identification: By reviewing the CSV file, the user can identify which model, training size, and parameter combination provided the most accurate predictions (i.e., the lowest MSE, RMSE, and MAE). This insight is crucial for determining the model that best captures the patterns in the dataset, ensuring that the predictive maintenance system is as effective as possible in anticipating equipment failures.

This approach reflects the sequential nature of the time-series data, serving as a simulator of the models' behaviour for the dataset, as if they were encountering continuous data in real time. With this methodology, it becomes possible to discern which model outperformed the others. Simultaneously, it facilitates the determination of the optimal training size and identification of the parameter combination that resulted in the most accurate predictions. Following the observations made in Section 3.3, the following combinations were considered for each model.

For the Linear and Polynomial Regression models, the only parameter to consider is the training size. As observed earlier, a larger training size tends to yield poorer model responses. A training size ranging from 6 measurements (30 min) to 96 measurements (equivalent to 8 h) was chosen for evaluation.

For the DES model, the parameters considered include  $\alpha$ ,  $\beta$ , trend, and damped\_trend. As highlighted before, there exists a correlation between  $\alpha$  and the number of training sets. A smaller  $\alpha$  value necessitates a larger training set for cumulative weights to approach 1. However, an excessively high training set for a small  $\alpha$  may result in the neglect of many points due to their minimal weight.

To address this, a code was developed to determine the number of measurements required to achieve a cumulative weight of 0.999 for each  $\alpha$  value. The results were rounded to the nearest multiple of 10 to enhance the proximity of the cumulative weight to 1.

The obtained results are summarized as follows:

- For an  $\alpha$  of 0.1, the algorithm considers the previous 70 measurements.
- An  $\alpha$  of 0.2 takes into account the latest 40 points.
- $\alpha$  of 0.3 and 0.4 consider the last 20 measurements.
- An  $\alpha$  greater than or equal to 0.5 considers only the previous 10 points.

However, the  $\beta$  component exhibits a similar relationship with the training size as the  $\alpha$ . Consequently, the algorithm selects the training size as the larger value needed between the  $\alpha$  and  $\beta$  parameters.

In the ARIMA model, the parameters consist of the components  $p$ ,  $i$ , and  $q$ . Given that the  $i$  component should remain unchanged to prevent making the data non-stationary, slight adjustments were made only to the  $p$  and  $q$  parameters. As for the training size, choices were made for half a day, 1 day, 2 days, and 3 days.

Finally, in the Prophet model, aiming to capture seasonal patterns, the same training sizes as the ARIMA model were selected. Parameters such as *changeoint\_prior\_scale* and *seasonality\_prior\_scale* were also considered, exploring values around their defaults, 0.05 and 10, respectively.

Regarding regression assessment metrics, smaller MSE, RMSE, and MAE collectively suggest a reduction in the magnitude of prediction errors. MSE measures the average squared magnitude of errors, RMSE quantifies the average magnitude of errors by taking the square root of MSE, and MAE provides the average magnitude of prediction errors. Smaller values for these metrics indicate that, on average, the model's predictions are closer to the actual observed values, reflecting improved predictive performance.

## 4.1 Evaluation analysis

As described before, a variety of models were employed to predict each measurement in the AV5 Water Cooling System, namely the pressure and temperature. The models predicted the temperature on both cold and hot sides, i.e., the AV5 Water Tower and the AV5 Furnace. This diversified configuration has resulted in the implementation of the alarm system with the devised algorithm at the Avintes plant, providing accurate and reliable predictions for the given scenario.

Beginning with Linear and Polynomial Regression models, the Linear Regression demonstrated much better performance, as expected. The metrics started to decrease with an increase in the

training size and reached a minimum point at a training size of 18 measurements for both temperatures, equivalent to an hour and a half. After this training size, all subsequent training sizes increased the RMSE and MAE values. For tower pressure, this occurred with a larger size of 78 measurements, and for AV5 inlet pressure, it occurred with 12 measurements.

Moving on to the DES, coincidentally, for all temperatures and pressures, the values of  $\alpha$  at 0.2,  $\beta$  at 0.9, and a damped additive trend showed the lowest RMSE. This configuration also resulted in the lowest MAE. In temperature, this model seems to perform better in AV5 temperature, with lower metrics in general. For pressures, the values are so close that the difference is negligible.

