

Article

Not peer-reviewed version

Enhancing Asset Reliability and Sustainability: A Comparative Study of Neural Networks and ARIMAX in Predictive Maintenance

[Salvador Perez-Garcia](#)*, [Cristina González-Gaya](#), [Miguel A. Sebastián](#)

Posted Date: 3 April 2025

doi: 10.20944/preprints202504.0230.v1

Keywords: sustainable maintenance; Industry 4.0; ARIMAX; neural network; hyperparameter



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Enhancing Asset Reliability and Sustainability: A Comparative Study of Neural Networks and ARIMAX in Predictive Maintenance

Salvador Perez-Garcia *, Cristina González-Gaya and Miguel A. Sebastián

Department of Construction and Manufacturing Engineering, Universidad Nacional de Educación a Distancia (UNED), C/Juan del Rosal 12, 28040 Madrid, Spain

* Correspondence: Correspondence: sperez199@alumno.uned.es

Abstract: Organizations strive to maximize efficiency in their manufacturing processes, yet they must also consider broader repercussions, as industrial activity directly impacts the environment and society. The adoption of innovative technologies and initiatives to mitigate this impact is therefore essential. Traditional asset maintenance plays a critical role in ensuring high equipment availability, but there is a clear need to evolve toward predictive and sustainable maintenance strategies to enhance reliability, safety, and equipment lifespan. This shift redefines maintenance itself, aligning it with Circular Economy principles and Industry 4.0 solutions to prevent unplanned downtime, reduce failures, and improve personnel and facility safety. This research examines the transition from traditional preventive maintenance to predictive and sustainable maintenance in a real-world industrial context, comparing the design of a neural network with the ARIMAX technique to develop reliable predictive models. The study aims to facilitate a paradigm shift by proposing a predictive model that reduces unplanned shutdowns and optimizes spare parts and labor utilization. The practical application focuses on a hydrogen compressor in the petrochemical industry, demonstrating the model's potential for operational and sustainability improvements.

Keywords: sustainable maintenance; Industry 4.0; ARIMAX; neural network; hyperparameter

1. Introduction

In September 2015, the United Nations General Assembly adopted the 2030 Agenda [1], which emphasized the three pillars of sustainable development: environmental protection, economic growth, and social inclusion. Due to limited progress, the follow-up resolution A/RES/79/1 (Pact for the Future) was approved in September 2024 [2], reinforcing the Agenda with heightened commitments to sustainable development, climate action, and digital cooperation. Parallely, industrial transformation advanced through initiatives like Industry 4.0, introduced by the German government at the 2011 Hanover Fair. This strategy aimed to modernize manufacturing via cyber-physical systems, the Internet of Things (IoT), and cloud computing [3]. These developments are particularly relevant to asset maintenance, which traditionally falls into three categories: (1) Corrective maintenance: Reactive repairs after unexpected failures; (2) Preventive maintenance: Scheduled component replacements based on time or usage intervals, and (3) Condition-based maintenance: Real-time monitoring using sensors to predict failures without physical inspection [4].

The rise of Industry 4.0 technologies (e.g., IoT, big data analytics) has revolutionized condition-based approaches by enabling continuous data collection, fault prognosis, and dynamic maintenance planning [5,6]. This shift enhances productivity, sustainability, and asset lifespan estimation while reducing unplanned downtime. This predictive approach is only applicable to monitored equipment, as it relies on sensor-derived data to accurately estimate asset lifespan [5]. By integrating real-time data, maintenance strategies—including condition monitoring, fault diagnosis, prognosis, and planning—are dynamically optimized [6]. The evolution of manufacturing paradigms, from mass

production to Lean Manufacturing, reflects growing societal demands for greener and more sustainable industrial practices. Modern companies must balance economic growth with social and environmental responsibility [7]. Sustainable manufacturing embodies this shift, emphasizing resource efficiency, cleaner technologies, and safeguards for both environmental integrity and human health [8,9]. To achieve these goals, a transition from a Linear to a Circular Economy is essential. This systemic change prioritizes waste reduction through policies promoting recycling, reuse, remanufacturing, and material recovery—all while maintaining economic viability [10–12].

The EN 16646 standard [13] emphasizes that effective maintenance policies must integrate diverse perspectives by understanding the interdependencies between operations and value creation across the asset lifecycle. Transitioning from preventive to predictive and sustainable maintenance is a multi-stage process involving: (1) Redefining key performance indicators (KPIs) and selecting target assets, (2) Sensor selection and deployment, (3) Data collection, analysis, and storage, (4) Predictive model development, (5) Staff training and capability building, (6) Solution validation and assessment, (7) Scaling the implementation to additional assets (as illustrated in Figure 1). For comprehensive guidance on implementation tools, methodologies, and critical success factors, refer to [14–18].

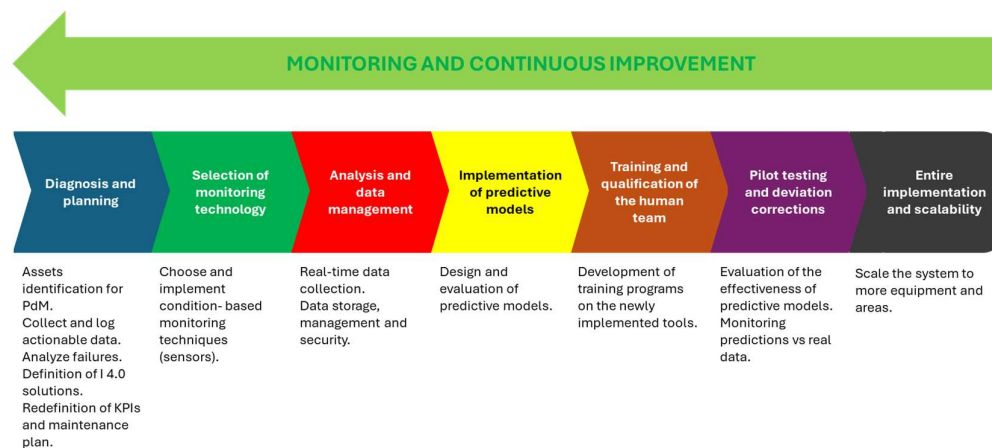


Figure 1. Stages of transitioning from a preventive maintenance model to a predictive one (Maintenance Plan).

1.1. Considerations in the Preparation of the Predictive Maintenance Plan

This paper focuses on predicting the operational behavior of a hydrogen compressor in a petrochemical facility producing caprolactam, a precursor for nylon 6 fibers. Given the continuous nature of the manufacturing process, any component failure can lead to full production stoppages. Unplanned shutdowns entail significant risks, including collateral damage to critical components (e.g., sensors, control valves, pipelines, and electrical systems), delayed customer deliveries, and increased costs for emergency spare parts and overtime labor.

Although the compressor has experienced few failures in recent years, emerging electromechanical issues—such as bearing wear due to overheating and power supply instability—highlight the need for proactive solutions. Additionally, missing data values in other equipment records, attributed to aging communication infrastructure and lack of signal redundancy, further complicate predictive efforts. These challenges have been formally assessed by multidisciplinary teams, confirming the urgency of addressing both hardware and data reliability gaps. Implementing these enhancements in the hydrogen compression process is expected to yield the following benefits:

- **Reduction in Unplanned Downtime:** Minimizing emergency stops caused by malfunctions or missing data values, thereby avoiding subsequent inspections, refurbishments, or replacements of damaged components.
- **Improved Decision-Making for Operators:** Providing reliable, data-driven insights to operators enables controlled shutdowns when necessary. In optimal scenarios, this allows continued operation until scheduled maintenance can address the issue.

- **Enhanced Sensor Infrastructure:** Rewiring sensors to ensure signal redundancy may require new hardware and software modifications to process duplicated data streams effectively.
- **Increased System Reliability:** Evaluating the acquisition of a twin compressor could further enhance reliability through redundancy, ensuring uninterrupted operation.

