UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dottorato di ricerca in Ingegneria industriale e del territorio

Ciclo XXXVIII

Development of an integrated methodology combining physical modelling and machine learning for the implementation of a Digital Twin in industrial applications.

Candidato: Luigi Leopardi

Relatore (Tutor): Prof. Matteo Giacopini

Coordinatore del Corso di Dottorato: Prof. A. Muscio.

Index

1 Introduction 5
2. Digital Twin 9
  2.1 Definition 9
  2.2 The Two Twins 10
  2.3 Communication with the Physical Twin 12
  2.4 Management of Optimizers 13
  2.5 Digital Twin layout 16
  2.6 Technological Requirements 18
    2.6.1 Response Time 18
    2.6.2 Quality of Acquired Data 19
  2.7 System interface and management components 20
3 Filter Press 23
  3.1 Press Structure 23
  3.2 Filtration Mechanism 25
    3.2.1 Closing Phase 26
    3.2.2 Filling and Pressurization 26
    3.2.3 Squeezing 29
    3.2.4 Cake Washing and Air Blowing 30
    3.2.5 Discharge 32
  3.3 Washing Mechanisms 34
4 Digital Twin Setup for the Filter Press 37
  4.1 Objective Definition 37
    4.1.1 Priority Definition and Correlation 38
  4.2 Modelling Strategies 40
5 Data Acquisition and Conditioning 43
  5.1 Label Translation 43
  5.2 Index and Spike Correction 46
  5.3 Cycle Identification and Alignment 47
  5.4 Data Package Generation 49
6 Deterministic Physical-Based Methods 51
  6.1 Aggregated Filtration Parameters 51
  6.2 Pressure Drop Evaluation Δp 52

6.3 Calculation of Parameter c (Concentration) 53
6.4 Evaluation of Non-Directly Measurable Parameters 55
6.4.1 Constant Flow Rate Condition – Analysis of R................................ 55
6.4.2 Constant Pressure Condition – Analysis of α 61
7. Cycle-Based Aggregation 65
7.1 Statistical Numerical Aggregation... 67
8. Neural Networks.. 69
8.1 Autoencoders 70
8.2 Feedforward Neural Networks 72
8.3 Convolutional Neural Networks.. 74
8.4 Autoencoders for Aggregation 76
8.4.1 Autoencoder Results on Time Series 79
9 Performance Index Definition 83
10 KPI Predictor–Optimizer 85
10.1 Recurrent Neural Networks (RNN) 85
10.2 Forecaster 87
10.2.1 Hybrid Forecasting Model 88
10.2.2 Predictor Results.. 89
10.2.3 KPI Analysis on Forecasts 91
11 Direct Solution and Anomaly Identification 93
12 Similarity Analysis.. 95
12.1 Clustering Methodologies 96
12.1 .1 K-means Algorithm 96
12.1.2 HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise 96
12.1.3 Self-Organizing Maps (SOM)... 97
12.1.4 Comparison of Methods 97
12.2 Clustering and Washing Impact Index 97
12.3 Applicability of the Incidence Index...101
13 Validation and Result Analysis 105
13.1 Test of direct formulation on historical data 106
14 Conclusions...111
References 115

1 Introduction

In recent years, the industrial sector has been undergoing a profound and widespread process of digital transformation, driven by the need to increase production efficiency, reduce operational costs, improve plant reliability, and mitigate environmental impact. This transition is strongly supported by the rapid development of enabling technologies such as industrial IoT platforms, distributed sensor networks, advanced automation systems, and increasingly accessible data-driven methodologies. In parallel, industrial stakeholders are progressively adopting strategies aimed at improving process transparency, reducing unplanned downtime, and implementing predictive approaches to maintenance and operational management. In this scenario, the availability of high-frequency operational data and the possibility of converting such information into actionable knowledge are becoming key factors for sustaining competitiveness and ensuring long-term asset performance.

Within this broader trend, the mining sector represents one of the industrial contexts where digital transformation is potentially most impactful, but also most complex. This is primarily due to the scale of the facilities, the high energy demand of the processes, and the intrinsic dependence of mining operations on the interaction with the surrounding environment. Unlike other industrial domains, mining activities inevitably involve a direct modification of the territory and the generation of large quantities of waste material. For this reason, the industry is increasingly required to monitor, evaluate, and control the impact that extractive processes have on local ecosystems, water resources, land morphology, and safety conditions. This growing attention is reflected not only in the evolution of regulatory frameworks and environmental standards, but also in the strategic priorities of industrial operators, which are progressively shifting toward more sustainable and responsible forms of production.

In this context, it is crucial to highlight that the demand to limit environmental impact must not negatively affect, or ideally must improve, plant productivity. On the contrary, sustainability and productivity can no longer be considered separate objectives, but rather interconnected requirements that must be simultaneously achieved through technological and organizational innovation. This is particularly relevant for mining plants, where production volumes are typically high and operational continuity is a primary constraint. Therefore, adopting new strategies for waste management, water recovery, and energy optimization requires robust solutions capable of maintaining stable throughput while reducing costs and risks.

The requirement to minimize environmental impact includes several key aspects. First, it involves the restoration and remediation of sites at the end of operations, ensuring that the territory can be safely reintegrated with minimal long-term consequences. Second, it requires the mitigation of any risk of material dispersion, which may occur due to slope instability, flooding events, and failures in containment systems. Third, environmental sustainability is also associated with non-invasiveness, meaning the reduction of permanent modifications to the territory and the minimization of new infrastructure requirements. In recent years, these aspects have become particularly critical due to well-known safety incidents associated with tailings management, as well as the increasing scarcity of water resources in several mining regions worldwide.

From a process standpoint, extractive operations frequently involve handling the extracted material in the form of slurry or, more generally, as a fluid. This is strongly associated with mineral processing technologies such as flotation, which are widely adopted for the recovery of valuable minerals from crushed ore. Flotation processes aim to separate fractions containing valuable material from a slurry obtained by mixing finely crushed ore with water, generating a suspension characterized by complex physical properties, including variable rheology and solid concentration. These processes are chemical–mechanical in nature and are typically carried out in dedicated flotation tanks, where the desired separation is obtained through controlled aeration and the use of chemical reagents.

As a result of flotation, two main biphasic fluids (solid/liquid) are produced. A first phase, rich in valuable material, is collected at the surface due to the attachment of mineral particles to air bubbles and subsequent flotation. A second phase, collected at the bottom of the flotation tanks, consists of waste solids and process water, and is generally referred to as tailings slurry. Importantly, both phases require mechanical filtration and dewatering steps in order to be further processed, transported, or stored. The concentrate stream, containing valuable material, often undergoes additional chemical and/or physical treatments to solidify or stabilize the product. Once these intermediate steps are completed, mechanical filtration is required to achieve the desired moisture content and enable subsequent handling.

In recent years, however, the waste stream collected at the bottom of flotation tanks has become an increasingly critical material requiring filtration, primarily due to environmental and safety considerations. Historically, tailings were commonly stored in large settling ponds or tailings dams, where solid–liquid separation occurred by gravitational sedimentation over long periods of time. This approach required the construction of large storage basins and imposed the need for continuous monitoring, maintenance, and safety management. In addition, tailings dams represent structures associated with non-negligible risk, especially under extreme weather events, operational errors, and long-term degradation mechanisms.

Moreover, the presence of large volumes of settling slurry leads to irreversible changes in the morphology of the territory, from both a climatic and a geological perspective. The permanent occupation of extensive areas, the potential contamination of soil and groundwater, and the consumption of water in the tailings storage system have motivated a progressive shift away from this practice. As a consequence, the use of settling ponds and large tailings dams is increasingly discouraged and, in many cases, progressively abandoned in favour of mechanical filtration technologies. This transition is often referred to as “filtered tailings” or “dry stacking”, where a mechanically dewatered solid fraction is produced and stored with improved safety and reduced environmental footprint.

Within this evolving operational and regulatory framework, filter presses have become a focal point of attention in mining plants. In the processing of valuable material, relatively smaller machines are typically employed, since the concentrate flow rate is often lower and high product quality is the primary requirement. Conversely, for waste material management, filter press productivity becomes a key bottleneck of the entire plant. This is because tailings generation rates are extremely high, and the plant must continuously treat and dewater very large quantities of slurry in order to avoid interruptions in upstream production. As a result, modern mining plants increasingly depend on high-throughput filtration systems, whose reliability and efficiency directly affect plant availability and global production capacity.

The separation of slurry into a solid fraction and water enables two strategic outcomes. First, it makes possible the recovery of process water, which can be treated and reused, reducing freshwater intake and improving overall water management. Second, it produces a solid fraction that is lighter, easier to transport, and simpler to store, enabling more stable and less invasive tailings disposal strategies. Therefore, the filter press system used to process waste slurry becomes a crucial element for environmentally sustainable and economically viable management of mining residues.

At the same time, filter presses are typically volumetric machines, operating cyclically and handling discrete batches. For this reason, their performance depends on the dynamic interaction between operational parameters, slurry characteristics, machine configuration, and wear conditions. These systems are affected by variability in feed properties, fluctuations in solids concentration, changes in mineral composition, cloth clogging phenomena, and mechanical degradation of critical components. Consequently, accurate performance monitoring and a deep understanding of the physical phenomena governing filtration behaviour are essential to maintaining competitive performance under varying operating conditions. Additionally, since filter presses often represent the bottleneck of the plant, even small performance losses can result in significant productivity reduction and increased operational costs.

In recent years, the demand for high throughput combined with low processing costs has driven a significant increase in machine size. Industrial filter presses are progressively characterized by larger filtration areas, increased numbers of plates, higher operating pressures, and more complex auxiliary systems. While this evolution improves nominal productivity, it also increases the complexity of the process and the difficulty of understanding the system state through direct observation. In such a context, data-driven methodologies and advanced monitoring frameworks become essential to ensure stable operation, reduce unplanned downtime, and optimize control strategies.

It is within this framework that the Digital Twin project addressed in this thesis is positioned. The original objective of the work is to develop a digital model of a representative machine taken as a case study [1], [2], with the purpose of capturing the essential characteristics of its behaviour and enabling quantitative analysis of its operating conditions. The Digital Twin paradigm, when applied to industrial assets, provides a structured approach for integrating physical understanding with operational data, supporting diagnostics, forecasting, and decision-making processes. By establishing a link between real-time plant information and a continuously updated digital representation, it becomes possible to identify inefficiencies, monitor degradation patterns, and support parameter tuning strategies aimed at maximizing performance while minimizing cost and environmental impact.

The research presented in this thesis is structured as follows:

1. **Analysis of the machine operating principles:** The first step consists of an analysis of the operating principles of the family of machines under investigation. The purpose of this phase is to identify the physical mechanisms involved in the filtration cycle, clarify the role of each operating phase, and highlight the main criticalities that affect productivity and reliability. This includes the identification of the variables that characterize cycle evolution and the definition of which process signals can be effectively used to infer the state of the filtration system.

2. **Definition of performance parameters of industrial relevance:** The second phase focuses on defining the performance parameters of greatest relevance for the target application sector, namely mining applications for tailings and waste filtration. In this phase, it is essential to identify which aspects are critical for the end user to optimize and which features require improvement, considering both productivity-related indicators and sustainability-driven metrics. Special attention is given to parameters that affect throughput, cycle duration, downtime, energy efficiency, and operational stability.

3. **Definition of the design boundaries of the Digital Twin project:** In the third phase, the Digital Twin system itself is designed, defining which elements should be included and

which should be excluded [3]. This includes setting the modelling assumptions, selecting the physical and numerical layers of the Digital Twin, establishing the required data structure, and identifying the expected outputs. The objective is to ensure a balance between model complexity, interpretability, scalability, and industrial feasibility.

4. **Model development:** The fourth phase concerns the implementation of the identified modelling strategies. This includes both deterministic models inspired by filtration physics and data-driven approaches aimed at extracting latent information from operational signals. The integration of different methodologies enables the Digital Twin to capture both interpretable parameters and nonlinear behaviours that emerge in real industrial scenarios.