A worse performance was noted with the increase of both  $\alpha$  and  $\beta$ . The higher they increase, the smaller the training size, which causes the model to consider very few points, leading to a disregard for the dataset's trend. It was also observed that the performance is generally better with an additive trend and damped. Most of the non-damped multiplicative trends showed the worst performance.

Considering the ARIMA model, for AV5 Pressure and Tower Pressure, an increase in the training size generally corresponds to improved predictive accuracy, with the best performance observed at a training size of 864 and 576, respectively. AV5 Temperature and Tower Temperature found the best performance again with a training size of 288, validating the idea that the best univariate model will be the same for both.

Finally, the Prophet model demonstrated better performance for increasingly larger training set values for all variables. Indeed, the training set for each variable should be 864. As expected, pressures showed the lowest RMSE and MAE values, with the model having *changeoint\_prior\_scale* of 0.01 and *seasonality\_prior\_scale* of 15 showing the best performance for both pressures. For temperatures, this time they had different parameters for better performance: *changeoint\_prior\_scale* of 0.1 and *seasonality\_prior\_scale* of 5 for AV5 temperature, and *changeoint\_prior\_scale* of 0.01 and *seasonality\_prior\_scale* of 15 for the tower side temperature.

By pitting these models against each other, one can scrutinize the performance metrics presented in Table 2 for the AV5 Water Tower and Table 3 for the AV5 Furnace. These metrics were extracted from the output CSV files, sorted in ascending order of RMSE, MSE, and MAE.

For the AV5 pressure, that is the pressure of the hot side of the heat exchanger, the best model is Prophet, with a training size of 864 points, a *changeoint\_prior\_scale* of 0.01 and *seasonality\_prior\_scale* of 15. For the AV5 temperature, the best model is the DES, with a training size of 40 measurements, a damped additive trend, an  $\alpha$  value of 0.2, and a  $\beta$  value of 0.9.

For the Tower pressure, that is the pressure of the cold side of the heat exchanger, the best model is ARIMA, with a training size of 576 points and  $p$ ,  $i$ , and  $q$  values of 0, 2, and 2, respectively. For the Tower temperature, the best model is, again, the DES, with a training size of 40 measurements, a damped additive trend, an  $\alpha$  value of 0.2, and a  $\beta$  value of 0.9.

## 5 Discussion

This study aimed to enhance predictive maintenance at BA Glass by addressing two primary research questions: RQ1. identifying the



TABLE 2 Top model performances for AV5 pressure and temperature prediction.

Dataset	Model and parameters	Train size	MSE	RMSE	MAE
AV5 pressure prediction	Prophet (0.01; 15)	864	8.42216E-05	0.009177234	0.009017344
	Arima	864	0.000257013	0.016031612	0.014458356
	Linear	12	0.000376765	0.019410441	0.017486384
	DES (0.2, 0.9, True, "add")	40	0.00102805	0.032063225	0.034972768
	Quadratic	18	0.001798349	0.042406945	0.039192513
AV5 temperature prediction	DES (0.2, 0.9, True, "add")	40	1.455394677	1.206397396	1.017156406
	Arima	288	2.961623158	1.720936709	1.422516798
	Linear	18	2.981266012	1.726634302	1.419559487
	Prophet (0.1; 5)	864	5.90992663	2.431034066	2.312245882
	Quadratic	96	10.12327961	3.181710171	2.621386333

TABLE 3 Top model performances for tower pressure and temperature prediction.