Sustainable maintenance management represents a critical strategy for achieving sustainable development [19], requiring organizational shifts in policies, strategies, and asset maintenance practices [20,21]. This approach expands the traditional triple-bottom-line framework of sustainable maintenance [22] by incorporating a fourth, technology-driven dimension [23]. The resulting four dimensions are defined as:

1. **Economic Dimension:** Focused on operational efficiency, downtime reduction, process quality, and overall productivity [24,25].
2. **Environmental Dimension:** Centered on resource optimization and waste minimization [26–29].
3. **Social Dimension:** Encompassing legal compliance regarding employee health/safety and customer obligations [30].
4. **Technological Dimension:** Addressing the integration of advanced technologies into maintenance systems [31–34].

The transition to sustainable maintenance necessitates adopting new performance metrics. Notably, the EN 15341 standard [35] outlines 183 maintenance management indicators, including, 3 directly tied to sustainability, 22 related to health, safety, and environment (HSE), 21 concerning personnel competence, and 20 focused on Industry 4.0 technologies. A comprehensive maintenance plan should integrate at least one indicator from each sub-function to enable holistic performance assessment. The adoption of condition-based maintenance is significantly enhanced by Industry 4.0 technologies, which are fundamentally characterized by cyber-physical systems. While multiple definitions exist [36–39], they universally emphasize, the seamless integration of computational and physical systems (mechanical, electrical, and/or human components), their collaborative function in communication and/or control processes, and the capability for real-time data analysis and decision-making.

Unlike traditional statistical forecasting methods such as autoregressive moving average (ARMA) [40], autoregressive integrated moving average (ARIMA) [41], Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) [42] and exponential smoothing [43] which are typically limited as they rely on a sole time series, neural networks provide a promising solution for handling massive data [44], and they are characterized by multiple layers that enable complex feature abstraction from data points [45].

For the purpose of anticipating possible problems, [46] indicates that smart maintenance within Industry 4.0 is oriented towards self-learning, failure prediction, diagnosis, and triggering maintenance programs. At the same time, [47] identifies the suitability of its application since the utilization of supervised models for making forecasts will ensure that the process operates correctly and efficiently without incurring high maintenance costs and reducing product or assets degradation. Machine learning refers to the set of techniques and methods used by computers to autonomously extract complex patterns and relationships from large datasets and make decisions or predictions about real-world events [48]. The concepts used in machine learning are cross-disciplinary and come from computer science, statistics, mathematics, and engineering. The main types of machine learning were outlined by [49], and predictions are made based on new or unanticipated data [50] using labeled dataset. Deep learning is a branch of machine learning that focuses on the learning of data representations and of hierarchical features. With the objective of extracting features, deep learning algorithms employ a system comprising numerous layers of nonlinear processing identities (neurons) that aid in data abstraction [51]. Although recurrent neural networks in their various forms have traditionally been used, architectures equipped with attention mechanisms commonly used in natural language processing, which are more complex and have a higher number of trainable

parameters, should provide better results [52]. The full discussion of Transformer-based neural networks for time series prediction can be found in [53].

For a comprehensive analysis of machine learning strategies and their application to predictive maintenance in sustainable manufacturing, readers may consult references [54,55]. The optimization of neural network hyperparameters is conducted through Design of Experiments (DoE) methodology, with the dual objective of maximizing model performance, while ensuring operational simplicity to facilitate adoption by practitioners with limited technical expertise in machine learning.

The implementation of this combination of solutions requires careful consideration and resolution of several difficulties, including the choice of technology to invest in, the management of current knowledge, its transfer, and the acquisition of new expertise [56] and [57] describes the implementation of a roadmap that optimizes the likelihood that these tools will be applied successfully. These improvements are benchmarked against traditional forecasting methods like ARIMAX, a generalized ARIMA framework that integrates exogenous variables to improve predictive capability.

The evaluation and integration of these new requirements are designed to enhance the likelihood of successfully implementing an intelligent and sustainable maintenance system, as opposed to relying solely on a traditional preventive maintenance approach.

1.2. Contribution of the Paper

In summary, this study aims to optimize resource utilization (technical, economic, and human) through systematic parameterization of a transformer-based neural network, with three key objectives:

Signal Reliability Assessment:

To evaluate whether statistical and technical tools can effectively determine if secondary variables provide reliable predictive information about process behavior, thereby creating signal redundancy.

Computational Efficiency Optimization:

To optimize neural network hyperparameters for maximum computational efficiency, minimizing training/retraining resource requirements, while benchmarking performance against simpler methodologies.

Practical Implementation Evaluation:

To assess the suitability of this forecasting tool for integration into maintenance plans, considering operational complexity and computational demands relative to alternative, simpler approaches.

1.3. Structure of the Document

The structure of this paper is organized as follows: Section 1 presents a comprehensive overview of Industry 4.0's foundations, sustainable manufacturing principles, and maintenance function evolution, highlighting their practical applications to real-world industrial challenges. Section 2 describes the facility's operational characteristics and details the data acquisition methodology from the target equipment. Section 3 provides a technical exposition of the Transformer-based neural network architecture employed in this study. Section 4 presents the statistical optimization framework, including the Design of Experiments methodology. Sections 5 and 6 analyze and discuss the obtained statistical results. Finally, Section 7 synthesizes the key findings, draws substantive conclusions, and proposes actionable directions for future research.

2. Description of the Equipment, Facilities, and Data Collection

A multistage (4-stages) hydrogen compressor is used to compress the gas from 0.034 MPa to 27.95 MPa, cooling the hydrogen after each stage to prevent its temperature from reaching above 423 K. According to [58], the present level of preventive maintenance corresponds to level 3, as there

exists a monitoring and alarm system that alerts operators in the event of any anomalies, in addition to the mandatory preventive maintenance every three months.

The objective is to implement Industry 4.0 technologies with the aim of enhancing maintenance up to level 4 and reducing the over maintenance through the utilization of real-time data monitoring based on predictive techniques. The acquisition of knowledge involves the training of a neural network that employs self-attention mechanisms to identify the hyperparameter configuration that most effectively aligns with the type of data generated by sensors. Due to the enormous volume of real-time sensor data generated in facility operations, a highly specialized distribution control system is implemented in order to enable efficient analysis and decision-making. Figure 2 illustrates how data is collected through the asset's sensors and connected via Ethernet to the control racks, which then send the signal to the unique database, which then shares the data through the OPC-UA server for subsequent processing by users. No data can be managed or stored either on the cloud or external servers due to the organization's current security policy, thus the raw sensor data is transmitted directly to clients, who process the monitoring in real-time through a SCADA system and adjust variable values as required to prevent out-of-range operating situations.

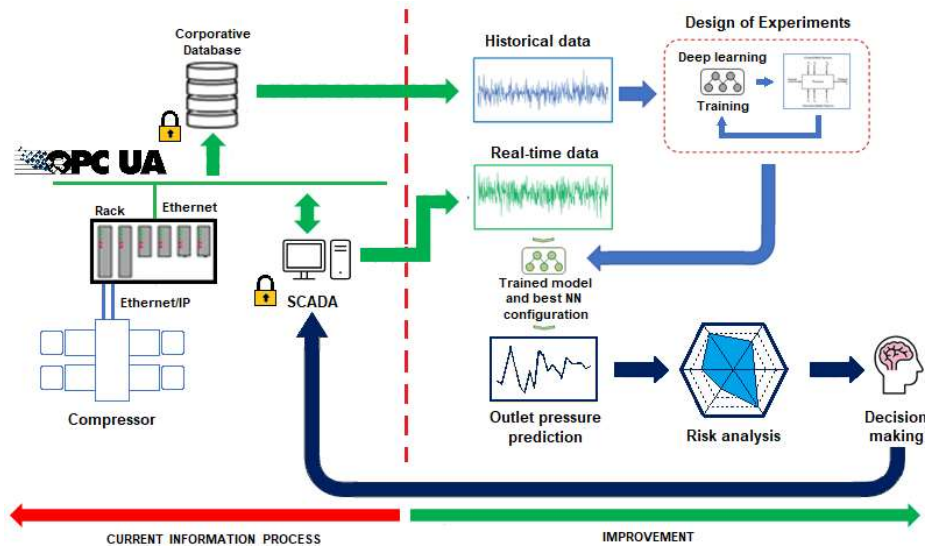


Figure 2. Current and improved information processes.