5. **Results analysis:** Finally, the last phase focuses on the validation and analysis of the obtained results. Model outputs are evaluated with respect to their ability to describe cycle dynamics, estimate meaningful state indicators, and support performance interpretation. The discussion emphasizes both the strengths and limitations of the proposed approach, outlining the potential industrial impact and the possible extensions for future developments.

Overall, this thesis aims to contribute to the development of a robust framework for the application of Digital Twin methodologies to filtration systems in the mining sector, providing both methodological tools and practical insights for improving efficiency, reliability, and sustainability in large-scale tailings management.

2. Digital Twin

2.1 Definition

In recent years, the concept of the Digital Twin has emerged as one of the most transformative paradigms in the fields of industry, engineering, and digital transformation [4] . Driven by the increasing availability of sensor data, advances in computational power, and the widespread adoption of artificial intelligence techniques, digital twins are rapidly becoming a cornerstone of modern industrial systems. A Digital Twin can be described as a virtual replica of a real object, system, or process, designed to mirror its physical counterpart with a high degree of fidelity. Unlike traditional simulation models, which are often static and limited to predefined scenarios, a digital twin is a living and evolving entity, capable of continuously updating its internal state through interaction with real-world data.

More specifically, a digital twin is not merely a static digital model, but a dynamic representation that can learn, adapt, and anticipate system behaviour over time. This capability is enabled by the integration of heterogeneous data sources, including sensors, control systems, historical databases, and advanced algorithms for data processing and analysis. Through the use of artificial intelligence and machine learning techniques, the digital twin can capture complex nonlinear behaviours, identify hidden patterns, and support predictive and prescriptive decision-making. For all practical purposes, a Digital Twin model can therefore be defined as a software product that enables the comprehensive representation, monitoring, and analysis of a mechanical system throughout its entire lifecycle [5].

A Digital Twin is generally understood as the combination of three fundamental and interdependent elements:

1. **The physical entity**: a single machine, a production line, an industrial plant, a building, a vehicle, or a large-scale infrastructure. This physical asset operates in the real world and continuously generates data through its interaction with the surrounding environment and operating conditions.
2. **The digital model**: a virtual representation of the physical entity that describes its structure, operating state, and dynamic behaviour. This model may rely on physics-based equations, data-driven methodologies, or hybrid approaches that combine both paradigms in order to accurately capture the system's behaviour.
3. **The bidirectional data flow**: enabled by sensors and the associated IT and communication infrastructure. This data exchange allows information to be transmitted from the physical system to the digital twin in near real time, while also enabling the digital twin to provide feedback, optimizations, predictions, and control strategies back to the physical asset.

According to a strict interpretation of the standard definition of a Digital Twin, any model that enables the holistic and integrated representation of a system in its entirety can be classified as such. This includes the coordinated use of all available technologies required to achieve the most accurate and comprehensive global modelling possible, encompassing mechanical, electrical, and control aspects. In this strict sense, a digital twin may be viewed as a black-box system which, once developed and calibrated, allows for the detailed simulation of a machine's behaviour under varying operating conditions and functional parameters. Such a model can be used to explore hypothetical scenarios, evaluate design alternatives, and assess system performance without interfering with the physical asset.

While achieving this level of modelling already represents a significant added value in terms of system understanding and optimization, the true potential of a fully developed Digital Twin lies in its ability to integrate hard or soft real-time data from the physical world. Continuous data integration enables the digital twin to remain synchronized with the evolving state of the physical system, thereby increasing the accuracy and reliability of its predictions. Furthermore, this real-time interaction allows the incorporation of evolutionary technologies, such as adaptive control strategies and machine learning models, within the virtual modelling framework. As a result, the digital twin can progressively improve its performance, support predictive maintenance, detect anomalies, and provide actionable insights for decision-makers, ultimately contributing to increased efficiency, reliability, and sustainability of industrial systems [6] .

2.2 The Two Twins

For a Digital Twin to exist, the presence of its physical counterpart is a fundamental requirement. As previously introduced, the physical twin constitutes the system, machine, or infrastructure that is intended to be represented and modelled within the digital domain. This physical component can be regarded as the portion of reality delimited by well-defined boundaries of interest, namely the set of elements that intercept the system's inputs and outputs and ultimately determine its operational behaviour. The explicit definition of these boundaries is a critical step in the Digital Twin development process, as it establishes the scope of the modelling activity and ensures consistency between the physical and virtual domains.

This delimitation enables the creation of a coherent and meaningful digital representation. In practice, it is neither necessary nor always desirable to model an entire machine or plant in its complete global entirety. Instead, it is essential to clearly identify and define the specific portion of the physical world that is to be replicated, including the relevant physical quantities, interfaces, and interactions. From this perspective, even a single subsystem of a complex industrial plant can be considered a valid object of a Digital Twin, provided that the operational parameters, functional relationships, and governing physical laws that describe its behaviour can be accurately characterized and observed [7]. This approach allows the Digital Twin to remain both computationally manageable and sufficiently accurate for its intended purpose.

Adopting such a perspective naturally leads to a modular interpretation of the Digital Twin concept. This modularity allows the modelling activity to be addressed in a more flexible, scalable, and progressive manner.

A concrete example is that of a hydroelectric power plant. Depending on the design objectives or operational requirements, a Digital Twin can be developed to represent the entire facility, including the reservoir, hydraulic conveyance system, turbines, generators, control systems, and auxiliary equipment. Alternatively, the modelling scope may be restricted to the turbine alone, treating upstream hydraulic quantities, such as flow rate and pressure, as inputs, and electrical quantities related to power generation as outputs. In a further refinement, a Digital Twin may be created exclusively for the hydraulic conveyance system, modelling the hydraulic characteristics of the pipeline as a function of its geometry, material properties, and operating pressure conditions.

In each of these cases, the individual blocks can be developed independently, as long as the physical, informational, and functional boundaries are clearly defined and the quantities required to correlate the model with the real system can be measured or estimated with sufficient accuracy [8]. This independence does not imply isolation: rather, it ensures that each module can be validated, calibrated, and refined separately before being integrated into a larger and more comprehensive digital framework.

This modular approach is often particularly effective in industrial contexts, as it enables shorter development times, greater specialization on individual components, and a more efficient allocation of modelling and computational resources. Moreover, it does not preclude the future integration of individual modules into a broader and more complex Digital Twin system. On the contrary, modularity facilitates gradual expansion and continuous improvement, allowing the Digital Twin to evolve alongside the physical asset throughout its lifecycle. Additionally, a modular architecture enables the clear identification of the measurement points required for model correlation and validation, thereby simplifying the implementation of bidirectional data exchange between the physical and digital domains Figure 1.

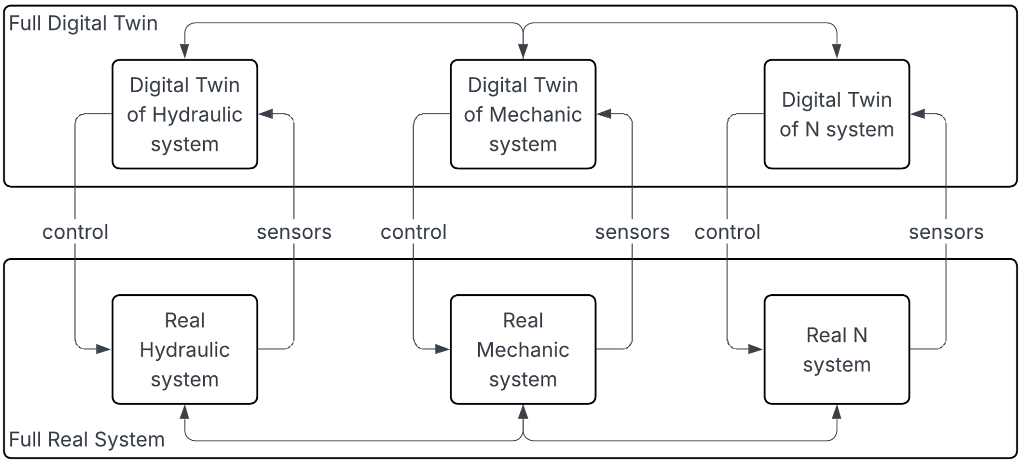

Figure 1 modular digital twin scheme

Even when used in isolation from the real world, a Digital Twin already represents a highly powerful analytical tool. Access to a virtual model capable of coherently reproducing a physical phenomenon enables in-depth analyses and comparisons that are difficult or impossible to obtain using the real system alone [5]. For instance, it becomes possible to evaluate system behavior under entirely new operating scenarios, including extreme, rare, or experimentally inaccessible conditions. This capability is particularly valuable in early design phases, where physical prototypes may not yet exist, or in operational contexts where experimentation on the real system would be impractical, unsafe, or excessively costly.

Similarly, a Digital Twin allows for the investigation of the internal behaviour of components or subsystems that are difficult to observe directly, either due to technological limitations, harsh operating environments, or safety constraints. A complete and well-correlated Digital Twin is capable of virtualizing the behaviour of the real system and performing predictive analyses under conditions that have never been experimentally tested. This predictive capability enables the anticipation of critical issues, the identification of performance bottlenecks, and the optimization of operational strategies, ultimately leading to an overall increase in system efficiency and reliability.

These advantages translate into tangible benefits in terms of both productivity and safety. The ability to simulate future or hypothetical scenarios reduces the risk of design errors, supports more effective maintenance planning, and enables more informed and data-driven decision-making processes. Furthermore, the virtual testing of potentially hazardous situations prevents the exposure of personnel and equipment to risky operating conditions, contributing to safer and more sustainable system management. In summary, the Digital Twin should not be regarded merely as a passive copy of reality, but rather as an advanced analytical and decision-support tool that significantly extends the capabilities for understanding, optimizing, and controlling complex physical systems.

2.3 Communication with the Physical Twin

Having access to an accurate digital model of a real system can provide significant benefits throughout the entire development and lifecycle of an industrial plant. However, the true added value of a system properly defined as a Digital Twin lies in its ability to establish a real-time bidirectional communication between the physical world and the virtual world [9]. It is this continuous, dynamic, and up-to-date interconnection that distinguishes a Digital Twin from a simple numerical model or an offline simulation, transforming it into an evolutionary and active tool within the production process.

The bidirectional exchange of information between the physical twin and the digital twin enables advanced functionalities such as:

- the adoption of evolutionary methodologies, including neural models and machine learning techniques;
- the progressive evolution and self-optimization of the model itself;
- direct tuning of machine parameters, either automatically or in a supervised manner;
- the continuous availability of an accurate estimation of the plant's operational state.

Thanks to this continuous data exchange, the Digital Twin becomes an ideal platform for the application of modelling techniques based on training/testing paradigms, such as statistical methods and machine learning technologies. The development of a self-learning model inherently requires large volumes of well-structured, indexed, and high-quality data. The data acquisition infrastructure described for a Digital Twin provides precisely this type of context, making such methodologies particularly effective and reliable over time.

More "traditional" technologies, such as the application of analytical models, the solution of numerical problems, or the mathematical fitting of signals, also benefit greatly from a continuous flow of information. Industrial reality is characterized by constant evolution: machinery undergoes wear, operating conditions change, and processed materials exhibit intrinsic variability. The continuous comparison between real behaviour and expected behaviour enables the early detection of significant deviations from reference conditions (set-points), allowing corrective actions to be undertaken both on the model side and on the plant side. Indeed, deviation analysis represents one of the primary diagnostic outputs of a Digital Twin system.

A misalignment between the digital model and the real system may indicate several conditions of interest, including the following [10].

Variation in operating conditions

It is possible that the plant's working conditions, such as processed material, ambient temperature, or operational load, change without such modifications having yet been incorporated into the virtual model. In this case, the Digital Twin provides immediate feedback regarding deviations from expected performance and allows one to:

- verify operational safety under the new conditions;
- re-optimize the system setup to maintain efficiency and quality;
- prevent configuration errors by suggesting technical or parametric adjustments.