Dataset	Model and parameters	Train size	MSE	RMSE	MAE
Tower pressure prediction	Arima	576	0.000122684	0.011076302	0.010207622
	Linear	78	0.000403563	0.020088886	0.019948447
	Prophet (0.01; 15)	864	0.000623703	0.024974049	0.020875929
	Quadratic	96	0.00885642	0.094108557	0.072050938
	DES (0.2, 0.9, True, "add")	40	0.035425682	0.188217114	0.078385025
Tower temperature prediction	DES (0.2, 0.9, True, "add")	40	1.629329848	1.276452055	1.071695927
	Linear	18	3.15036397	1.774926469	1.456565323
	Arima	288	4.492378029	2.119523066	1.703327873
	Prophet (0.01; 15)	864	5.666512855	2.380443836	2.062677351
	Quadratic	96	11.00533821	3.317429459	2.732734257

most suitable types of sensors for the diverse machinery within the BA Glass Avintes factory floor; and RQ2. determining the most accurate prediction models for predicting the state of components in the Water Cooling System of the AV5 furnace.

Regarding RQ1, we identified two sensors as particularly suitable for the AV5 production line machines: the *Sure Cross*<sup>®</sup> *QM30VT2 Vibration and Temperature Sensor* and the *WISE-2410 LoRaWAN Wireless Condition Monitoring Sensor*. The *QM30VT2* was selected for its Modbus compatibility, allowing direct connection to a Modbus-485 network, making it a reliable and cost-effective option. In contrast, the *WISE-2410* offered the advantage of a wireless, battery-operated design, providing flexibility in placement but requiring additional infrastructure for connectivity.

The discussion regarding the types of sensors for factory floor sensorization led to the acquisition and subsequent exploration of two sensor options: wired sensors with RS 485 connection and wireless sensors directly linked to a Modbus TCP gateway. This analysis highlights two key considerations, with the initial being focused on financial aspects. The acquisition of wired sensors

presents the advantage of lower costs and eliminates the need for additional investments in gateways. However, the connection of multiple wireless sensors to a single gateway can lead to a reduction in the cost per sensor over time. Nevertheless, if each wireless sensor is more expensive than its wired counterpart, the financial advantage diminishes, and wired sensors maintain a preferable position.

The subsequent crucial consideration, and what emerges as the genuine advantage of wireless sensors, lies in their ease of installation, mitigating concerns associated with managing extensive cables on the plant floor. The implementation of the wired sensor encountered challenges, particularly in extending cables to reach the sensor and connecting it to power outlets. Additionally, the presence of exposed cables introduces a potential risk of inadvertently damaging the sensors. In contrast, wireless sensors distinguish themselves with a straightforward implementation and removal process, along with the optimization of space on the plant floor.

This dichotomy essentially boils down to a choice between cost and practicality, depending on the short-term preferences of BA Glass. However, considering the goal of achieving predictive

maintenance, this conducted research, encompassing the search, implementation, and integration of sensors into *PowerStudio* SCADA, suggests that transitioning to a completely wireless sensorization would be more advantageous in the long-term context.

For the RQ2, we evaluated various predictive models for forecasting the state of the AV5 furnace's water cooling system components. The Prophet model proved to be the best for predicting the pressure on the hot side of the heat exchanger, while the DES model was most effective for predicting temperature on both the hot and cold sides. The ARIMA model was found to be the best for predicting pressure on the cold side of the heat exchanger.

Before deploying any forecasting models, gaining a comprehensive understanding of the dataset's condition is paramount. This necessitates the availability of an extensive dataset that enables the observation of patterns and trends over a significant period. Additionally, attention should be given to outliers, acknowledging their role not only as representations of measurement errors or noise but also as indicators of human intervention, whether in the shutdown of machines or engagement in maintenance processes. Striking the right balance between the dataset's expansiveness for pattern detection and the detection and removal of outliers is vital for ensuring the accuracy and reliability of predictions.

In the context of predictive maintenance models, it becomes evident that the preference for different models is contingent upon the specific characteristics of the dataset. For forecasting models applied to datasets exhibiting daily seasonality, such as temperature data with a slow and steady oscillation of values over time, the recommendation leans towards using DES with an additive damped trend. Additionally, linear models tend to perform well with these datasets and show a preference for smaller training sizes.

On the contrary, when dealing with more stable datasets characterized by lower amplitude oscillations in the trend but occurring at higher frequencies, exemplified by pressure data, the recommendation shifts towards the utilization of ARIMA or Prophet models, especially with a more substantial training size. In these datasets, avoiding a small-sized training set is advised, as it might only capture rapid variations, leading to inaccurate predictions of future values.