Two separate wirings comprise the sensors: one monitors internal variables, while the other monitors external ones. The variables are grouped based on the connection in Table 1. The variable to be predicted is "H₂ post intercooling pressure" or simply outlet pressure.

Table 1. Grouping sensors by wiring connection and nominal values of each variable.

External Variables	Internal Variables
Fresh H ₂ flow (500 Nm ³ /h)	H ₂ outlet pressure. Stages 1 – 4
Cooling flow (Nm ³ /h) Undefined.	(0.39/1.97/8.29/28.4 MPa)
H ₂ inlet temperature (318 K)	H ₂ temperature. Stages 1 - 4 (418/398/403/403 K)
Compressor consumption (300 A)	Compression ratio. Stages 1 - 4. (11.5/5.0/4.2/3.5
rpm (min ⁻¹) Undefined.	Dimensionless)
	H ₂ post intercooling pressure. (27.95 MPa)
	H ₂ post intercooling temperature (373 K)

The hydrogen pressure is performed by an absolute pressure stainless-steel transmitter designed for explosion environments. This device features a piezoresistive metallic diaphragm and can operate over 30 MPa with a PROFIBUS – PA compatible analog output (4 – 20 mA). The current

measurement device is a 3-phase connection and 0-400 A measurement range with PROFIBUS-PA compatibility.

The equipment produces data points every minute, with the output pressure nominally set at 27.95 MPa and controlled within limits of 25.99 and 29.91 MPa. With the objective of stabilizing the process and improving the predicting capacities of the neural network, it's necessary to operate on a fifteen-minute basis, and it is imperative to adjust the control limits to align with the new average values. This makes it easier to predict the next few hours of operation, as opposed to predicting the equipment's behavior over the next few minutes. By applying the central limit theorem, the new limits for the average out pressure have been established at 27.44 and 28.45 MPa utilizing the expression $\sigma = (29.91 - 25.99) / (6\sqrt{15})$ to estimate the new standard deviation. Thus, the new limits control will be set at $\pm 3\sigma$.

Controlled and emergency stops are the two types of stoppages that an operator can perform. The latter is carried out in a shorter time and can cause greater degradation of the compressor components. Besides, the training processes for control operators are extensive due to the absence of any technological support for decision-making. Every action requires collaboration with highly experienced operators until trainees attain the necessary expertise and confidence to accurately assess the situation, thereby minimizing the risk of a full system shutdown.

The data presented in Figure 3 reflects fifteen-minute averages of both the outlet hydrogen pressure and the main motor's current consumption. As illustrated, five scheduled equipment shutdowns—associated with preventive maintenance—are evident, with the most prolonged interruption occurring during the summer months.

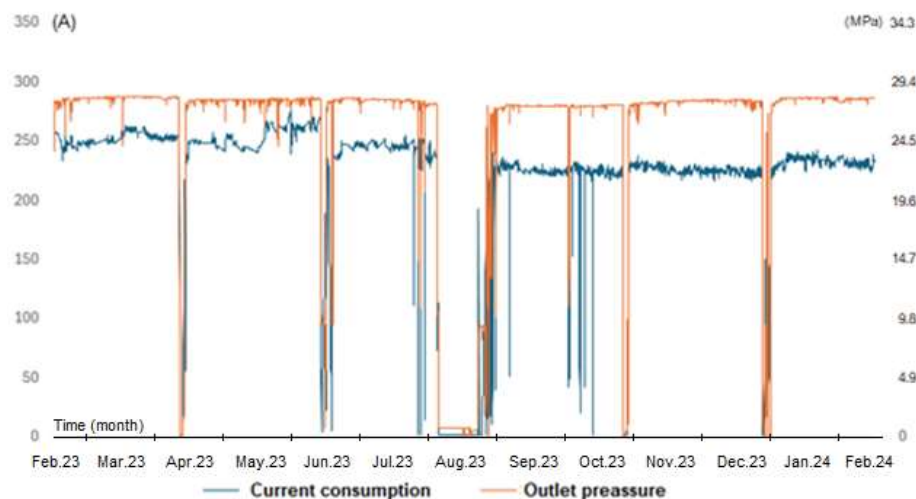


Figure 3. Compressor outlet pressure and current consumption time series.

3. Self-Attention Neural Network

The network known as Transformer, originally designed for natural language processing, uses only attention mechanisms to predict the output sequence instead of Long Short-Term Memory (LSTM) networks along with attention mechanisms of type [59] or [60] as were previously implemented in these types of solutions. The innovative approach of Transformers employs three distinct attention architectures: self-attention for input data in the encoder, masked attention for the predicted sequences in the decoder, and finally, cross-attention, wherein the information from the encoders and the decoders, specifically from masked attention, are combined.

The model inspired by [61] captures dependencies between distinct points in the time sequence, although instead of employing all parts of the Transformer network, including encoders and

decoders, to predict the behavior of time series values, only the self-attention mechanism of the encoder is used along a feed-forward layer to obtain the output tensor.

Therefore, Figure 4 shows the network used, in which the normalized sequence of data to be predicted is introduced into a normalization layer before it enters multi-head attention block. Batch normalization is a regularization technique that is applied to input layers consisting of normalizing the activation values of the units in that layer so that there is a mean of 0 and a standard deviation of 1 [62]. This improves stability and facilitates the convergence speed while avoiding problems such as exploding or vanishing gradients. You should not confuse batch normalization with input data normalization, which involves scaling input data to a specific range to improve numerical representation.

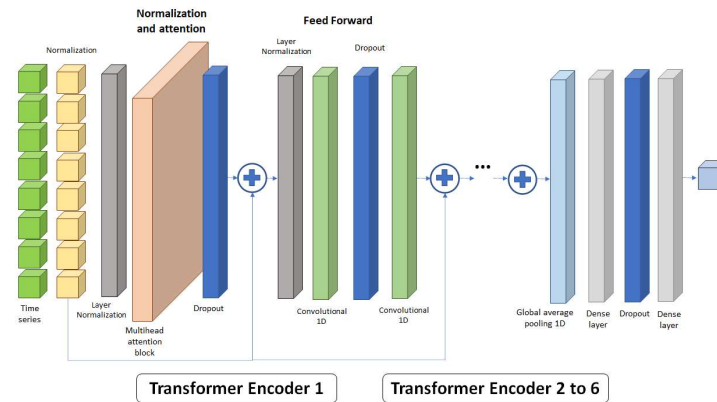


Figure 4. Architecture of the neural network used in the prediction model.

The attention block uses three independent networks: query (Q), key (K), and value (V) to acquire information from the input data and obtain a proper representation of it. The block calculates the attention weight by dividing the dot product between the query and key vectors by a scaling factor (d), transforming it using the softmax function to obtain numerical values between 0 and 1, and finally multiplying it by the value vector (Equation 1) to condense the information into a vector for each data point or token.

This set of layers and blocks forms the normalization and multi-head attention block. Multiple attention blocks can be used to detect further associations and relationships among data points and/or groups of data across different levels. These vectors are then combined using an additional neural network into a single compressed vector. Furthermore, a dropout layer is applied to its output [63] in order to prevent excessive overfitting of the model.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

The obtained vector is then added to the normalized input sequence and introduced into the Feed Forward layers, which consist of a normalization layer, a dropout layer, and two 1D convolutional layers. Finally, the output is added to the input of this layer, forming the encoder structure. Due to their feature extraction capability, convolutional networks are commonly used in data from numerical series.