Degradation of machine performance

A persistent correlation error may be symptomatic of a deterioration of the real system, caused by:

- mechanical failures or structural damage;
- progressive wear of components;
- measurement errors, sensor drift, or malfunctions in control systems.

In such cases, the Digital Twin performs a predictive and proactive function, enabling the identification of anomalies before they result in unplanned downtime or reductions in process quality. This capability supports predictive maintenance strategies, reduces costs associated with corrective maintenance, and increases overall operational safety.

Model errors or limitations

In some instances, the digital model may fail to fully represent certain operating configurations or dynamic scenarios. In this case, the misalignment between the real and virtual systems becomes an indicator of the model's validity limits. In machine learning–based contexts, this condition represents an opportunity to expand the training dataset and improve the predictive capability of the digital twin. For physics-based or hybrid models, it instead requires a critical revision of the underlying assumptions, adopted simplifications, and modelled parameters.

In conclusion, bidirectional communication between the physical twin and the digital twin constitutes the core of the Digital Twin paradigm. It not only ensures continuous and faithful model updating, but also enables advanced monitoring, optimization, diagnostics, and predictive maintenance functionalities that fundamentally transform the management of industrial plants. Through this integration, the Digital Twin becomes not merely an analytical tool, but a true cognitive partner of the physical system, capable of learning, adapting, and supporting both operational and strategic decision-making.

2.4 Management of Optimizers

The development of optimizers represents the natural evolution of the Digital Twin infrastructure as defined in the previous sections. Once the real system, the virtual model, a bidirectional communication interface, and a robust capability for interpreting system outputs are available, particularly through the systematic analysis of deviations between real and virtual behaviour, it becomes possible to exploit this information not only for monitoring and diagnostic purposes, but also to actively influence and guide the behaviour of the physical plant. At this stage, the Digital Twin transitions from a descriptive and interpretative tool into an active decision-support system.

Up to this point, the described architecture has primarily leveraged the information flow from the physical system to the virtual model. This flow enables state estimation, anomaly detection, and performance evaluation by continuously correlating real measurements with simulated outputs. The introduction of an optimizer, however, fundamentally extends this architecture by allowing the feedback loop to be closed. By exploiting the reverse communication channel, from the digital model to the machine, it becomes possible to automatically define setup configurations, adjust operating parameters, or suggest operational strategies aimed at improving system performance [11]. In this sense, the optimizer acts as the logical bridge between system understanding and system control (Figure 2).

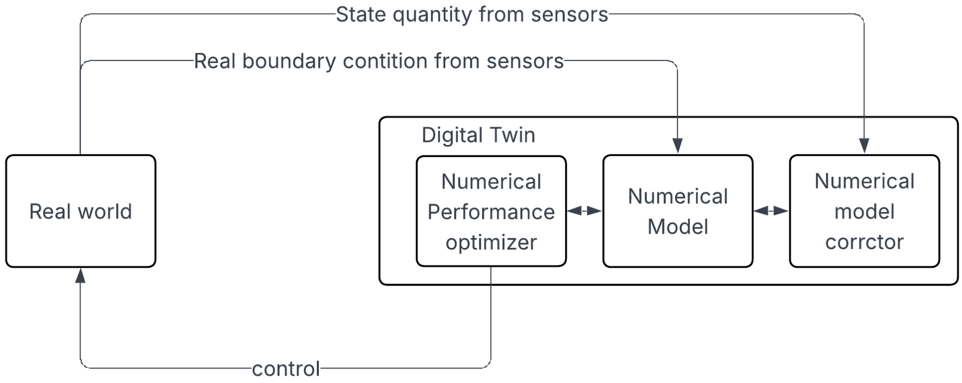

Figure 2 Digital Twin behaviour scheme

From a methodological perspective, decision-making systems responsible for setting or modifying operating parameters within a Digital Twin framework can be broadly classified into two main categories:

- Direct decision-making methods
- Iterative methods (true optimizers)

These two classes differ significantly in terms of complexity, computational requirements, and applicability, and are often complementary rather than mutually exclusive.

Direct decision-making methods

Direct decision-making methods are applicable when the condition to be addressed is known, well modelled, and clearly associated with a specific and recognizable physical manifestation. In such scenarios, the output derived from the correlation between the physical system and the virtual model exhibits an immediate and unambiguous mapping to the corrective action that must be undertaken. The Digital Twin, in this case, acts as a real-time observer capable of detecting predefined events or deviations and triggering predefined responses.

Typical examples of direct decision-making include:

- the identification of a mechanical failure or a clearly characterized anomaly, followed by the automatic activation of maintenance or shutdown strategies;
- the execution of safety interventions as a consequence of exceeding operational thresholds or violating design constraints.

In these situations, the decision logic is essentially linear: the correlation analysis identifies a characteristic scenario, the decision-making system interprets it based on predefined rules, a direct command is sent to the machine. This architecture is particularly suitable for situations involving discontinuous events, such as faults, alarms, or safety-critical conditions, where fast and deterministic responses are required.

Direct decision-making architectures offer several advantages. They are computationally lightweight, easy to validate, and capable of operating under strict real-time constraints. Moreover, their deterministic nature simplifies certification and integration into existing industrial control systems Figure 3. However, their main limitation lies in the requirement for prior knowledge of the triggering causes and the availability of explicit decision rules. As a result, they are inherently limited in their ability to manage complex, nonlinear, or evolving phenomena, for which the cause–effect relationship cannot be easily defined in advancev.

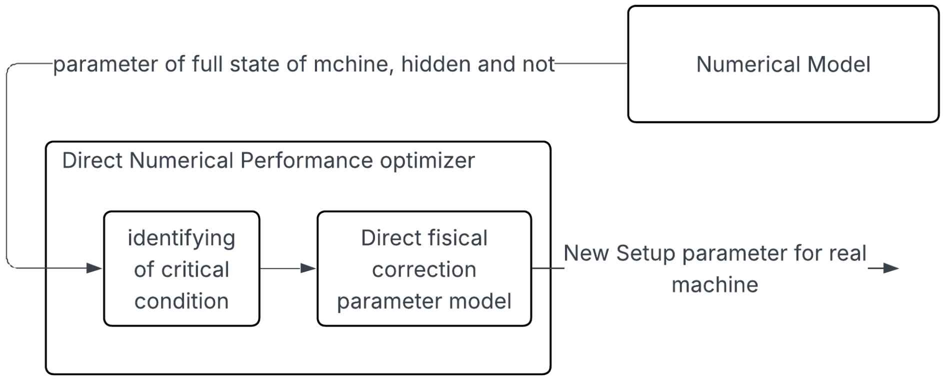

Figure 3 schematic representation of a direct decision-making architecture.

Iterative methods and true optimizers

Iterative decision-making methods, by contrast, are more closely aligned with the numerical concept of an optimizer and are particularly well suited to problems characterized by gradual evolution, trade-offs between multiple objectives, or uncertainty in system behaviour. In this paradigm, the virtual twin is exploited as a controlled experimentation environment, within which different operating configurations can be systematically evaluated in order to identify the one that optimizes a predefined objective function. Typical objectives include efficiency maximization, quality improvement, energy consumption reduction, cost minimization, or throughput maximization.

Within this approach, continuous correlation with the physical system plays a crucial role. First, it ensures that the virtual model remains aligned with the actual state of the machine, preventing the optimizer from converging toward solutions that are optimal only in a simulated but unrealistic context. Second, it enables the progressive validation of simulated operating choices through comparison with real-world outcomes. Finally, it allows optimization trajectories to be dynamically corrected when changes in operating conditions, external disturbances, or performance degradation are detected.

Iterative optimizers represent powerful and flexible tools; however, their computational nature introduces inherent limitations. Because multiple candidate solutions must be evaluated, often through repeated simulations, these methods require non-negligible computational time and are therefore less suitable for applications demanding immediate response or hard real-time performance. Nevertheless, their use is not precluded even in critical contexts, provided that appropriate precautions are adopted. For example, optimization may be performed on a reduced-order model, within constrained operating regions, or at a slower supervisory level rather than directly in the control loop (Figure 4).

In scenarios such as the early onset of a failure, performance drift, or the presence of highly stochastic phenomena that prevent the adoption of a purely rule-based strategy, iterative optimization executed on the Digital Twin can still provide significant value. By rapidly evaluating alternative scenarios in the virtual domain, the optimizer can support operators or automated control systems in selecting the most robust or least risky operational strategy, even when absolute optimality cannot be guaranteed.

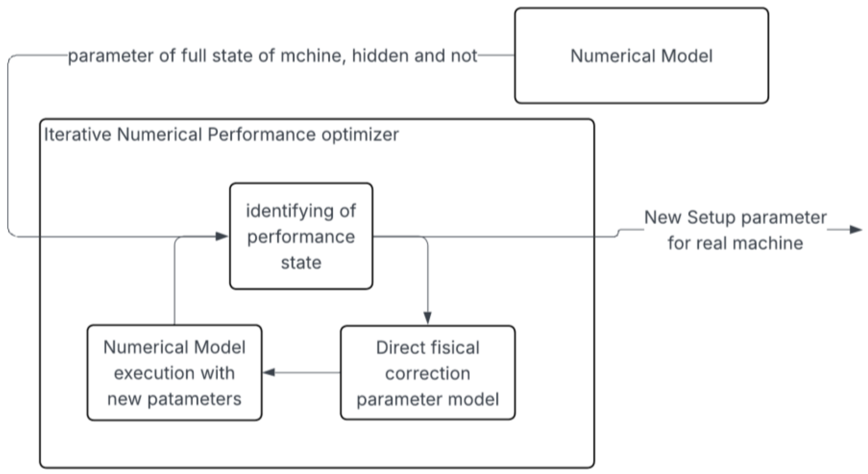

Figure 4 Iterative decision scheme

The integration of an optimizer within a Digital Twin framework completes the paradigm by transforming the digital twin into a tool that is not only diagnostic and predictive, but also proactive and prescriptive. It enables not only an understanding of how the system is currently operating, but also a systematic determination of how it should operate in order to achieve optimal or near-optimal performance under given constraints. In modern industrial contexts, the combination of bidirectional communication, evolutionary modelling, and iterative optimization represents one of the most powerful levers for increasing efficiency, reliability, adaptability, and overall decision-making capability in complex production processes.

2.5 Digital Twin layout

Having examined the fundamental components of a Digital Twin, it is now appropriate to describe in greater detail the possible practical system architectures and the modalities through which such systems can interact with the human user. From an information technology perspective, a Digital Twin can be interpreted as a multilayer infrastructure in which physical data acquisition, data transmission, virtual modelling, and user interaction are tightly integrated (Figure 5).

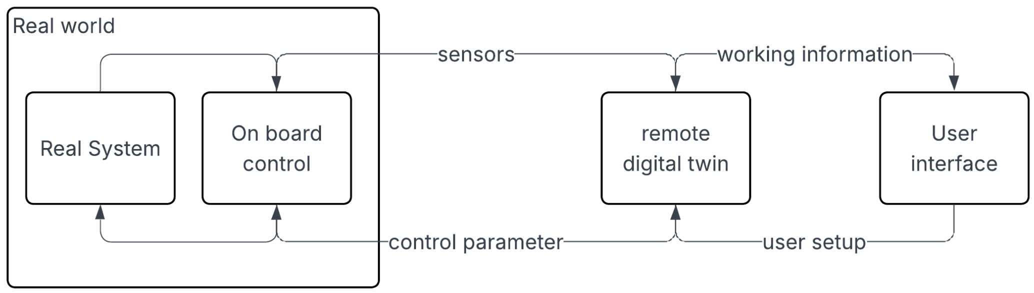

Figure 5 Digital Twin IT implementation scheme

For the correct and reliable functioning of the overall infrastructure, it is necessary to ensure an adequate technological maturity with respect to several key aspects, which can be broadly categorized into technological components and system interface and management components.