## 5.1 Feedback from BA Glass operational staff

While this study presents a comprehensive analysis of predictive maintenance implementation at BA Glass, an essential aspect that could further validate the project's success is feedback from the operational staff who interact directly with the system. Although this feedback was not formally gathered and analyzed as part of the current study, preliminary discussions with BA Glass operational staff have provided some valuable insights.

The operational staff highlighted several strengths of the implemented predictive maintenance system:

- **Enhanced Decision-Making:** Staff members reported that the integration of the predictive maintenance algorithm with the *PowerStudio* SCADA system significantly improved their ability to anticipate and address potential equipment failures. The real-time alerts and data-driven insights

enabled them to make more informed decisions, thereby reducing unplanned downtime and improving overall operational efficiency.

- **Ease of Use:** The usability of the alarm system was particularly noted. Workers appreciated the clear, actionable alerts generated by the SCADA system, which translated complex predictive data into simple, understandable notifications. This user-friendly interface minimized the learning curve and allowed staff to quickly adapt to the new system.
- **Operational Efficiency:** Feedback also indicated that the predictive maintenance approach helped streamline maintenance processes. By shifting from a reactive to a proactive maintenance strategy, the operational staff observed a reduction in emergency repairs and unscheduled maintenance tasks, which previously disrupted production schedules.

However, the feedback also identified areas for potential improvement:

- **Sensor Performance Monitoring:** Some staff expressed a desire for more detailed real-time monitoring of sensor performance. While the current system provides essential data, they suggested that additional diagnostics on sensor health and functionality could further enhance the system's reliability.
- **Comprehensive Training:** Although the system was designed with usability in mind, there were suggestions for more comprehensive training sessions. Operational staff indicated that while they could navigate the system effectively, deeper training on interpreting the predictive analytics and understanding the underlying models could empower them to use the system more effectively.
- **Scalability and Expansion:** Finally, staff members were interested in expanding the predictive maintenance approach to other areas of the factory beyond the AV5 furnace. They recognized the value of the system and suggested that similar implementations in other critical machinery could further reduce downtime and enhance operational efficiency across the entire facility.

## 5.2 Contextualizing the findings

The findings align with previous research emphasizing the importance of sensor selection and predictive modelling in enhancing maintenance strategies. For instance, studies have highlighted the role of sensor compatibility and data transmission capabilities in effective predictive maintenance [Zonta et al. \(2020\)](#); [Basri et al. \(2017\)](#). Our study reinforces these insights by demonstrating the practical application of these principles in a real-world industrial setting.

Moreover, the identification of the most suitable predictive models for different components of the water cooling system adds to the existing body of knowledge. Prior research has explored various forecasting models for industrial maintenance, but our study provides specific insights into the effectiveness of models like Prophet, DES, and ARIMA in the context of a glass manufacturing plant's cooling system [Addo-Tenkorang and Helo \(2016\)](#); [Plunkett et al. \(2013\)](#).

This study advances the current understanding of predictive maintenance by providing a detailed analysis of sensor suitability and predictive model accuracy in a specific industrial context. The integration of real-time monitoring and predictive analytics, as demonstrated in this study, highlights the potential for proactive maintenance strategies to reduce downtime and operational costs. The practical insights gained from the sensor selection and model evaluation processes can inform similar initiatives in other industrial settings, contributing to the broader adoption of predictive maintenance practices.

By addressing the specific needs of a glass manufacturing plant, this study provides a concrete example of how predictive maintenance can be tailored to meet the unique challenges of different industrial environments. The findings underscore the importance of selecting appropriate sensors and predictive models based on the specific characteristics of the machinery and operational processes. This tailored approach can significantly enhance the effectiveness of predictive maintenance strategies, leading to improved reliability, efficiency, and safety in industrial operations.

### 5.3 Potential shortcomings and limitations

While the study provides valuable insights, several limitations should be considered. First, the evaluation of sensors was based on specific operational needs and environmental conditions of the *BA Glass* Avintes factory. The findings may not be universally applicable to other industrial settings with different requirements and conditions. Furthermore, insufficient sensor data related to the AV5 Water Cooling System constrained the exploration of a diverse range of failure cases. The analysis could only encompass instances of failure in the pumps on the cold side of the heat exchanger. This restriction hindered the ability to investigate a broader spectrum of failure scenarios, which could have provided more comprehensive insights into the predictive maintenance algorithm's performance.