Convolutional networks are a type of dense, feed-forward neural networks that have been successfully applied in many domains [64], such as classification problems or time series classification [65]. This structure of the model enables the recognition of specific patterns and the abstraction of information from the input sequence. This architecture is called an encoder, and it can be repeated up to six times in succession. A global average pooling 1D layer (GAP1D) is added to the output of the encoders for polling, preserving all relevant information in a single layer that is easily understood by dense layers equipped with another dropout layer to obtain the time series forecast.

Once the self-attention mechanism network architecture has been defined, it is necessary to empirically analyze it and determine the value of hyperparameters that maximize the result (lower

MAE or RMSE) with the lowest computational cost. Statistical analysis allows only considering relevant factors, which result in a simpler model to train.

Python and specifically its Keras and Tensorflow libraries have been used to design the neural network, which are intended for the creation of deep learning models and networks. They also allow great flexibility by allowing user-customizable configurations.

4. The Concepts and Procedures for the Utilization of Design of Experiments (DoE), ARIMAX and Bootstrap

Most practical situations where it is intended to control the output or response of a real process are conditioned by the existence of controllable variables (factors) and other unknown factors. The DoE is a methodology proposed by Ronald A. Fisher in the 1920s and 1930s and it encompasses a set of mathematical and statistical techniques that seek the effects on a measured response variable (Y) by manipulating simultaneously some input variables (Xs).

In the application of DoE techniques, the main guidelines may include the following: (1) stating the objectives of the research, (2) defining the response variable(s), (3) determining factors and levels, (4) deciding the experimental design type, (5) conducting the experiment, (6) data analysis and (7) validation of the results. With the intention of estimating first and second order effects and cross-product, two-level factorial designs are used, and the polynomial function must contain quadratic terms. An orthogonal design is shown in the equation below:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{1 \leq i < j}^k \beta_{ij} x_i x_j + \varepsilon \quad (2)$$

where k is the number of variables, β_0 is the constant term, β_i represents the coefficients of the linear parameters, x_i represents the variables, β_{ii} represents the coefficients of the quadratic parameters, β_{ij} represents the cross-product, and ε is the residual associated with the experiment (random error).

The variables are usually expressed in different units and vary in different ranges, which makes it necessary to code them. If this is not accomplished, it is impossible to assess the relevance of a factor through the magnitude of the model coefficients. As a result, all factor levels are set at two levels (-1, +1) with equal spaced intervals between them.

Hence, the factorial experiments with multiple factors with two levels ("high" and "low"), suggest that the complexity of the model grows exponentially (2^k), and the experiment requires a lot of resources and time. Since a low number of main effects and lower order interactions are significant to the response variable, and higher order interactions are not significant, a Fractional Factorial Design may be chosen.

As an extension of the widely used ARIMA (Autoregressive Integrated Moving Average) forecasting model, the ARIMAX allows for the inclusion of one or more exogenous variables, thereby enabling more satisfactory predictions. The model can be expressed as:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d y_t = c + \beta X_t + (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t \quad (3)$$

where y_t is the value of the time series at time t, X_t is the exogenous variable at time t, β is the coefficient that measures the impact of the exogenous variable, ϕ_1, \dots, ϕ_p are autoregressive coefficients, $\theta_1, \dots, \theta_q$ are moving average coefficients, c is de average level of the time series when all other variables are zero (AR(p), MA(q), β), L in the lag operator and finally, ε_t is the error term at time t. The proper use of this model requires the time series to be stationary, that is it must have constant mean and variance over time. To verify this condition, the Dickey-Fuller (ADF) test is employed, where the null hypothesis (H_0) states that the series is non-stationary, against the alternative hypothesis (H_1) that the series is stationary. If the series is found to be non-stationary, it must be differenced until stationarity is achieved. The determination of the AR and MA orders is guided by the ACF and PACF plots. Additionally, if the ACF exhibits a slow, gradual decay, this is a clear indicator of non-stationarity in the series.

Similarly, in cases where it is desired to estimate the parameters of an unknown data distribution, it is common to use the Bootstrap technique, introduced by Efron in 1979 [66]. This

method involves creating new datasets from an existing sample by employing resampling with replacement. The Bootstrap principle starts with a random sample from an unknown distribution $X = \{x_1, x_2, \dots, x_n\}$, and B resamples $X_1^*, X_2^*, \dots, X_B^*$, each of size n , are generated. For each resample X_b^* , the statistic of interest is computed $\hat{\theta}_b^*$, and the empirical distribution of $\hat{\theta}^*(X)$ is used to approximate the sampling distribution of the statistic. There is extensive literature on this topic, so I encourage interested readers to consult [67]. The desired estimator is then computed on the datasets generated, and this estimator will approximate the (unknown) estimator evaluated on the actual population.

5. Use of Artificial Neural Network

This methodology relies on historical data, where hyperparameter values are adjusted simultaneously to identify those most influential on the network's performance, aiming to minimize specific outputs. These outputs include the mean absolute error (MAE), defined by $MAE = \sum_{t=1}^T \frac{|\hat{y}_t - y_t|}{T}$, computing time per epoch, and the number of trainable parameters. Here, y_t represents the true output, \hat{y}_t denotes the estimated output T corresponds to the number of data points. To manage computational complexity and resource efficiency, a reduction strategy is often necessary to limit the number of experimental runs, as the complexity scales with the number of hyperparameters. Table 2 outlines the factors evaluated in this study to determine the optimal network size for minimizing prediction error on the provided dataset. It is important to note that these factors are presented as normalized and encoded values, ranging from -1 to 1.

Table 2. Hyperparameters considered in the model.

Hyperparameter	-1	1	Hyperparameter	-1	1
A Sequence size	5	35	E No. encoder blocks	2	6
B Head size	64	192	F mlp units	64	192
C Number of heads	2	6	G Learning rate	0.0005	0.0015
D F. Forward dimension	2	6			

The sequence size (A) denotes the number of data points extracted from the temporal sequence that are utilized for network training to generate the forecast. The head size (B) corresponds to the number of neurons comprising each attention head. The number of heads (C) represents the quantity of independent attention mechanisms embedded within the transformer block. The feed-forward dimension (D) specifies the count of one-dimensional convolutional layers. The number of encoder blocks (E) indicates the total number of encoder layers that constitute the network. The MLP units (multi-layer perceptron units - F) define the number of neurons within the fully connected dense network, while the learning rate (G) determines the magnitude of the adjustment step taken during each iteration to minimize the loss function.

To minimize the computational demands on the hardware, relevant hyperparameters are optimized to their ideal values, while non-essential factors are set to their minimum thresholds. The neural network is subsequently trained using historical data, concurrently with the collection of new data from the compressor. The model generates predictions in the form of a time series. Based on the model's training, the predicted data, and their operational expertise, control room technicians evaluate the potential implications of equipment operating outside acceptable ranges or experiencing unplanned shutdowns. Decisions are then made in accordance with established operational protocols.

Given the considerable number of factors under consideration (7), the adoption of a fractional factorial design is considered necessary. The resolution of the design is determined by the number of experimental runs required and the acceptable degree of aliasing among the factors, including their second-order, third-order, and higher-order interactions. Consequently, a resolution IV design is selected as the minimal configuration capable of capturing the most significant main effects and second-order interactions. While this design results in aliasing between factors and second-order interactions, it is deemed the most suitable option based on the established assumptions. By

employing this design, only 16 experimental trials are required, as opposed to the 128 trials needed for a full factorial design, and two replications are included.

Figure 5 illustrates the process followed. The data have been divided into three subsets: training (70%), validation (15%), and testing (15%). The hyperparameters are adjusted according to the determinations made in the DoE, and the MAE values for the validation data are calculated. Once the model with the optimum MAE is identified, the residuals are determined to establish the prediction limits. The testing data are used to assess whether the network possesses the desired predictive capability.