The Technological components includes components of a primarily technological nature and represents the foundational layer of the Digital Twin architecture. In particular, the performance parameters that characterizing a robust technological infrastructure are as follows:

- **Continuous access to measurement data related to the system under analysis**, which is essential to maintain an up-to-date and representative virtual counterpart of the physical system. This requirement implies the availability of persistent data streams and reliable communication channels between the physical asset and the digital environment.
- **Reliability and accuracy of the measurements**, as the validity of the Digital Twin's outputs and predictions strictly depends on the quality of the input data. Measurement uncertainty, sensor drift, and data loss must therefore be carefully addressed through appropriate sensor selection, calibration procedures, and fault detection mechanisms.
- **Number and relevance of the measurement channels**, which must be sufficient to capture the key physical variables governing the behaviour of the system. An inadequate or poorly selected set of sensors may lead to incomplete or misleading representations of the system dynamics.
- **Temporal and quantitative resolution with respect to the observed phenomena**, ensuring that the sampling frequency and measurement granularity are compatible with the time scales and amplitudes of the underlying physical processes. This aspect is particularly critical in systems characterized by fast dynamics or highly nonlinear behaviour.
- **An adequate infrastructure for the virtual environment**, including computational resources, data storage capabilities, and software platforms for simulation, data processing, and model updating. The complexity of the Digital Twin models and the volume of data involved often require scalable and high-performance computing solutions.

System interface and management components

In addition to the technological foundations, a Digital Twin must also include system interface and management components, which enable effective interaction between the digital system and human users. These components encompass visualization tools, dashboards, decision-support systems, and control interfaces, and they play a crucial role in translating complex data and model outputs into actionable information.

While the technological components are largely governed by objective criteria related to sensing, communication, and computation, and their adequacy has been widely investigated in the scientific literature, particularly in relation to the static and dynamic characteristics of the observed phenomena, the design of the interface and management layer is more strongly influenced by contextual factors. These include the specific operational environment, the skills and responsibilities of the end users, and the intended level of system autonomy.

In particular, the degree of automation directly affects how the Digital Twin interacts with human operators: in low-automation scenarios, the Digital Twin primarily serves as a monitoring and decision-support tool, whereas in highly automated systems it may actively influence or directly control the behaviour of the physical asset. Consequently, the interface design must balance transparency, usability, and control authority to ensure both system effectiveness and user trust.

2.6 Technological Requirements

The technological requirements necessary for the implementation of a Digital Twin project are strongly correlated with the intended objectives of the system and with the environmental conditions in which it operates. Such conditions include, but are not limited to, the physical location of the system, the available installation space, and the accessibility of supporting services such as power supply, network connectivity, and data storage infrastructure.

In order to correctly size and design an adequate technological infrastructure, it is essential to establish a clear correlation between the characteristics of the operating environment and the functional requirements imposed by the Digital Twin objectives. This alignment ensures that the technological solutions adopted are both technically feasible and economically sustainable, while still meeting the desired performance targets.

In most practical applications, two key aspects require particularly careful analysis.

The first is the system time response, which determines the capability of the Digital Twin to reflect the dynamic behaviour of the physical asset in a timely manner. This aspect directly influences decisions regarding sensor sampling rates, data transmission latency, computational performance, and model update frequency.

The second critical aspect is the quality of the acquired data, which encompasses measurement accuracy, resolution, reliability, and robustness against noise and data loss. High-quality data are fundamental for ensuring the credibility of the Digital Twin, especially when it is used for predictive analysis, optimization, or automated decision-making.

Together, these factors play a decisive role in defining the overall effectiveness of the Digital Twin infrastructure and must be carefully evaluated during the early design stages of the project.

2.6.1 Response Time

The first point of analysis in the design of an appropriate architecture is the evaluation of the response time required by the system [12]. A useful conceptual distinction can be made between **hard real-time** and **soft real-time** Digital Twins.

A hard real-time Digital Twin is required when interaction with the physical system imposes strict temporal constraints. In this case, the digital twin does not merely perform monitoring functions, but actively contributes to process control. If the model update or command generation occurs beyond a defined deadline, the system can no longer be considered reliable, and potentially hazardous situations may arise. This approach is typical of highly automated industrial environments, collaborative robotics, predictive control systems for critical machinery, and aerospace applications. For example, a digital twin that computes corrective trajectories in real time for a production robot must strictly guarantee deadline compliance: even a single delay could result in collisions or equipment damage. Such applications require deterministic platforms, communication protocols with guaranteed low latency, and algorithms optimized from a temporal standpoint.

In contrast, in soft real-time Digital Twins the temporal deadline is important but not strictly binding. Delays in model updates may degrade simulation quality, but without compromising system safety or overall functionality. This mode is typical of digital twins used for long-term predictive maintenance, scenario simulation, energy optimization, or non-critical monitoring. For instance, a digital twin of a smart building can update parameters such as temperature or energy consumption with delays of several seconds without negative effects: the model remains fully effective for trend analysis or anomaly prediction. In these cases, flexibility, scalability, and the ability to process large volumes of data are generally prioritized over strict determinism.

The adoption of a hard or soft real-time Digital Twin therefore depends on the impact of timing on system safety and correctness. Where direct control of critical processes is involved, a hard real-time Digital Twin is required; where analysis and decision support prevail, a soft real-time approach is sufficient. This distinction guides the selection of technologies, models, and architectures to be adopted.

Even if not in a strictly formal manner, it is already possible to observe that hard real-time requirements tend to favour what has previously been defined as a direct approach or one based on deterministic modelling, whereas systems requiring a soft real-time approach more naturally lend themselves to iterative model-based strategies.

The definition of the required response time also directly influences the supporting hardware infrastructure for the models. In fact, it is difficult to rely on remote infrastructures for hard real-time systems; instead, a local architecture deployed close to the machine, often with simplified models, is generally preferred. Conversely, in the case of a soft real-time optimizer, a remote architecture becomes feasible, allowing access to greater computational resources compared to a purely local system.

2.6.2 Quality of Acquired Data

The definition of the requirements needed during the setup of the data acquisition infrastructure represents a fundamental step to ensure the effectiveness and reliability of a Digital Twin model. The quality and quantity of the collected data constitute the foundation upon which the digital twin's ability to faithfully and continuously represent the behaviour of the real system is built [13]. For this reason, each acquisition channel must be technologically suited to support the specific type of signal, ensuring performance that is consistent with both the physical characteristics of the monitored process and the requirements of the computational model.

In particular, three fundamental parameters must be precisely defined:

- **sampling time of the data logger**, i.e., the sampling frequency required to correctly capture the temporal evolution of the signals;
- **resolution**, intended as the system's capability to discriminate minimal variations in the measured quantity;
- **measurement range**, namely the operating interval within which the sensor maintains accuracy and linearity.

These are complemented by additional requirements related to accuracy, precision, long-term stability, and sensor drift control, which are essential to prevent systematic errors that could directly affect model performance.

Another crucial aspect concerns the temporal synchronization of acquisitions: in complex systems, the correct alignment of data originating from different sensors is essential to coherently represent physical interactions. Similarly, appropriate filtering techniques and noise management strategies must be implemented to improve raw data quality and reduce artifacts in simulations.

With regard to data quantity, the infrastructure must be designed to ensure adequate coverage of the different operating conditions of the system, including steady-state regimes, transient phases, and potential anomalous conditions. The amount of available data directly influences the robustness of the parameter identification process, model validation, and the Digital Twin's ability to perform accurate predictions. When real-time simulations are required, it also becomes necessary to implement continuous data acquisition and update mechanisms, with particular attention to latency and communication network availability.

An effective Digital Twin model also requires a robust data governance infrastructure: each data item must be traceable, enriched with the relevant metadata (timestamp, measurement units, acquisition conditions, sensor information), and stored in standardized formats to enable integration with other enterprise systems. Data quality management further includes procedures such as data cleaning, outlier detection, missing data handling, and normalization, all of which are indispensable to ensure consistency and reliability in the modelling process.

In summary, within the context of a Digital Twin, requirements related to data quality and quantity extend well beyond the purely hardware aspects of acquisition channels, encompassing the entire information flow architecture, from sensing to transmission, from storage to processing. Careful design of these aspects is essential for the digital twin to effectively fulfil its monitoring, prediction, and decision-support functions.

2.7 System interface and management components

Beyond the technological foundations consisting of sensors, models, communication infrastructures, and computational capabilities, an effective Digital Twin must include a structured set of interface and management components[14]. This layer represents the point of contact between the digital system and human users, enabling data interpretation, operational interaction, and support for decision-making processes. Without an appropriate interface and well-designed management mechanisms, even the most accurate Digital Twin from a modelling perspective risks remaining a scarcely usable or unreliable tool in the operational context.

The interface and management layer therefore serves a dual function. On the one hand, it translates complex information, often multidimensional and dynamic, into representations that are understandable and usable. On the other hand, it allows users to influence system behaviour, for example by modifying parameters, simulating alternative scenarios, or issuing commands to the physical system through the Digital Twin.

Visualization tools constitute one of the central elements of a Digital Twin interface. These include interactive dashboards, time-series plots, state maps, three-dimensional models, and animations that represent the evolution of the physical system over time. The primary objective is not merely to display raw data, but to provide a concise and contextualized view of the system state[15].

Dashboards must be designed with the specific needs of end users in mind. Line operators, maintenance managers, process engineers, and decision-makers have different goals and levels of expertise, and therefore require different views of the same Digital Twin. In this regard, interface customization and the ability to configure key performance indicators (KPIs) represent critical success factors.

Another relevant aspect concerns support for situational awareness. Through clear and real-time updated visualizations, the Digital Twin can help users quickly identify anomalies, deviations from expected behaviour, or potentially critical situations, thereby reducing cognitive load and improving decision timeliness.

Alongside visualization, modern Digital Twins increasingly integrate decision support systems (DSS). These systems leverage analytical models, what-if simulations, optimization algorithms, and artificial intelligence techniques to provide operational recommendations to users[16].

The added value of a DSS lies in its ability to connect the current state of the system to possible future actions and their predicted effects. For instance, a Digital Twin may suggest preventive maintenance actions, alternative control strategies, or adjustments to process parameters to improve efficiency or reduce risks. However, the effectiveness of such recommendations largely depends on how they are presented to the user.

It is essential that recommendations be accompanied by understandable explanations that make the underlying reasoning transparent. So-called explainability thus becomes a key requirement, especially in industrial and safety-critical contexts, where operators must be able to evaluate and justify the decisions made.

The degree of system automation profoundly influences the design of control interfaces. In low-automation systems, the Digital Twin primarily plays a monitoring and decision-support role, while the human operator remains the main actor responsible for actions on the physical system. In these cases, the interface should prioritize informational clarity, ease of data exploration, and the ability to compare alternative scenarios.

Conversely, in highly automated systems, the Digital Twin may have the capability to directly influence or even control the behaviour of the physical asset. This requires interfaces that support automation supervision, the definition of operational constraints, and human intervention in the event of abnormal situations. In this context, the concepts of human-in-the-loop or human-on-the-loop become central: the user does not necessarily execute every action, but retains a supervisory and validation role.

Within this framework, it is essential to clearly define responsibilities and control authority. A poorly designed interface may generate ambiguity, excessive reliance on automation, or, conversely, rejection of automated functionalities, thereby compromising the overall effectiveness of the system.

Unlike core technological components, which can be evaluated according to relatively objective criteria, the interface and management layer is strongly dependent on the application context. The operational environment, organizational procedures, users' level of training, and applicable regulations significantly influence design choices.

For this reason, the design of Digital Twin interfaces should follow a user-cantered approach, involving end users from the early stages of development. Techniques such as task analysis, iterative prototyping, and usability testing make it possible to identify critical issues and progressively improve the user experience[17].

A particularly relevant aspect is the development of user trust in the system. Information transparency, consistency of representations, and predictability of Digital Twin behaviour contribute to establishing an effective collaborative relationship between humans and the digital system.