Second, the predictive models were tested using historical data from the AV5 furnace's Water Cooling System, which may not fully capture all potential future scenarios. Additionally, the study did not account for potential external factors, such as changes in operational practices or environmental conditions, which could impact the predictive accuracy of the models.

Finally, in terms of predictive models, the study did not utilize multivariate models. Despite observing strong relationships between temperatures and a moderate correlation with tower pressure, we failed to explore multivariate models. This limitation highlights the potential for further research to compare the effectiveness of univariate and multivariate models, providing a more comprehensive understanding of their applicability in predictive maintenance scenarios.

## 6 Conclusion

This study aimed to enhance predictive maintenance practices at *BA Glass* by addressing two critical research questions: RQ1. identifying the most suitable sensors for diverse machinery within

the *BA Glass* Avintes factory floor; and RQ2. determining the most accurate prediction models for the AV5 furnace's Water Cooling System components. The investigation provided several key insights and contributions to the field of predictive maintenance.

Firstly, the evaluation of sensors led to the identification of the *Sure Cross*<sup>®</sup> *QM30VT2 Vibration and Temperature Sensor* and the *WISE-2410 LoRaWAN Wireless Condition Monitoring Sensor* as particularly suitable for the factory's needs. The *QM30VT2* offered a reliable, cost-effective solution with direct Modbus compatibility, while the *WISE-2410* provided flexibility through its wireless, battery-operated design, albeit with additional infrastructure requirements.

Secondly, the study identified the most effective predictive models for different components of the AV5 furnace's Water Cooling System. The Prophet model was found to be optimal for predicting pressure on the hot side of the heat exchanger, whereas the DES model excelled in forecasting temperatures on both the hot and cold sides. The ARIMA model proved most suitable for predicting pressure on the cold side.

Despite encountering limitations such as restricted data availability and the exclusion of multivariate models, the study demonstrated the practical application of predictive maintenance strategies in an industrial setting. The integration of real-time monitoring and predictive analytics facilitated a proactive approach to maintenance, aiming to reduce downtime and operational costs.

The findings of this study contribute to the broader understanding of predictive maintenance by providing practical guidance on sensor selection and model implementation. Future research should focus on addressing the identified limitations, particularly by exploring multivariate models and a wider range of failure scenarios to enhance the predictive maintenance framework.

Furthermore, the chosen univariate and multivariate models would be transferred to components with similar behaviour in the factory, such as other furnaces and production lines, to validate whether the insights gained from this modelling context could be applied to other components. This would follow an approach similar to transfer learning, where the knowledge and parameters acquired from one set of components are leveraged to improve the performance of models applied to analogous components. Such an extension of the research could provide valuable insights into the generalizability and adaptability of predictive maintenance models across various machinery and components in this and the other factories of the *BA Glass* group.

Another enhancement would involve establishing a direct connection between the prediction tool and the SCADA system, allowing for the generation of alarms without the need for a CSV file to manually toggle alarm variables. The suggested method for this integration involves the use of an *OPC Router*, facilitating seamless communication between systems.

In conclusion, this study lays a strong foundation for the continued advancement of predictive maintenance practices at *BA Glass* and similar industrial environments. By leveraging the identified sensors and predictive models, organizations can move from reactive to proactive maintenance strategies, ultimately achieving greater operational efficiency and reliability.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

MB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing—original draft. TM: Conceptualization, Project administration, Resources, Supervision, Writing—review and editing. GG: Funding acquisition, Writing—review and editing. RP: Conceptualization, Formal Analysis, Methodology, Supervision, Writing—review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by project openZDM – “Open Platform for Realizing Zero Defects in Cyber Physical Manufacturing” - GA n. 101058673, funded by the European Commission under the Horizon Europe program.