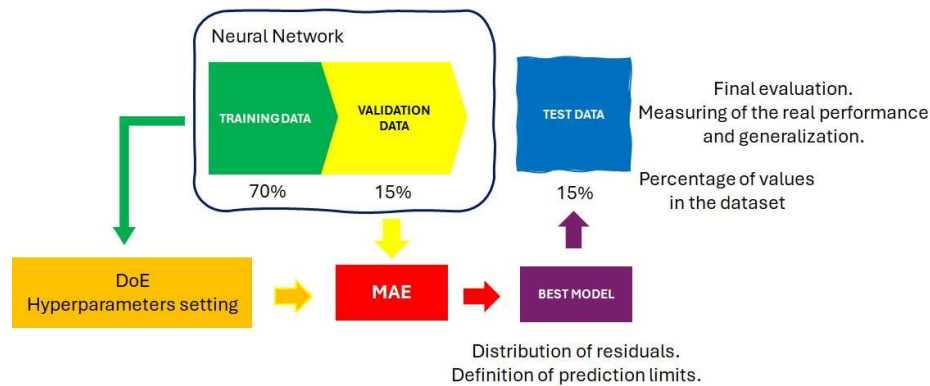


Figure 5. Identification of the best ANN model.

The columns of Table 3 display the MAE obtained for the validation and testing data for each configuration of the neural network and its replicates. Additionally, the average computation time per epoch in seconds is included, along with the number of trainable parameters, which provides an indication of the network's complexity.

Table 3. MAE. Results of the experiment for validation and test data.

No.	Run 1		Run 2		Run 3		s/Epoch (average)	Trainable param.
	Val.	Test	Val.	Test	Val.	Test		
1	0.1452	0.1930	0.1291	0.1746	0.1224	0.1688	3.243	14 513
2	0.0833	0.1297	0.1161	0.1429	0.0814	0.1247	2.455	17 609
3	0.1049	0.1371	0.0989	0.1328	0.1045	0.1311	2.735	20 497
4	0.1483	0.1992	0.1693	0.2115	0.1509	0.2029	17.391	57 545
5	0.1344	0.1919	0.1143	0.1798	0.1252	0.1852	3.017	32 665
6	0.1304	0.1561	0.0903	0.1308	0.1034	0.1330	2.799	17 681
7	0.1225	0.1418	0.1174	0.1396	0.1230	0.1505	1.310	6 065
8	0.1111	0.1473	0.1186	0.1519	0.1118	0.1528	1.355	3 225
9	0.1154	0.1564	0.1174	0.1585	0.1313	0.1646	1.460	17 585
10	0.0899	0.1619	0.0892	0.1653	0.0528	0.1448	3.500	32 713
11	0.1024	0.1672	0.1009	0.1651	0.0532	0.1467	17.294	75 793
12	0.1041	0.1384	0.0925	0.1245	0.1092	0.1370	2.835	20 425
13	0.1182	0.1760	0.1225	0.1821	0.1632	0.2055	2.333	32 689
14	0.1579	0.1904	0.1018	0.1667	0.1600	0.1962	2.060	14 489
15	0.1177	0.1388	0.1039	0.1255	0.1173	0.1532	1.775	20 377
16	0.1595	0.2081	0.1192	0.1745	0.1176	0.1724	3.435	14 609

From the response surface model for the MAE of the validation and testing data, it is inferred that the significant factors are the size of the data series (A), the number of encoder blocks (E), the size of the multilayer perceptron (F), and the learning rate (G). Additionally, the normality of the data allows for the application of ANOVA, leading to the conclusion that the mean errors of the two data groups (validation and testing) are significantly different, which may be attributed to various underlying causes. The presence of *data drift* or *data shift* refers to a change in the distribution of the

predictors, a change in the distribution of the target variable, or a change in the relationship between the predictors and the target variable. This phenomenon leads to a progressive decline in the model's predictive accuracy [68]. An additional factor contributing to this issue is the model's tendency to overfit the validation and training data, which compromises its ability to generalize effectively to new testing datasets. Resolving this challenge is non-trivial, as [69] highlights that the root causes often lie in the use of excessively complex models with a high number of trainable parameters or the model's inadvertent assimilation of noise inherent in the data series.

To further analyze the network, a design is implemented in which the discarded variables (B, C, and D) are fixed at their maximum, median, and minimum values, verifying whether the variables excluded by the DoE truly have no impact on the network's performance. Figure 6 presents the boxplot of the validation and test values, highlighting the non-significant factors and enabling clear observation of distinct processes.

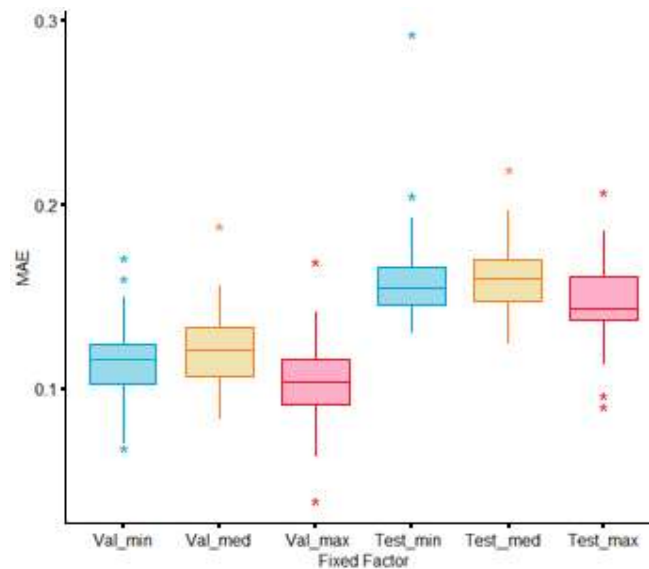


Figure 6. Box-plot of MAE for validation and test data.

After analyzing the normality of the obtained MAE values, ANOVA is performed on the normally distributed data, and the non-parametric Kruskal-Wallis test is applied to the non-normal distributions (test data). In both cases, it is concluded that the most complex architecture yields the lowest MAE, although the input data series must be the shortest. Likely, the use of a larger number of data points in the input series introduces additional noise into the network, resulting in poorer predictions. Therefore, the optimal architecture corresponds to Network (A = 20, B = 192, C = 6, D = 6, E = 2, F = 64, G = 15, $MAE_{Val} = 0.06289$ y $MAE_{Test} = 0.09573$).

Figure 7 illustrates the residuals of the validation and test data, which represent the difference between the observed actual values and those estimated by the neural network. It is readily apparent that the residuals exhibit bias, specifically 0.91 MPa, indicating that the predicted values are systematically lower than the actual values. Furthermore, the distributions of these residuals are non-normal.

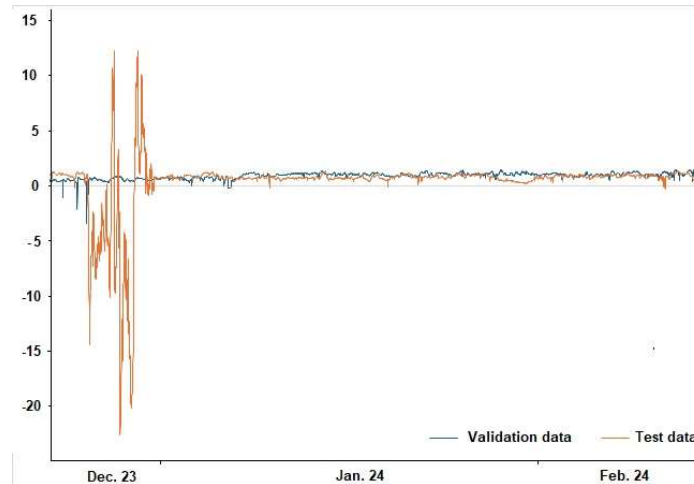


Figure 7. Residuals. Validation and test data.

Additionally, Table 4 presents the characteristics of the residuals for both series, as well as the residuals corresponding to the test data, excluding those associated with the year-end preventive maintenance shutdown.

Table 4. Characteristics of the residuals.

		Val. data	Test data	Test data *
Residuals ($y_t - \hat{y}_t$)	Bias	0.9165	0.3561	0.7906
	Standard Deviation	0.2826	2.6777	0.2039
	Normally distributed	No	No	No

* The first 1,000 data points of the series are excluded (maintenance time).