Interface and management components therefore represent a key element for the success of Digital Twins, as they determine how technological capabilities are actually exploited in real-world contexts. A Digital Twin is not only an accurate model of the physical system, but also a tool for interaction, communication, and decision support.

Balancing usability, transparency, and level of control, while accounting for the degree of automation and user characteristics, is a complex but fundamental challenge. Proper design of this layer enables the value of the Digital Twin to be maximized, improving operational performance, safety, and user trust.

3 Filter Press

In the modern mining sector, the management of processing residues represents one of the most delicate and strategic challenges. The increase in extraction volumes, the need to reduce environmental impacts, and the progressive adoption of closed-loop processes require solid–liquid separation technologies that are increasingly efficient, reliable, and scalable. Within this context, large-capacity filter presses have assumed a central role, becoming essential tools for the production of filtered tailings and for reducing the water content of waste materials. Their adoption not only enables the optimization of tailings pond management, but also improves the geotechnical stability of deposits, significantly reducing the environmental footprint of the entire mining operation [18].

State-of-the-art filter presses adopt technological solutions that allow the treatment of high slurry flow rates through fast and repetitive cycles, while simultaneously maintaining high standards of safety and operational continuity. Thanks to fully automated handling systems, high-performance filter media, and structures designed to withstand the stresses typical of abrasive materials, these machines are capable of operating under harsh conditions while ensuring result uniformity and minimizing plant downtime. The growing focus on sustainability has also driven manufacturers to develop increasingly accurate water recovery systems, with the aim of reducing the overall water demand of mining plants and promoting a more responsible approach to resource management.

Another key aspect is the ability of these filter presses to be integrated into highly automated processes supervised by advanced control systems. The adoption of distributed sensors, continuous pressure monitoring, predictive diagnostics, and connectivity with data acquisition and analysis systems enables real-time optimization of the filtration cycle, adapting it to variations in the characteristics of the incoming material. This capability allows the production of filter cakes with consistently low moisture content, which is essential for the safe handling and storage of tailings, as well as for potential relocation or secondary reuse processes.

The increasingly large dimensions reached by these machines reflect the mining industry's trend toward high-productivity plants, in which large volumes must be managed with the smallest possible number of filtration lines. Massive supporting structures, high-pressure hydraulic systems, and high-strength frames make it possible to maintain consistent performance even under intensive cycles and particularly demanding loads. In parallel, the introduction of fast-opening devices, panel shaking systems, and automated cleaning procedures contributes to reducing the time between successive cycles, further increasing overall productivity.

In summary, filter presses employed in mining applications represent the outcome of a technological evolution that combines production capacity, sustainability, and operational robustness. Their presence in tailings treatment plants is no longer merely an option, but a strategic choice to ensure production continuity, reduce environmental risk, and achieve more efficient use of water resources. This chapter will examine in detail the operating principles, structural architectures, and operational logics of these machines, highlighting their crucial role within modern mining processes.

3.1 Press Structure

The machine under investigation is a filter press designed for the treatment of waste slurry generated by mining processes for metal extraction. This type of machinery can be broadly classified into two main families based on the plate pack closing technology: **head-cylinder filter presses** and **tie-rod filter presses**.

The former, generally characterized by smaller dimensions, are only marginally used for the treatment of large material volumes, such as mining waste streams. The latter, which fall more directly within the scope of the present analysis and are characterized by larger dimensions, are the most widely adopted solution for the treatment of mining extraction residues. Among tie-rod filter presses, two subfamilies can be identified, commonly referred to in commercial terminology as **"P" type** and **"F" type** machines.

The former ("P" type) are machines used in processes that, due to the physical nature of the material being treated, require long filtration times, typically on the order of several hours per cycle. These machines are therefore designed to maximize the amount of material that can be filtered within the press, prioritizing filtration volume while accepting longer operational times between successive filtration cycles. They are characterized by a large number of plate packs, occupying nearly the entire length of the machine, and by a moving head that is actuated along the full stroke by the closing cylinders. In this configuration, the cylinder housings are fixed to the stationary head, while the piston rods are connected to the moving head via tie rods. At equal overall machine dimensions, "P" type presses generally provide a larger filtration volume than "F" type presses, at the expense of significantly longer opening and closing times.

Filter presses of the second type ("F") exhibit, for comparable machine dimensions, smaller filtration volumes but significantly reduced actuation times. These machines are typically employed in plants where the filtration phase has a duration comparable to the mechanical movement times, generally on the order of several minutes. From the perspective of plate pack logic, they behave similarly to "P" type machines; however, the moving head is actuated over most of its stroke by rack-and-pinion hydraulic drives, which provide relatively fast motion. Once the moving head reaches the closing position, the tie rods are engaged and perform only the final loading stroke required to ensure sealing (Figure 6).

Figure 6 at left type P machine at right type F

Regardless of the actuation and closing system, both machine types are conceptually designed to form filtration chambers between the individual plates, as illustrated in Figure 7.

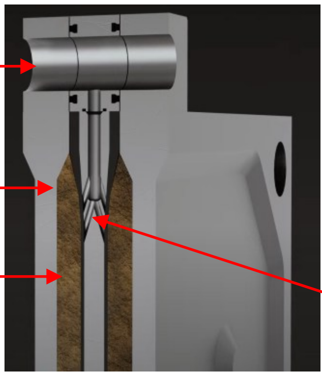

Figure 7 filtration chamber scheme

Once stacked and alternated with the corresponding filter cloths, the individual plates form the filtration chambers, within which the slurry to be filtered is introduced.

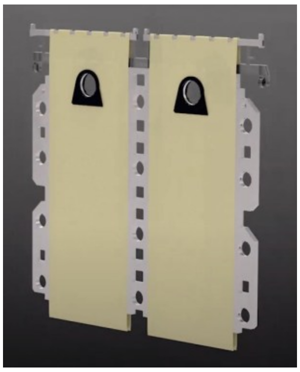

Figure 8 filtration cloth

The filtration process takes place within these chambers. The machine is therefore equipped with a set of subsystems responsible for filling, discharge, handling, and maintenance of the filtration chambers. These subsystems can be grouped into the following categories:

- **Mechanical and hydraulic systems**, responsible for the movement and actuation of components;
- **Filter cloth and plate washing systems**;
- **Slurry feeding and pressing systems**.

All machines of this type are volumetric machines, processing a defined volume of material during each operating cycle. As cycles progress, both the filter cloths and the plates are subject to degradation and to the accumulation of contaminants. These phenomena respectively require periodic component replacement and continuous cleaning operations. With regard to maintaining machine cleanliness, two interventions are currently carried out systematically: a spray washing at the end of each cycle, and a high-pressure washing operation performed directly on the filter cloths by means of a dedicated robotic system.

In the following sections, the filtration process and the washing processes are described in greater detail.

3.2 Filtration Mechanism

With reference to the filters under analysis, namely those dedicated to the treatment of waste materials, and therefore characterized by large dimensions and high throughput, the filtration mechanism is purely mechanical. The process can be schematically divided into the following phases:

- **Closing**
- **Filling**
- **Pressing**
- **Squeezing**
- **Cake washing**
- **Blowing**
- **Discharge**

3.2.1 Closing Phase

The closing phase consists in bringing the plates into contact and subsequently applying load in order to ensure proper sealing. In the case of the "F" type machine, the moving head requires a long and rapid opening stroke to allow plate separation and ensure fast discharge. For this reason, the closing stroke is implemented through three main phases:

- **Mechanical actuation via hydraulic motor**;
- **Closing by pull-in of flow-controlled hydraulic cylinders**;
- **Sealing pressurization via pressure-controlled hydraulic cylinders**.

Most of the opening and closing stroke is actuated by two rack-and-pinion mechanisms driven by hydraulic motors. Once the moving head comes into contact with the plate pack and is in a position to engage the hydraulic tie rods, the rack-and-pinion drives are disengaged and the head is subsequently driven by the hydraulic cylinders. During hydraulic closing, actuation is initially performed under flow control to ensure proper alignment of the head; once a predefined counter-pressure threshold is reached, the control mode is switched to pressure control until the maximum allowable clamping force for the hydraulic system is achieved.

The control mode transition based on the counter-pressure measured by the cylinders ensures adequate uniformity of contact prior to load increase, thereby limiting the risk of mechanical instability.

3.2.2 Filling and Pressurization

Once the machine is closed, it is ready for the filling phase and for the first filtration stage, namely pressing. It is important to specify that the hydraulic closing load does not provide any direct contribution to the filtration process, but solely serves to ensure hydraulic sealing between the plates. The mechanical pressing action responsible for filtration is instead achieved through the compression of the slurry inside the machine, which is generated by the feed pumps.

This clarification explains why the filling and filtration actions are, in practice, two consecutive phases carried out with the same mechanical configuration of the machine, but with different objectives and pump control strategies. The filling phase corresponds to the time interval during which the machine is flooded, with the pumps configured to operate at maximum flow rate. During filling, a condition of constant slurry inflow rate is established, while the counter-pressure increases as a result of the accumulation of solid material. Once the pressure set-point is reached, the pumps switch to steady-state operation under constant pressure control (Figure9).

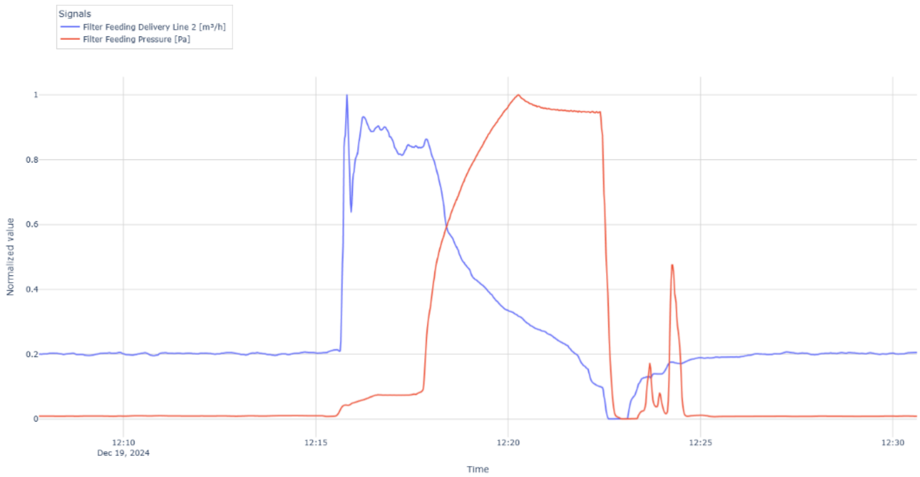

Figure 9 Pressure and flow during cycle

The filling and pressing phases can therefore be identified as two sub-phases of the filtration process, namely a **constant-flow filling phase**, followed by a **constant-pressure filtration phase**.

The filtration mechanism adopted in this type of machine follows purely mechanical principles. It consists of compressing the slurry within a volume delimited by partially permeable membranes, or, in other terms, membranes that are highly permeable to water and only weakly permeable to the solid phase. The filtering action is thus induced by the forced pumping of the slurry through a mechanical resistance that acts predominantly on the solid fraction.

The filtering medium is therefore composed of both the membrane itself and the accumulation of solid material on its surface; the latter represents a non-negligible and fundamental contribution to the formation and behaviour of the filtering medium. When considering filtration as governed exclusively by mechanical phenomena, it is possible to relate pressure drop and flow rate through **Darcy's law [19]**.

$$Q = \frac{A\Delta p}{\mu R}$$

The resistance to flow through the filtration medium can be observed as two resistances connected in series: one associated with the membrane and the other with the solid phase accumulated on its surface. The former remains constant throughout the filtration cycle, while the latter increases as the thickness of the filter cake grows.