## References

- Addo-Tenkorang, R., and Helo, P. T. (2016). Big data applications in operations/supply-chain management: a literature review. *Comput. Industrial Eng.* 101, 528–543. doi:10.1016/j.cie.2016.09.023
- Alves, J. F. (2012). *BA-Marca com história no vidro de embalagem*. Vila nova de Gaia: BA Vidro, S.A.
- Alves, J. P. (2023). Raw materials process management improvement in the glass industry. Master's thesis, Porto, Portugal: University of Porto.
- Basri, E. I., Razak, I. H. A., Ab-Samat, H., and Kamaruddin, S. (2017). Preventive maintenance (pm) planning: a review. *J. Qual. maintenance Eng.* 23, 114–143. doi:10.1108/jqme-04-2016-0014
- Bucay-Valdiviezo, J., Escudero-Villa, P., Paredes-Fierro, J., and Ayala-Chauvin, M. (2023). Leveraging classical statistical methods for sustainable maintenance in automotive assembly equipment. *Sustainability* 15, 15604. doi:10.3390/su152115604
- Caicedo-Castro, I. (2023). Course prophet: a system for predicting course failures with machine learning: a numerical methods case study. *Sustainability* 15, 13950. doi:10.3390/su151813950
- Cao, K., Liu, Y., Meng, G., and Sun, Q. (2020). An overview on edge computing research. *IEEE access* 8, 85714–85728. doi:10.1109/access.2020.2991734
- Circutor (2013). Powerstudio SCADA user manual.
- Cortés-Leal, A., Cárdenas, C., and Del-Valle-Soto, C. (2022). Maintenance 5.0: towards a worker-in-the-loop framework for resilient smart manufacturing. *Appl. Sci.* 12, 11330. doi:10.3390/app122211330
- Dillon, T., Wu, C., and Chang, E. (2010). “Cloud computing: issues and challenges,” in 2010 24th IEEE international conference on advanced information networking and applications (Ieeee), Perth, WA, April 20–23, 2010, 27–33.
- EN, B. (2017). *13306: 2017; maintenance—maintenance terminology*. London, UK: British Standards Institution.
- Hu, Y., Zhang, H., Li, C., Liu, S., and Zhang, Y. (2013). “Exponential smoothing model for condition monitoring: a case study,” in 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), Chengdu, China, July 15–18, 2013 (IEEE), 1742–1746.
- Huang, S., Wang, B., Li, X., Zheng, P., Mourtzis, D., and Wang, L. (2022). Industry 5.0 and society 5.0—comparison, complementation and co-evolution. *J. Manuf. Syst.* 64, 424–428. doi:10.1016/j.jmsy.2022.07.010
- inray Industriesoftware GmbH (2024). Opc router.
- Laaper, S., Yauch, G., Wellener, P., and Robinson, R. (2018). *Embracing a digital future*. United Kingdom: Deloitte Insights.
- Lu, Y., Zheng, H., Chand, S., Xia, W., Liu, Z., Xu, X., et al. (2022). Outlook on human-centric manufacturing towards industry 5.0. *J. Manuf. Syst.* 62, 612–627. doi:10.1016/j.jmsy.2022.02.001
- Márquez, F. P. G., and Papaalias, M. (2020). Introductory chapter: an overview to maintenance management. *Maint. Manag.* doi:10.5772/intechopen.86892
- Mo, H. (2023). Comparative analysis of linear regression, polynomial regression, and arima model for short-term stock price forecasting. *Adv. Econ. Manag. Political Sci.* 49, 166–175. doi:10.54254/2754-1169/49/20230509
- Müller, J., Balch, C., Bartolucci, C., Buncek, M., Cotta, J., Curaj, A., et al. (2020). *Enabling technologies for industry 5.0—results of a workshop with europe's technology leaders*. Brussels, Belgium: Directorate-General for Research and Innovation.
- Nunno, L. (2014). *Stock market price prediction using linear and polynomial regression models*. Albuquerque, NM: University of New Mexico.
- Pinto, R., Pinheiro, J., Gonçalves, G., and Ribeiro, A. (2023). “Towards industry 5.0: a capacitation approach for upskilling and technology transfer,” in Distributed Computing and Artificial Intelligence, Special Sessions I, 20th International Conference. Editors R. Mehmood, V. Alves, I. Praça, J. Wikarek, J. Parra-Dominguez, R. Loukanova, et al. (Cham: Springer Nature Switzerland), 342–351.
- Pinto, R., Žilka, M., Zanolli, T., Kolesnikov, M. V., and Gonçalves, G. (2024). Enabling professionals for industry 5.0: the self-made programme. *Procedia Comput. Sci.* 232, 2911–2920. doi:10.1016/j.procs.2024.02.107
- Plevris, V., Solorzano, G., Bakas, N. P., and Ben Seghier, M. E. A. (2022). “Investigation of performance metrics in regression analysis and machine learning-based prediction models.” In 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022), Oslo, Norway, June 5–9, 2022 (European Community on Computational Methods in Applied Sciences).
- Plunkett, T., Macdonald, B., Nelson, B., Hornick, M., Sun, H., Mohiuddin, K., et al. (2013). *Oracle big data handbook*. New York, NY: McGraw-Hill Osborne Media.
- Pombal, T., Ferreira, L. P., Sá, J., Pereira, M. T., and Silva, F. (2019). Implementation of lean methodologies in the management of consumable materials in the maintenance workshops of an industrial company. *Procedia Manuf.* 38, 975–982. doi:10.1016/j.promfg.2020.01.181
- Ponte, R., Pinto, R., and Meireles, T. (2024). Energy consumption analysis in scada: a case study in the glass container industry. *Procedia Comput. Sci.* 237, 717–724. doi:10.1016/j.procs.2024.05.158