To evaluate the relationship between the actual and predicted data, the degree of correlation between the original data and their predictions can be measured. For this purpose, the Pearson and Spearman correlation coefficients, as well as Kendall's tau [70], can be employed. These coefficients measure the relationship between two variables in different ways. Refer to Table 5.

Table 5. Correlation coefficients between the actual and estimated data.

Data	Pearson	Spearman	Kendall tau
Validation	-0.3160	-0.4022	-0.2732
Test	0.9243	0.5135	0.3695

While the correlation between the actual and predicted data in the validation group exhibits an inverse relationship — that is, there is a tendency for one variable to decrease as the other increases — the test data, in contrast, shows a direct correlation. From this divergent behavior, it can be inferred that the compression process captured in the test and validation data differs.

The conclusion of this study involves determining the prediction interval, using the residuals corresponding to the validation data for this purpose. Figure 8 shows the distribution of the residuals with bias correction (null mean). Due to the non-normality of the distribution, the Bootstrap technique is employed to determine the 2.5th and 97.5th percentiles of the distribution, enabling the calculation of a 95% confidence interval. After performing 10 000 repetitions of 5 116 data points each, the $P_{2.5}$ and $P_{97.5}$ percentiles are estimated, with the median being the representative value for all of them ($P_{2.5} = -0.5269$ and $P_{97.5} = 0.4181$). Owing to the asymmetry of the residual distribution, the prediction limits are also asymmetric relative to the expected mean (27.3931, 28.3681 MPa).

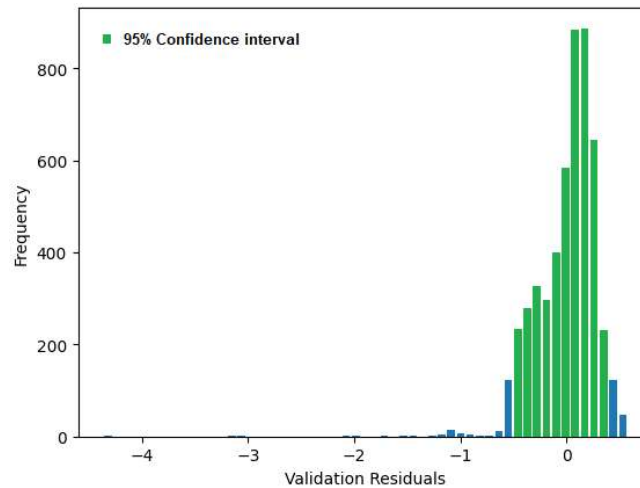


Figure 8. Distribution of validation residuals incorporating bias correction.

6. ARIMAX Model-Based Approach

As in the previous case, the same dataset is used to: (1) estimate the model parameters (training), (2) determine the forecast boundaries and bias (validation), and (3) finally evaluate the forecasting performance using test data. Table 6 displays the ADF test results for stationarity assessment across the time series, while Figure 9 plots the ACF and PACF used to determine the AR and MA orders.

Table 6. ADF test results.

Time series	ADF statistic	p-value	Stationary
Outlet pressure			
Training	-4.34435	0.000372	Yes
Validation	-7.04869	0.000000	Yes
Test	-3.53525	0.007127	Yes
Current consumption			
Training	-4.80803	0.000052	Yes
Validation	-6.36995	0.000000	Yes
Test	-3.82928	0.002622	Yes

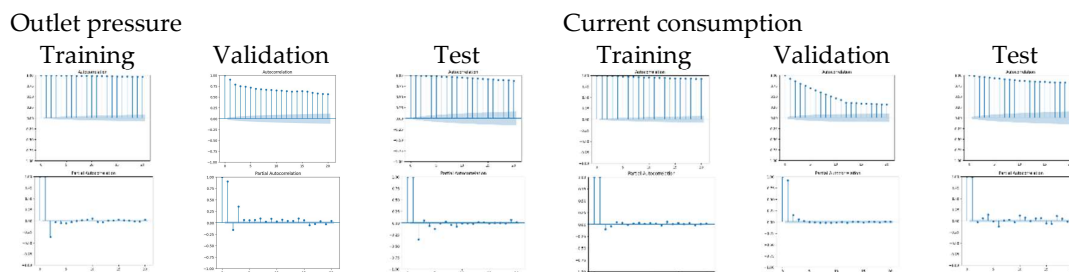


Figure 9. ACF and PACF for endogenous and exogenous variables.

From Figure 9, the autocorrelation and partial autocorrelation functions indicate the series needs first-order differencing ($d = 1$), with preliminary evidence supporting at least AR(1) and MA(1) terms in the model specification. The final model corresponds to an ARIMA(1, 1, 2), with the values of statistically significant parameters shown in Figure 10. The values were derived by incrementally testing ARIMA models up to fifth-order terms and retaining only those coefficients that achieved statistical significance ($P \leq 0.05$)

	coef	std err	z	P> z	[0.025	0.975]
IICIA	0.0016	5e-05	32.224	0.000	0.002	0.002
ar.L1	0.7311	0.006	120.556	0.000	0.719	0.743
ma.L1	-0.3850	0.006	-64.410	0.000	-0.397	-0.373
ma.L2	-0.1222	0.002	-49.488	0.000	-0.127	-0.117
sigma2	0.0862	4.16e-05	2071.043	0.000	0.086	0.086

Figure 10. Coefficients and signification for ARIMAX model.

The MAE for the training data is 3.4692, which is clearly higher than the values obtained by the neural network (see Table 3). This suggests greater imprecision in the predictions, indicating that the model is unable to capture the underlying patterns of the data and may be experiencing underfitting (an overly simplistic model). Based on the fitted model and using validation data, the residuals exhibit a bias of 0.6007. After bias removal, the percentiles show $P_{2.5} = -0.4466$ and $P_{97.5} = 0.1741$.

Figure 11 depicts the last 1,000 test data points, including the forecast limits, control limits, and the prediction provided by the neural network and ARIMAX with bias correction (+0.916 MPa and +0.600 MPa). The evident underfitting of the ARIMAX prediction model manifests in excessively stable forecasts that prove unresponsive to either perturbations or structural changes in the input time series, consequently resulting in increased prediction error.

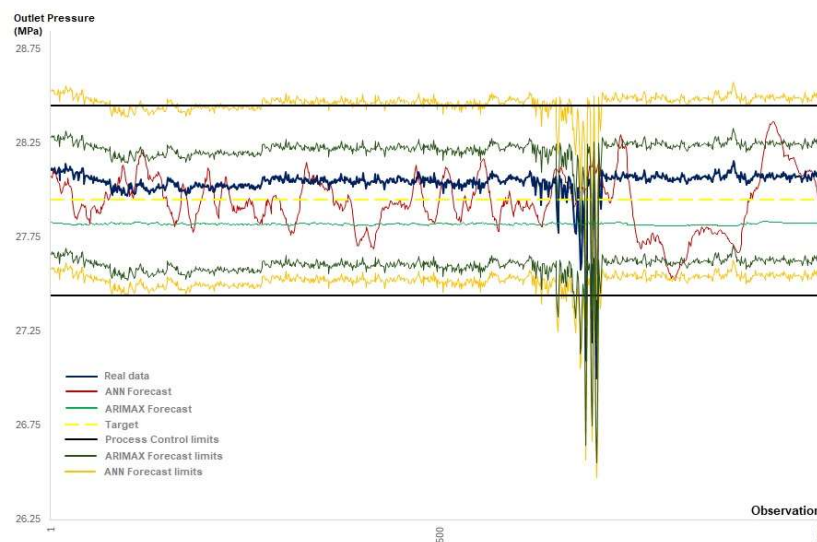


Figure 11. Last 1 000 data points.

Several inferences can be drawn from the neural network prediction model. First, the prediction lags delay the actual series. Second, disturbances are smoothed, exhibiting lower amplitude compared to the actual ones. Third, the neural network displays greater variability than the real data, indicating that the network has captured noise during its training and is reflecting it. This highlights the need for further study and a deeper understanding of the process and the current control system.