Under the assumption of an incompressible solid, the fluid-dynamic resistance varies linearly with the volume of accumulated solid material.

$$R_c = \alpha w$$

In the case of a compressible solid, the specific cake resistance αbecomes dependent on the pressure drop across the filter medium. Under these conditions, an average value of αcan be introduced, which can be evaluated through an integral mean as follows:

$$\Delta p = \Delta p_{\mathrm{c}} + \Delta p_{\mathrm{m}}$$

$$\frac{1}{\alpha_{\mathrm{av}}} = \frac{1}{\Delta p_{\mathrm{c}}} \int_0^{\Delta p_{\mathrm{c}}} \frac{d(\Delta p_{\mathrm{c}})}{\alpha}$$

Clearly, the specific cake resistance can be expressed as $\alpha = f(\Delta p)$, and this functional relationship can be derived from experimental correlations, semi-empirical models, or a combination of both.

In any case, the machine currently operates under the assumption of an incompressible solid, and the operating parameters are set in order to achieve working points within the incompressibility regime. At present, this is accomplished through laboratory testing of the slurry to be treated, combined with empirical knowledge derived from machine operation.

Once the permeabilities of the involved media have been defined, it becomes possible to develop Darcy's relation in terms of filtrate volume and pressure drop.

$$\frac{\mathrm{d}t}{\mathrm{d}V} = \alpha\mu c \frac{V}{A^2 \Delta p} + \frac{\mu R}{A \Delta p}$$

From this formulation, the material constants can be identified through two auxiliary constants:

$$\frac{\mathrm{d}t}{\mathrm{d}V} = a_1 \frac{V}{A^2 \Delta p} + b_1 \frac{1}{A \Delta p}$$

Considering the operation of the machine, the filtration period can be divided into two phases:

- **Machine filling phase**;
- **Pressing phase**.

In the first phase, the machine is being filled and therefore operates under conditions of constant and maximum achievable flow rate. In the second phase, the system enters the pressing stage, operating at constant pressure drop Δp, which is imposed as a design parameter of the machine and is generally considered constant over the machine's service life.

Once these two operating phases have been defined, Darcy's law can be evaluated in two steps: the first under constant flow rate conditions and the second under constant pressure conditions.

Constant flow rate operation

$$\Delta p = \alpha\mu c \frac{Q^2}{A^2} t + \mu R \frac{Q}{A}$$

$$\Delta p = a_1 v^2 t + b_1 v \qquad v = \frac{Q}{A}$$

Under the assumption of an incompressible solid and constant flow rate operation, the pressure drop can be observed to vary linearly with time.

Constant pressure drop operation

During the constant pressure operation phase, still under the assumption of an incompressible solid, Darcy's law can be integrated as follows:

$$\int_0^A \mathrm{d}t = \frac{a_1}{A^2\Delta p}\int_0^V V\mathrm{d}V + \frac{b_1}{A\Delta p}\int_0^V \mathrm{d}V$$

$$t = a_1 \frac{V^2}{2A^2\Delta p} + b_1 \frac{V}{A\Delta p}$$

The above relationship can be expressed in linear form as follows, namely in terms of flow rate as a function of the filtrate volume.

$$\frac{t - t_s}{V - V_s} = \frac{\alpha\mu c}{2A^2\Delta p}(V + V_s) + \frac{\mu R}{A\Delta p}$$

As in the constant flow rate condition, from the experimental identification of the slope and the intercept of the linear trend in the corresponding plot, it is possible to derive the characteristics of the solid component of the slurry and the specific resistance of the membrane.

Knowledge of the parameters α(specific resistance of the solid), R(membrane resistance), and c(mass of solid per unit volume of filtrate) makes it possible to establish a relationship between the cycle time and the amount of filtrate produced, and therefore with the amount of processed slurry.

Based on the characteristics described above, the following objectives can be identified:

- **Implementation of a dynamic computation system**, operating throughout the machine's service life, for the continuous estimation of the constants a_1 and b_1, enabling adaptive adjustment of the filtration time and avoiding idle times or premature machine opening;
- **Computation of the parameter α as a function of** Δp, in order to prevent and/or assess compressibility phenomena of the solid phase.

3.2.3 Squeezing

The squeezing process is an optional phase characteristic of so-called **membrane filter presses**, in which mechanical compression of the filter cake is performed by means of a pneumatic diaphragm. In the machine under investigation, this phase is not implemented; however, a brief description of its operation is provided for the sake of completeness (Figure 10).

Following the pressing phase, filter cakes are obtained with a significantly high solid concentration, yet still containing a non-negligible amount of residual water. The subsequent processes are primarily aimed at further reducing the water content. The squeezing phase applies a mechanical compression to the cake in order to compact the solid phase and expel the remaining water. When this phase is included in the filtration cycle, a dedicated pneumatic diaphragm is present within the filtration chamber and is inflated accordingly.

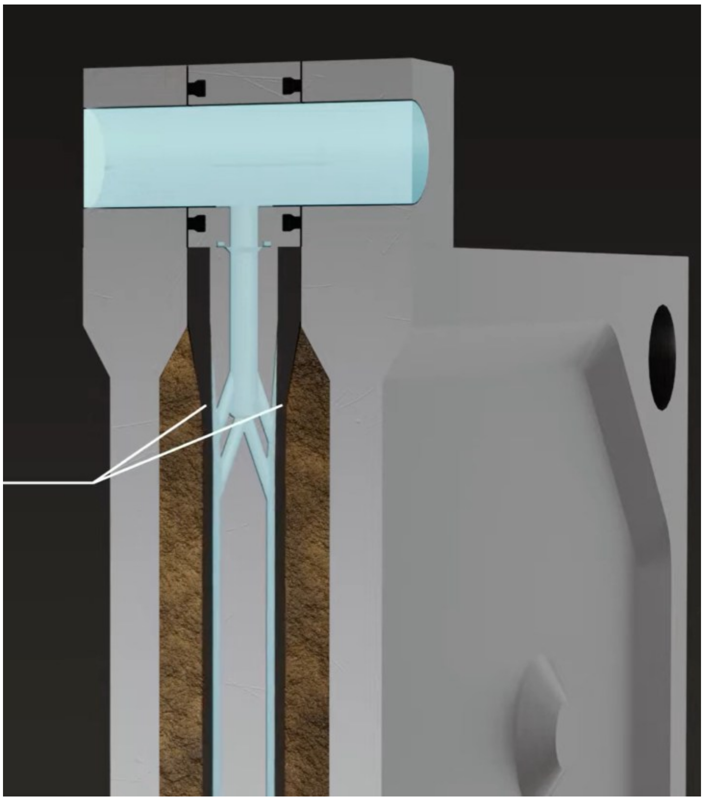

Figure 10 filtration chamber during pressing

3.2.4 Cake Washing and Air Blowing

The cake washing and blowing phases represent complementary operations within the filtration cycle and are based on a common operating principle. In both cases, the filter cake formed inside the filtration chambers is subjected to a forced flow, of water and air, respectively, that passes through the thickness of the solid layer deposited on the filter cloths. The purpose of these operations is to improve the final quality of the solid phase by acting on its chemical composition and on its residual moisture content.

The primary objective of the cake washing phase is the removal of harmful particles or undesired elements that are readily soluble in water. This operation is adopted, for example, when it is necessary to reduce the concentration of specific chemical compounds present in the solid phase, in order to meet product quality requirements, regulatory constraints, or needs related to subsequent disposal or reuse stages of the filtered material.

From an operational standpoint, cake washing is performed by exploiting the filtrate discharge lines, which are used in reverse to introduce clean pressurized water into the filtration chambers. The washing water flows through the cake thickness, permeating its porous structure and dissolving the soluble species present. The washing liquid, enriched with the removed substances, is then conveyed to the discharge lines, allowing the contaminants to be separated from the solid phase.

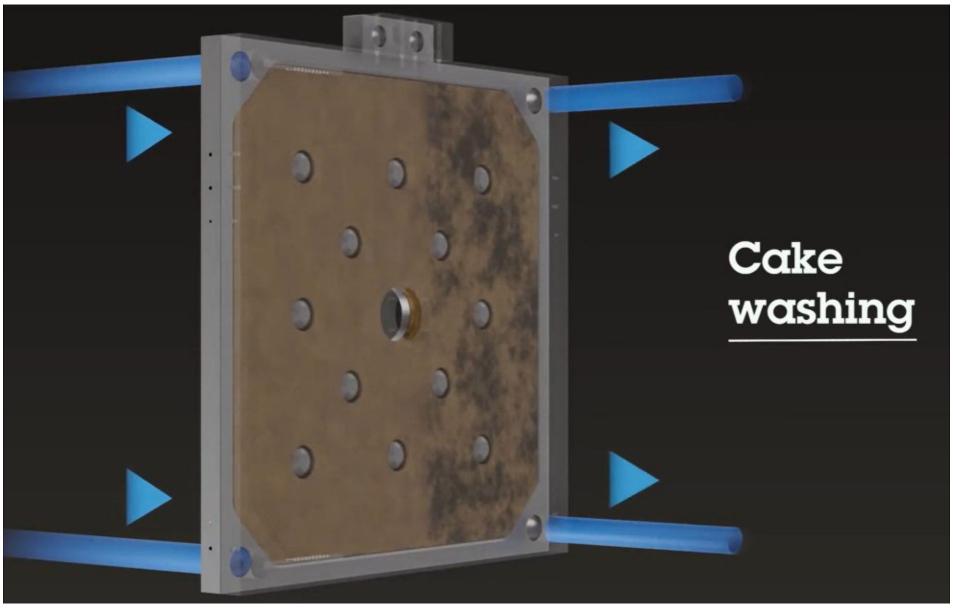

Figure 11 cake washing flux

The effectiveness of the washing process depends on several factors, including the inlet water pressure, the total volume of fluid employed, the permeability of the filter cake, and the physico-chemical properties of the substances to be removed. In this case as well, process parameters are generally set on the basis of empirical considerations and well-established operational experience (Figure 11).

Following the washing phase, the cake blowing operation is carried out, which is conceptually analogous but implemented through the use of compressed air. The primary purpose of the blowing phase is to reduce the residual moisture content of the cake by promoting the mechanical removal of water retained within the porous structure of the solid. This phase is particularly important for improving the handling properties of the discharged material and for reducing costs associated with transport and storage.

Blowing is generally performed using the same lines employed for washing, by introducing pressurized air that flows through the cake and forces the residual water outward. In most cases, the process is carried out in counter-current with respect to the washing phase, in order to maximize the effectiveness of the drying action and to promote the removal of residual water accumulated in the innermost regions of the cake(Figure 12).

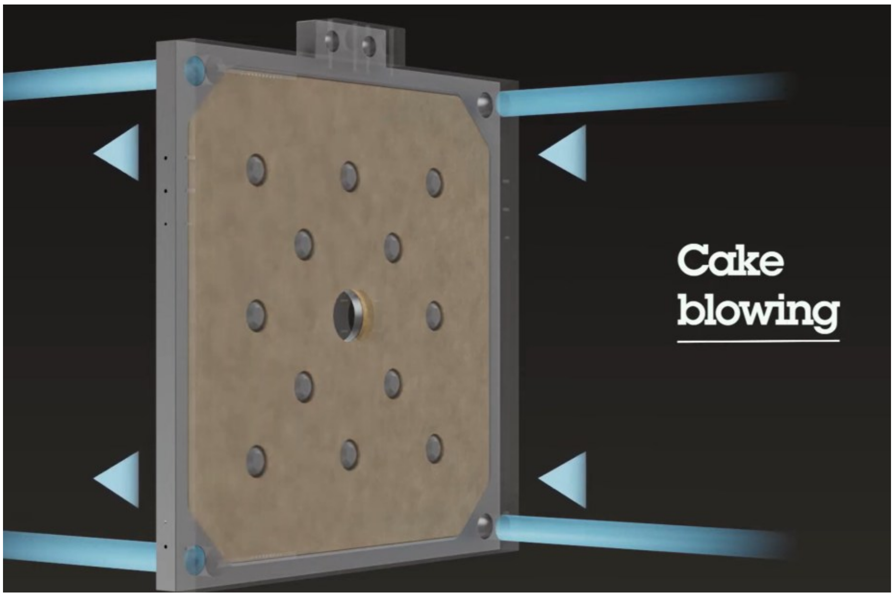

Figure 12 cake blowing flow

The effectiveness of the blowing phase also depends on various operating parameters, such as air pressure, process duration, cake porosity, and the degree of compaction achieved during filtration. Insufficient blowing may result in an excessively wet cake, whereas excessive blowing can lead to unnecessary energy consumption or undesired mechanical stresses on the filter cloths.