## Acknowledgments

The authors gratefully acknowledge the support provided by the Foundation for Science and Technology (FCT/MCTES) within the scope of the Associated Laboratory ARISE (LA/P/0112/2020), the R&D Unit SYSTEC through Base (UIDB/00147/2020) and Programmatic (UIDP/00147/2020) funds, and BA Glass Portugal, S. A. as well.

## Conflict of interest

Author TM was employed by BA Glass.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Psarommatis, F., May, G., and Azamfirei, V. (2023). Envisioning maintenance 5.0: insights from a systematic literature review of industry 4.0 and a proposed framework. *J. Manuf. Syst.* 68, 376–399. doi:10.1016/j.jmsy.2023.04.009
- Riady, S. R. (2023). Stock price prediction using prophet facebook algorithm for bbca and tlkm. *Int. J. Adv. Data Inf. Syst.* 4, 1–8. doi:10.25008/ijadis.v4i2.1258
- Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., et al. (2015). Industry 4.0: the future of productivity and growth in manufacturing industries. *Boston Consult. group* 9, 54–89.
- Sharma, K., Bhalla, R., and Ganesan, G. (2022). “Time series forecasting using fb-prophet,” in Algorithms Computing and Mathematics Conference, August 29–30, 2022, 59–65.
- Silvestri, L., Forcina, A., Introna, V., Santolamazza, A., and Cesarotti, V. (2020). Maintenance transformation through industry 4.0 technologies: a systematic literature review. *Comput. Industry* 123, 103335. doi:10.1016/j.compind.2020.103335
- Testa, M., Malandrino, O., Sessa, M. R., Supino, S., and Sica, D. (2017). Long-term sustainability from the perspective of cullet recycling in the container glass industry: evidence from Italy. *Sustainability* 9, 1752. doi:10.3390/su9101752
- Wasaya, A., Malik, S., Abbas, M. Z., and Shahid, H. (2021). “An approach towards prepaid metering system using PowerStudio SCADA,” in 2021 4th International Conference on Energy Conservation and Efficiency (ICECE), Lahore, Pakistan, March 16–17, 2021 (IEEE), 1–5. doi:10.1109/ICECE51984.2021.9406294
- Xu, X., Lu, Y., Vogel-Heuser, B., and Wang, L. (2021). Industry 4.0 and industry 5.0—inception, conception and perception. *J. Manuf. Syst.* 61, 530–535. doi:10.1016/j.jmsy.2021.10.006
- Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., and Li, G. P. (2020). Predictive maintenance in the industry 4.0: a systematic literature review. *Comput. and Industrial Eng.* 150, 106889. doi:10.1016/j.cie.2020.106889