The limited dispersion of the original data is likely due to potential over-control by operators on the process (tampering), which must be identified and managed. The causes for this operator behavior, as outlined by [71], are: (1) limited capacity of workers to solve problems and make decisions, (2) differences in each operator's approach to identifying problems, (3) lack of time and motivation, and (4) shortcuts taken by those responsible for identifying and addressing issues, as there is no record of actions taken on the process.

Once the suitability of this technology for predicting future equipment performance has been established, it becomes necessary to incorporate new KPIs in the Maintenance Plan to monitor emerging technologies and enhance both process sustainability and environmental impact. New KPIs

must be implemented to assess maintenance performance, including: (1) information and communication technology metrics to evaluate these technological tools' effectiveness; (2) personnel competency indicators to monitor workforce adaptation to these solutions; and (3) health, safety, and environmental (HSE) measures to enhance organizational sustainability.

7. Conclusions

This paper demonstrates that by utilizing existing signals and employing neural networks, redundancy can be achieved, enabling the control of equipment in the event of a communication failure. Furthermore, it illustrates how the use of complex neural networks can predict the actual behavior of a hydrogen compressor and advance maintenance management within organizations. This approach fosters proactivity and predictability, avoiding unnecessary shutdowns, over-costs, over-control of the process, and the unnecessary use of resources.

Consistent with these findings, this model demonstrates superior generalization capabilities and provide more robust solutions than conventional forecasting systems when faced with pattern shifts in the data series.

The analysis of the neural network has helped to understand and interpret the differing behavior of the process by identifying similarities and/or differences between the validation and test data series. Thus, before implementing complex maintenance plans aimed at deploying actions and tools to enhance the predictive capability of models through the adoption of cyber-physical systems, IoT, security, data storage, etc., it is essential to understand the causes of the compressor's differing behavior, without overlooking the human factor as a potential source of variability in the process.

The use of predictive models, no matter how complex, enhances the understanding of the process and helps establish action plans focused on efficiently improving the maintenance function. It also increases the sustainability of organizational activities by avoiding the waste of resources on unnecessary actions. However, allowing the system to operate autonomously may lead to undesirable outcomes.

Finally, introducing and replicating such solutions without proper training and understanding among the personnel involved can lead to conflicting situations among control operators. This may occur when the forecasts exhibit greater variability, prompting attempts to over-adjust the process in response to the actual behavior of the compressor.

Author Contributions: S.P.: conceptualization, methodology, investigation, writing—original draft preparation, and visualization; C.G-G.: conceptualization, methodology, investigation, writing—review and supervision; M.A.S.: conceptualization, methodology, investigation, writing—review and supervision. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This paper has been produced within the scope of the doctoral activities carried out by the lead author at the International Doctorate School of the Spanish National Distance-Learning University (EIDUNED_ Escuela Internacional de Doctorado de la Universidad Nacional de Educación a Distancia). The authors are grateful for the support provided by this institution.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ACF. Autocorrelation function.

ADF. Augmented Dickey – Fuller.

ANN. Artificial neural network.

ARIMAX. Autoregressive Integrated Moving Average with EXogenous variables.

DoE. Design of experiment.

KPI. Key performance indicator.

MAE. Mean absolute error.

PACF. Partial autocorrelation function.

PdM. Predictive maintenance.

References

1. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. *Resolution Adopted by the General Assembly on 25 September 2015*, **2015**, 42809, 1-13.
2. United Nations. The Pact for the Future. *Resolution adopted by the General Assembly on 22 September 2024*, **2024**, New York.
3. Kagermann, H., Wahlster, W., & Helbig, J. Recommendations for implementing the strategic initiative Industrie 4.0. *Frankfurt: Acatech-National Academy of Science and Engineering*, **2013**.
4. Jasiulewicz-Kaczmarek, M., & Gola, A. Maintenance 4.0 technologies for sustainable manufacturing-an overview. *IFAC-PapersOnLine*, **2019**, 52(10), 91-96.
5. Tiddens, W., Braaksma, J., & Tinga, T. Exploring predictive maintenance applications in industry. *Journal of quality in maintenance engineering*, **2022**, 28(1), 68-85.
6. Wang, J., Zhang, L., Duan, L., & Gao, R. X. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. *Journal of Intelligent Manufacturing*, **2017**, 28, 1125-1137.
7. Uusitalo, T., Rana, P., Granados, M. H., & Macchi, M. Sustainable manufacturing: challenges, approaches and a roadmap. *Value Networks in Manufacturing: Sustainability and Performance Excellence*, **2017**, 27-41.
8. Jayal, A. D., Badurdeen, F., Dillon Jr, O. W., & Jawahir, I. S. Sustainable manufacturing: Modeling and optimization challenges at the product, process and system levels. *CIRP Journal of Manufacturing Science and Technology*, **2010**, 2(3), 144-152.
9. Garetti, M., & Taisch, M. Sustainable manufacturing: trends and research challenges. *Production planning & control*, **2012**, 23(2-3), 83-104.
10. Ghisellini, P., Cialani, C., & Ulgiati, S. A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. *Journal of Cleaner production*, **2016**, 114, 11-32.
11. Elia, V., Gnoni, M. G., & Tornese, F. Measuring circular economy strategies through index methods: A critical analysis. *Journal of cleaner production*, **2017**, 142, 2741-2751.
12. Jawahir, I. S., & Bradley, R. Technological elements of circular economy and the principles of 6R-based closed-loop material flow in sustainable manufacturing. *Procedia Cirp*, **2016**, 40, 103-108.
13. EN 16646. Maintenance. Maintenance within physical asset management. *European Committee for Standardization*. **2014**.
14. Serradilla, O., Zugasti, E., Rodriguez, J., & Zurutuza, U. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. *Applied Intelligence*, **2022**, 52(10), 10934-10964.
15. Davari, N., Veloso, B., Costa, G. D. A., Pereira, P. M., Ribeiro, R. P., & Gama, J. A survey on data-driven predictive maintenance for the railway industry. *Sensors*, **2021**, 21(17), 5739.
16. Jardine, A. K., Lin, D., & Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, **2006**, 20(7), 1483-1510.
17. De Jonge, B., & Scarf, P. A. A review on maintenance optimization. *European journal of operational research*, **2020**, 285(3), 805-824.
18. Damant, L., Forsyth, A., Farcas, R., Voigtländer, M., Singh, S., Fan, I. S., & Shehab, E. Exploring the transition from preventive maintenance to predictive maintenance within ERP systems by utilising digital twins. In *Transdisciplinary Engineering for Resilience: Responding to System Disruptions*, **2021**, 171-180. IOS Press.
19. Yan, J. Machinery prognostics and prognosis oriented maintenance management. *John Wiley & Sons*, **2015**. Singapore.
20. Kumar U, Galar D. Maintenance in the era of industry 4.0: issues and challenges. *Quality, IT and Business Operations: Modeling and Optimization*, **2018**, pp 231-250.
21. Ghaleb, M., & Taghipour, S. Evidence-based study of the impacts of maintenance practices on asset sustainability. *International Journal of Production Research*, **2023**, 61(24), 8719-8750.
22. Jasiulewicz-Kaczmarek, M., Legutko, S., & Kluk, P. Maintenance 4.0 technologies–new opportunities for sustainability driven maintenance. *Management and production engineering review*, **2020**, 11.