Overall, the cake washing and blowing phases represent fundamental steps for controlling the quality of the solid product discharged from the filter press. Their proper integration within the filtration cycle makes it possible to obtain a material with chemical and physical characteristics that comply with process specifications, while simultaneously highlighting the potential interest in optimization strategies based on more precise and adaptive control of the operating parameters.

3.2.5 Discharge

Once the filtration operations have been completed, the subsequent phase of the filter press operating cycle is represented by cake discharge, which takes place through machine opening and the gravitational fall of the solid material. This phase plays a particularly important role, as it enables not only the removal of the filtered product, but also an indirect assessment of the overall performance of the filtration cycle that has just been completed.

The discharge procedures differ depending on the type of filter press considered. In **F-type filters**, opening of the plate pack allows the simultaneous discharge of all cakes formed in the different filtration chambers. By contrast, in **P-type filters**, discharge occurs sequentially, with cakes falling one chamber at a time, in accordance with the different plate opening kinematics adopted in this configuration. This difference results in variations both in the overall discharge time and in the dynamics of cake detachment from the filter cloths.

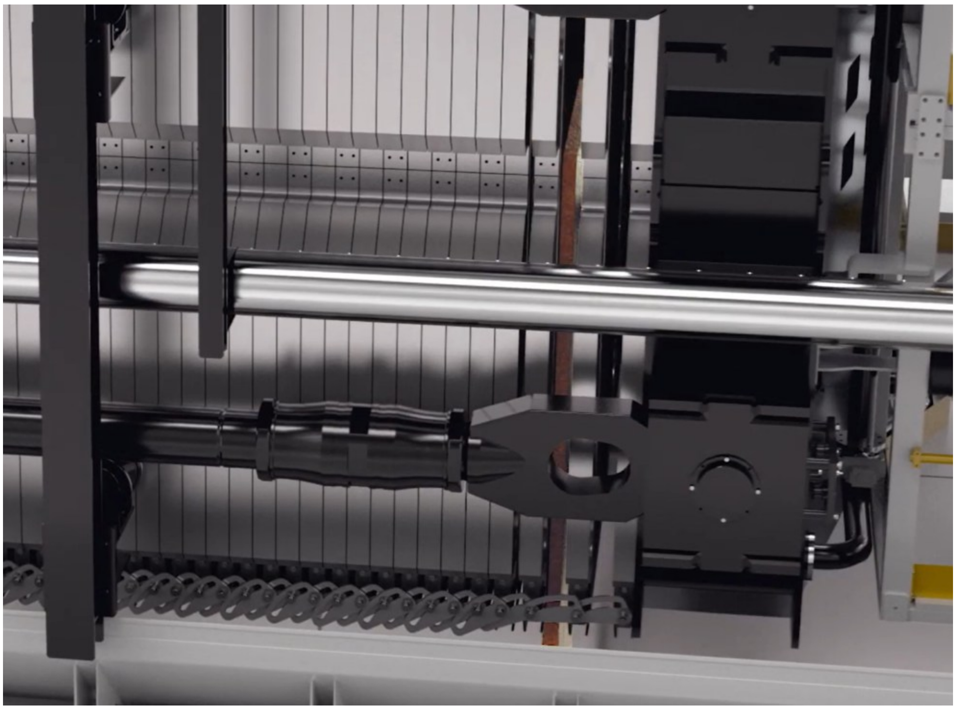

Figure 13 machine opening process

The solid material discharged from the filter press is conveyed into a collection screw conveyor positioned beneath the machine. The screw conveyor is responsible for continuously transferring the filter cake onto a belt conveyor, which subsequently transports the material to a dedicated storage area or to further processing stages. The belt conveyor is equipped with dedicated instrumentation for monitoring the discharged product: in particular, the mass of the material is measured through an integrated weighing system, and the surface moisture content of the cake is detected by means of an infrared hygrometer.

The discharge phase is of significant operational and diagnostic interest, as it represents the stage at which the actual performance of the filtration cycle can be directly assessed. The quantities measured and observed during this phase make it possible to evaluate several key aspects of the process, including:

- the amount of solid material discharged, which is indicative of solid–liquid separation efficiency;
- the quality of the discharged solid phase, in terms of residual moisture content and cake consistency;
- the ability of the cake to detach from the filter cloths, a parameter closely related to the cleanliness state of the cloths and to the operating conditions of the filtration cycle.

Incomplete or irregular cake detachment may indicate issues such as excessive adhesion of the solid to the cloth, insufficient cleaning of the filtering surfaces, or non-optimal filtration conditions.

In **F-type filters**, with the machine fully open, an additional cloth-shaking phase is also implemented, aimed at promoting the detachment of any residual material remaining adhered to the filtering surfaces. The system consists of two electric motors driving an equal number of camshafts, positioned above the plate pack, one on the right side and one on the left side. During operation, the cams mechanically engage the filter cloths, inducing dynamic excitation in the form of shaking.

The shaking duration is set as a constant value, defined primarily on the basis of empirical and operational considerations.

3.3 Washing Mechanisms

The washing phases represent a fundamental element of the maintenance process of filter presses, as they ensure the long-term preservation of filtration performance and the correct mechanical functionality of the machine. During normal operation, a progressive accumulation of contaminants occurs on all components in contact with the filtration chambers, making the adoption of dedicated and properly designed cleaning systems necessary.

In particular, the main fouling phenomena that can be observed involve:

- the **filter cloths**, which are subject to the accumulation of solid residues;
- the **plate sealing surfaces**, on which solid material originating from the filtered product tends to deposit.

The accumulation of residues on the filter cloths leads to a twofold criticality. On the one hand, filtration performance deteriorates due to partial clogging of the cloth pores; on the other hand, increased adhesion between the filter cake and the cloth surface results in difficulties during cake detachment at the end of the filtration cycle. Similarly, the presence of solid material on the plate sealing surfaces is particularly critical for the correct closing of the plate pack. Under these conditions, residues may compromise sealing between mating surfaces, causing fluid leakage or, in more severe cases, mechanical instability of the plate pack during the compression phase.

These considerations highlight the need for careful monitoring of fouling phenomena and for the implementation of appropriate mitigation strategies. The mechanical solutions currently adopted on filter presses are mainly based on two different washing systems:

- **rain washing**;
- **high-pressure washing**.

Each of these methodologies has been developed to address both fouling mechanisms described above, albeit with different specific effectiveness. Rain washing is primarily oriented toward cleaning the plate sealing surfaces, whereas high-pressure washing is mainly dedicated to cleaning the filter cloths.

Rain washing

Rain washing consists of a cleaning procedure performed with the machine in the open configuration, during which the plate sealing surfaces are flooded by means of water jets. The system is composed of six water supply pipes, arranged coaxially with the machine axis and positioned along the edges of the plate pack. These pipes are equipped with nozzles appropriately oriented to direct the water jets toward the plate sealing surfaces.

The process is carried out at every filtration cycle and is controlled in terms of delivered volume. For each cycle, the machine is kept open and the washing system is activated for a predefined duration, such that a fixed volume of water is supplied. The washing volume is set *a priori* based on empirical considerations and consolidated operational experience.

From a hydraulic standpoint, the system exhibits good repeatability over time. Aside from component wear, the hydraulic characteristics of the circuit remain substantially stable, as do the operating parameters and performance of the pressure vessels, which are generally sufficient to ensure maintenance of the required supply pressure.

In contrast, the behaviour of the water jets is inherently stochastic and difficult to predict. This variability can be attributed to several factors, including:

- non-uniform and not perfectly repeatable positioning of the plates with respect to the nozzles;
- variations in distance and orientation between the nozzles and the surfaces to be cleaned;
- interaction between the water jet and the material to be removed, which may vary depending on the physical characteristics of the residue (adhesion, particle size distribution, moisture content) and its spatial distribution on the surface.

These aspects make rain washing an intrinsically poorly controllable process, characterized by significant variability in the achieved cleaning effectiveness, despite the stability of the hydraulic supply conditions.

High-pressure washing

High-pressure washing relies on mechanisms that are generally more controllable than those of rain washing, but it acts directly on the filter cloths. The process involves the use of an automated system consisting of a robotic unit that, with the machine fully open, inserts a washing bar equipped with high-pressure water nozzles between the filter cloths.

In the analysed configuration, the system is composed of two washing curtains that descend simultaneously, allowing two filtration chambers, and therefore four filter cloths, to be cleaned during each washing cycle. This system is particularly effective in removing solid residues adhered to the cloth surfaces, significantly restoring cloth permeability and improving the quality of cake detachment.

Despite its high effectiveness, high-pressure washing cannot be performed frequently due to the significant machine downtime it entails, typically on the order of one to two hours. This constraint makes it necessary to strike a compromise between cleaning efficiency and plant operational availability. As in the case of rain washing, high-pressure washing is scheduled based on the number of filtration cycles, and the intervention parameters are set as fixed values, determined primarily through operational experience.

4 Digital Twin Setup for the Filter Press

In accordance with the definition of Digital Twin provided in Chapter 2, the project concerning the filter press was structured by defining the objectives to be achieved and the strategies to be adopted.

4.1 Objective Definition

The present work aims to identify and analyse strategies for improving the filtration process in filter presses, with the objective of increasing overall efficiency, operational sustainability, and long-term reliability. The definition of the objectives has been carried out by considering both the production requirements of the plant and the logistical, environmental, and energy constraints that characterize the installation context. In particular, the plant under investigation operates in geographically remote and high-altitude areas, conditions that amplify the impact of operational inefficiencies and make the optimization of available resources strategically important.

The objectives have been defined and ranked according to a priority scale based on their direct impact on system performance [20], [21].

Increase in plant productivity

The primary objective of this work is the increase in the overall productivity of the filter press. This goal is not limited to a mere reduction of cycle time, but rather encompasses a broader process-oriented perspective, considering the plant's ability to produce a larger quantity of filtered material for the same amount of time, resources, and hardware configuration.

Productivity enhancement can be pursued through the optimization of individual operational phases of the cycle, particularly washing, blowing, and discharge, which significantly affect the overall process duration without directly contributing to solid–liquid separation. A more efficient and adaptive management of these phases makes it possible to reduce non-productive time, improve operational continuity, and increase the number of cycles executed per unit time, while maintaining or improving the quality standards of the final product.

Increase in cake dryness

The second priority objective concerns increasing the degree of dryness of the filter cake discharged from the press. A high dry solids content represents a key quality parameter of the solid phase and directly influences handling, transportation, and storage operations.

A drier cake allows for a reduction in the total volumes to be managed, improves the mechanical stability of the material, and decreases the costs associated with downstream stages of the filtration process. Achieving this objective requires careful optimization of the filtration and blowing phases, as well as a balance between dryness level, process duration, and resource consumption. In this context, the goal is not the indiscriminate maximization of cake dryness, but rather the identification of the optimal trade-off between product quality and operational sustainability.

Reduction of air and water consumption

Another relevant objective is the reduction of compressed air and water consumption, which are essential resources for the cake washing and blowing phases. This aspect is particularly critical given that the plant operates in remote and high-altitude environments, where water supply and compressed air generation are complex and costly.

In such contexts, inefficient resource utilization can lead to significant increases in operating costs and greater exposure to logistical constraints. Reducing consumption requires a more conscious, data-driven management of process parameters, avoiding overly conservative approaches based solely on empirical safety margins. The objective is therefore to make the process more robust, adaptive, and resilient to variable operating conditions.