23. Bilge, P., Emec, S., Seliger, G., & Jawahir, I. S. Mapping and integrating value creation factors with life-cycle stages for sustainable manufacturing. *Procedia CIRP*, **2017**, 61, 28-33.
24. Holgado, M., Macchi, M., & Evans, S. Exploring the impacts and contributions of maintenance function for sustainable manufacturing. *International Journal of Production Research*, **2020**, 58(23), 7292-7310.
25. Jasiulewicz-Kaczmarek, M., Żywica, P., & Gola, A. Fuzzy set theory driven maintenance sustainability performance assessment model: A multiple criteria approach. *Journal of Intelligent Manufacturing*, **2021**, 32(5), 1497-1515.
26. Wakiru, J., Pintelon, L., Muchiri, P. N., & Chemweno, P. Maintenance optimization: application of remanufacturing and repair strategies. *Procedia CIRP*, **2018**, 69, 899-904.
27. Alarcón, M., Martínez-García, F. M., & de León Hijes, F. C. G. Energy and maintenance management systems in the context of industry 4.0. Implementation in a real case. *Renewable and Sustainable Energy Reviews*, **2021**, 142, 110841.
28. Bányai, Á., & Bányai, T. Real-time maintenance policy optimization in manufacturing systems: An energy efficiency and emission-based approach. *Sustainability*, **2022**, 14(17), 10725.
29. Pérez, M. P., Rodríguez, Á. T. P., & de la Paz Martínez, E. M. Contribution to the sustainability of the plastics industry in Cuba through lean-green maintenance management. *IFAC-PapersOnLine*, **2022**, 55(10), 1918-1923.
30. Karuppiah, K., Sankaranarayanan, B., & Ali, S. M. On sustainable predictive maintenance: Exploration of key barriers using an integrated approach. *Sustainable Production and Consumption*, **2021**, 27, 1537-1553.
31. Silvestri, L., Forcina, A., Introna, V., Santolamazza, A., & Cesarotti, V. Maintenance transformation through Industry 4.0 technologies: A systematic literature review. *Computers in industry*, **2020**, 123, 103335.
32. Antosz, K., Jasiulewicz-Kaczmarek, M., Paško, L., Zhang, C., & Wang, S. Application of machine learning and rough set theory in lean maintenance decision support system development. *Eksploracja i Niezawodność*, **2021**, 23(4), 695-708.
33. Karki, B. R., & Porras, J. Digitalization for sustainable maintenance services: A systematic literature review. *Digital Business*, **2021**, 1(2), 100011.
34. Shaheen, B. W., & Németh, I. Integration of maintenance management system functions with industry 4.0 technologies and features—A review. *Processes*, **2022**, 10(11), 2173.
35. EN 15341. Maintenance. Maintenance Key Performance Indicators. *European Committee for Standardization*. **2019**.
36. Rodrigues, J. A., Farinha, J. T., Mendes, M., Mateus, R., & Cardoso, A. M. Short and long forecast to implement predictive maintenance in a pulp industry. *Eksploracja i Niezawodność*, **2022**, 24, 1.
37. Lee, E. A. Cyber physical systems: Design challenges. In 2008 11th IEEE international symposium on object and component-oriented real-time distributed computing *IEEE*, **2008**, pp. 363-369.
38. Jazdi N. Cyber physical systems in the context of Industry 4.0. In 2014 IEEE international conference on automation, quality and testing, robotics *IEEE*. **2014**, pp. 1-4.
39. Alguliyev R, Imamverdiyev Y, Sukhostat L. Cyber-physical systems and their security issues. *Computers in Industry 100*. **2018**, pp 212-223.
40. Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento Jr, C., Prendinger, H., & Henriques, E. M. Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers & Industrial Engineering*, **2018**, 115, 41-53.
41. Francis, F., & Mohan, M. Arima model based real time trend analysis for predictive maintenance. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology -ICECA. *IEEE*, **2019**, pp. 735-739.
42. Beard, E., Marsden, J., Brown, J., Tombor, I., Stapleton, J., Michie, S., & West, R. Understanding and using time series analyses in addiction research. *Addiction*, **2019**, 114(10), 1866-1884.
43. Rodrigues, J. A., Farinha, J. T., Mendes, M., Mateus, R., & Cardoso, A. M. Short and long forecast to implement predictive maintenance in a pulp industry. *Eksploracja i Niezawodność*, **2022**, 24, 1.
44. Jan B, Farman H, Khan M, Imran M, Islam IU, Ahmad A, et al. Deep learning in big data Analytics: a comparative study. *Comput Electr Eng*, **2019**, 75, 275-87
45. Wason R. Deep learning: evolution and expansion. *Cognit Syst Res*, **2018**, 52, 701-8.

46. Kanawaday A, Sane A. Machine learning for predictive maintenance of industrial machines using IoT sensor data. In 2017 8th IEEE international conference on software engineering and service science *IEEE*, **2017**, pp 87-90.
47. Zeng, A., Chen, M., Zhang, L., & Xu, Q. Are transformers effective for time series forecasting? In Proceedings of the AAAI conference on artificial intelligence, **2023**, 37, 9, pp. 11121-11128.
48. Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, **2022**, 23(1), 40-55.
49. Murphy, K. P. Machine learning: a probabilistic perspective. *MIT press*, **2012**.
50. Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. Machine learning on big data: Opportunities and challenges. *Neurocomputing*, **2017**, 237, 350-361.
51. Sultan, H. H., Salem, N. M., & Al-Atabany, W. Multi-classification of brain tumor images using deep neural network. *IEEE access*, **2019**, 7, 69215-69225.
52. Goldberg, Y. Neural network methods for natural language processing. *Springer Nature*, **2022**.
53. Ahmed, S., Nielsen, I. E., Tripathi, A., Siddiqui, S., Ramachandran, R. P., & Rasool, G. Transformers in time-series analysis: A tutorial. *Circuits, Systems, and Signal Processing*, **2023**, 1-34.
54. Wang, Y., & Zhao, Y. Multi-scale remaining useful life prediction using long short-term memory. *Sustainability*, **2022**, 14(23), 15667.
55. Abidi, M. H., Mohammed, M. K., & Alkhalefeh, H. Predictive Maintenance Planning for Industry 4.0 Using Machine Learning for Sustainable Manufacturing. *Sustainability*, **2022**, 14, 3387.
56. Cotrino A., Sebastián M. A., & González-Gaya C. Industry 4.0 HUB: a collaborative knowledge transfer platform for small and medium-sized enterprises. *Applied Sciences*, **2021**, 11(12), 5548.
57. Cotrino A., Sebastián M. A., & González-Gaya C. Industry 4.0 roadmap: Implementation for small and medium-sized enterprises. *Applied sciences*, **2020**, 10(23), 8566.
58. Haarman, M., Mulders, M., Vassiliadis, C. Predictive maintenance 4.0: Predict the unpredictable. In PwC documents, no. *PwC & mainnovation*, **2017**, p. 31.
59. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint*, **2014**. arXiv:1409.0473.
60. Luong, M. T., Pham, H., & Manning, C. D. Effective approaches to attention-based neural machine translation. *arXiv preprint* **2015**, arXiv:1508.04025.
61. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Advances in neural information processing systems*, **2017**, 30.
62. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, 86(11), pp 2278-2324.
63. Wang, Z., Yan, W., & Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In 2017 International joint conference on neural networks *IEEE*, **2017**, pp. 1578-1585.
64. Ioffe, S., & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, **2015**, pp. 448-456.
65. Zaremba, W., Sutskever, I., & Vinyals, O. Recurrent neural network regularization. *arXiv preprint*, **2014**, arXiv:1409.2329.
66. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **1979**, 7, pp. 1-26.
67. Michelucci, U., & Venturini, F. Estimating neural network's performance with bootstrap: A tutorial. *Machine Learning and Knowledge Extraction*, **2021**, 3(2), 357-373.
68. Moreno-Torres, J. G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N. V., & Herrera, F. A unifying view on dataset shift in classification. *Pattern recognition*, **2012**, 45(1), 521-530.
69. Ying, X. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, **2019**, February (Vol. 1168, p. 022022). IOP Publishing.
70. Tohka, J., & Van Gils, M. Evaluation of machine learning algorithms for health and wellness applications: A tutorial. *Computers in Biology and Medicine*, **2021**, 132, 104324.
71. Smeds, M. Exploring Tampering: Towards an Understanding of Why Improvement Efforts Sometimes Fail (*Doctoral dissertation, Linköping University Electronic Press*). **2022**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.