Reduction of energy consumption

Closely related to the previous objectives is the reduction of the overall energy consumption of the plant. Energy usage is strongly influenced by both the duration of the operational phases and the utilization of high-power auxiliary systems, such as compressors, pumps, and handling equipment.

Improvements in productivity, reductions in non-productive time, and optimization of air and water usage yield a direct benefit in terms of energy consumed per unit of processed material. In an industrial context increasingly oriented toward environmental sustainability and energy footprint reduction, this objective represents a key factor in enhancing the overall competitiveness of the plant.

Enablement of predictive maintenance strategies

An additional objective of this work is the enablement of predictive maintenance strategies, aimed at reducing unplanned downtime and improving long-term system reliability. The analysis of process variables measured during filtration, washing, blowing, and discharge phases enables the identification of early indicators of performance degradation, such as progressive clogging of filter cloths, wear of mechanical components, or inefficiencies in auxiliary systems.

The adoption of a predictive maintenance approach allows maintenance activities to be planned more effectively, reducing both corrective maintenance costs and the impact of downtime on overall productivity. In this sense, the availability of reliable data and their systematic interpretation represent key enablers for the evolution of the system toward more advanced management models.

Reduction of design and configuration time

Finally, this work aims to contribute to the reduction of design and configuration time for the filter press and its associated process parameters. Currently, many operational settings are defined based on prior experience and iterative testing, resulting in extended commissioning times and limited transferability of solutions between different plants.

The development of data-driven analysis and optimization methodologies enables a faster and more effective design and configuration process, reducing the number of required iterations and improving solution repeatability. This objective is particularly relevant from an industrial perspective, as it allows for accelerated plant start-up phases and improved efficiency of the overall engineering process.

4.1.1 Priority Definition and Correlation

Once the macroscopic objectives of the project have been defined, namely those parameters whose improvement yields a direct benefit for the end user or for the plant buyer, it becomes necessary to perform a correlation analysis between these objectives and the machine subsystems.

The purpose of this phase is to identify modelling priorities, thereby defining the operational boundaries of the project and focusing efforts on the subsystems that most strongly influence the required performance. The correlation between global objectives and modelling objectives makes it possible to determine which subsystems have the greatest impact on achieving the targeted performance levels.

Within this framework, the following priority ranking of the global objectives has been defined, from the most to the least relevant:

1. **Increase in plant productivity**
2. **Increase in cake dryness**
3. **Reduction of air and water consumption**
4. **Reduction of energy consumption**
5. **Enablement of predictive maintenance strategies**
6. **Reduction of design and configuration time**

	PRODUCTIVITY INCREASE	DRY MATTER INCREASE	ENERGY SAVINGS	FLUID SAVINGS (AIR/WATER))	PREDICTIVE MAINTENANCE	REDUCTION OF DESIGN TIME
A) FILTRATION PERFORMANCE	X	X				
B) WASHING CONTROL	X	X	X			
C) WASHING AND BLOWING	X			X		
D) COMPLETE HYDRAULIC MODEL						X
E) CLOSING MODEL					X	

Table 1 externa and internal objective correlation

Subsequently, a correlation table was constructed between the identified macro-objectives and the machine sub-processes, each of which exhibits a sufficiently well-defined scope to be modelled individually. The increase in productivity is highlighted as the primary objective, as it represents the issue of greatest industrial relevance and is strongly correlated with the core processes of the machine. The entire research activity is therefore oriented toward maximizing performance in terms of productivity, while remaining within the operational and performance limits of the filter press.

	PRODUCTIVITY INCREASE
A) FILTRATION PERFORMANCE	X
B) WASHING CONTROL	X
C) WASHING AND BLOWING	X

Table 2 productivity focus

Focusing on the objective of increasing productivity, and considering the subsystems that are physically connected to this goal, it is possible to define their degree of influence and to establish the following priority order for analysis:

1. **Control of washing operations**
2. **Filtration performance**
3. **Cake washing and cake blowing**

Analysis of the machine operation highlights that the time associated with the filtration process itself is difficult to optimize, as it is strongly constrained by the characteristics of the processed solid and by the mechanical limits of the machine, in terms of geometry and allowable feeding pressure. The washing and blowing processes have a limited impact on the overall operating time and are instead more relevant to the quality of the final product.

By contrast, the washing phases involve high water consumption and exhibit a wide margin for improvement in terms of both time and resource usage, as they are currently based on constant parameters of a predominantly empirical nature. There is, in fact, no objective evidence justifying the need to stop the machine to perform high-pressure washing, nor a clear quantification of the amount of water actually required for rain washing. A deeper understanding of the cleanliness state of the filter cloths and plate sealing surfaces would make it possible to dynamically adjust the intervention thresholds of these processes, enabling, in a first stage, an improvement in productivity through the reduction of unnecessary machine downtime, and, in a second stage, an optimization of water consumption.

In light of the hypothesis that the greatest impact on productivity, given the limited physical improvement margins of the other processes, is associated with the insufficient knowledge of the machine's fouling state, the subsequent research activity focuses on the identification of a modelling approach capable of estimating a latent parameter representative of the fouling level. This constitutes a paradigmatic case in which a Digital Twin model enables the observability of a condition that is not directly measurable in reality and, on this basis, allows for the implementation of an optimizer. In the specific case under study, this approach enables the control of washing occurrences in order to improve performance with respect to the targeted industrial objective.

4.2 Modelling Strategies

The modelling activities developed in the present work are defined on the basis of the priority analysis framework identified in the previous chapter (Section 5.1). In particular, the selection of modelling methodologies and the level of detail adopted for the individual subsystems are driven by the impact that these subsystems have on the macroscopic objectives of the project, with specific reference to increasing plant productivity and, consequently, to the optimization of washing-related downtime.

From a methodological standpoint, two main modelling approaches can be distinguished:

- **Deterministic, physics-based methods**
- **Numerical or statistical methods**

Methods classified as physical or deterministic are based on the application of well-established theoretical models from the literature to describe the behaviour of the system and its sub-processes. Within this framework, the virtual twin is constructed starting from known physical relationships that correlate process inputs and outputs, while accounting for the constructive and operational characteristics of the machine.

The underlying assumption of this approach is that a sufficiently accurate modelling of the system makes it possible to highlight quantities that are not directly measurable, such as the specific resistance of the filter cloths. This parameter, once properly interpreted and processed through a direct model, can be mapped onto an index representative of the machine's fouling level. Such an index can then be used as a control variable for managing washing cycles, enabling the transition away from strategies based solely on time thresholds or empirical criteria.

Alongside the deterministic approach, a second family of methodologies based on numerical, statistical, or machine learning models is also considered. In this case, the final objective remains unchanged, namely the identification and estimation of latent system variables that are useful for improving operational performance and controlling washing processes. However, unlike physics-based modelling, this approach does not rely on explicit theoretical laws, but rather on mathematical formulations derived from the analysis of historical operational data collected from the plant.

It is important to emphasize that the work carried out within the deterministic modelling framework is not discarded; on the contrary, it is “transferred” into the statistical approach, providing valuable input information to the numerical models. Variables calculated or estimated through physical models can indeed be used as meaningful features, contributing to the enhancement of the descriptive and predictive capabilities of machine learning algorithms.

One of the main critical issues associated with statistical modelling concerns the definition of a reliable output state variable required for model training. In the present case, the objective is to estimate a hidden system variable that is not physically measurable, such as the fouling level of the machine. This results in an intrinsic inability to access a directly observable historical ground truth that could be used as a reference for supervised training.

To overcome this limitation, the proposed approach is based on the use of auxiliary variables and indirect operational information, with the aim of retrospectively identifying the conditions under which washing interventions became necessary in historical operation. The underlying assumption is that washing activation occurred in correspondence with conditions of excessive system fouling. By exploiting this information, it is possible to indirectly label specific regions of the operational domain and use them for training the statistical model.

The trained model is therefore capable of recognizing, within new operational data, the conditions that precede or necessitate a washing intervention. From a prospective standpoint, this capability enables the anticipation of critical states and supports more efficient and adaptive control strategies, with direct benefits in terms of productivity, reduced resource consumption, and minimization of unnecessary machine downtime.

5 Data Acquisition and Conditioning

Data acquisition and subsequent data conditioning represent the first fundamental phase of the modelling activity. The data used in this work are obtained directly from the machine and are provided by the infrastructure operator in their raw form, as extracted from the plant's acquisition and supervision system.

Working with raw data inherently involves the presence of unstructured and redundant information, as well as data affected by measurement errors, temporal discontinuities, and inconsistencies arising both from the plant's operating conditions and from the signal recording procedures. For this reason, a preliminary phase of data cleaning, organization, and cataloguing is required, with the aim of making the dataset suitable for subsequent analysis and modelling activities, both deterministic and statistical in nature.

In particular, the data conditioning activity is intended to ensure the temporal consistency of the acquired variables, improve the informational quality of the signals, and reduce the impact of noise and anomalies that could compromise the reliability of the developed models. This phase is crucial, as the quality of the input data directly affects the robustness and accuracy of the results obtained.

The main operations performed during the data cleaning and cataloguing phase are listed below:

1. **Label translation**
2. **Index correction and spike removal**
3. **Cycle identification and alignment**
4. **Generation of data packets**

5.1 Label Translation

For plant setup–related reasons, the research activity was in fact conducted partly in parallel with the construction of the plant and of the data exchange infrastructure. As a result, the datasets used in this work originate from different sources and from different acquisition systems. This circumstance created the need to develop a framework capable of temporally aligning the data and recognizing equivalent variables across heterogeneous datasets.

The data provided by the machine cover the period from its commissioning up to the completion of the transmission and data acquisition infrastructure.

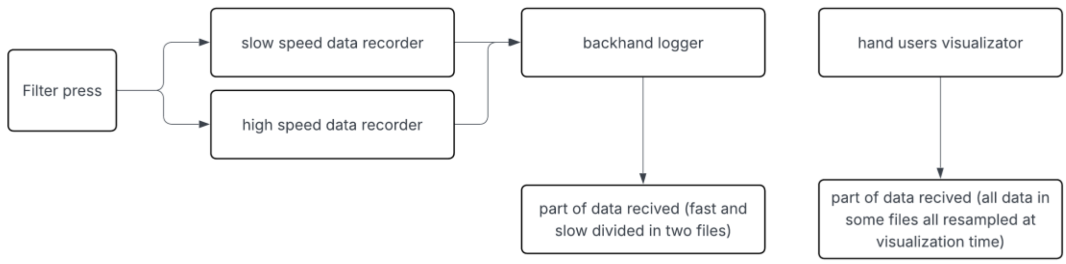

Figure 14 Data acquiring scheme

The first issue addressed concerns the reassembly of data originating from low-frequency and high-frequency acquisition systems (Figure 14). Indeed, the collected data are divided into two categories and are therefore stored in two separate databases: data acquired every second and data acquired every thirty seconds. This distinction in sampling frequency is driven by the dynamics of the observed phenomena; for instance, certain variables such as the level of very large tanks do not require high sampling rates.

At this stage, the two databases are merged, and the lower-frequency dataset is resampled, by means of linear interpolation, onto the timeline of the higher-frequency dataset. This operation yields a coherent dataset that includes the entire set of information available from the machine. On the other hand, the resulting dataset is larger in dimensionality and inevitably more demanding in terms of storage and handling. However, this condition is still negligible with respect to the final dataset size and the overall computational requirements. It can be observed that only approximately 5% of the variables require resampling.

A more complex issue to address in the acquisition process concerns the alignment of successive data packets based on variable labels. During the evolution of the infrastructure, the provided data packets were extracted from different system layers, leading to discontinuities in variable nomenclature. Clearly, nomenclature inconsistencies also arise from the lack of full visibility into the labelling logic adopted by the various systems (Figure 15). To address this issue, a dedicated framework was developed to consistently map variable names back to a predefined reference nomenclature established at the origin.

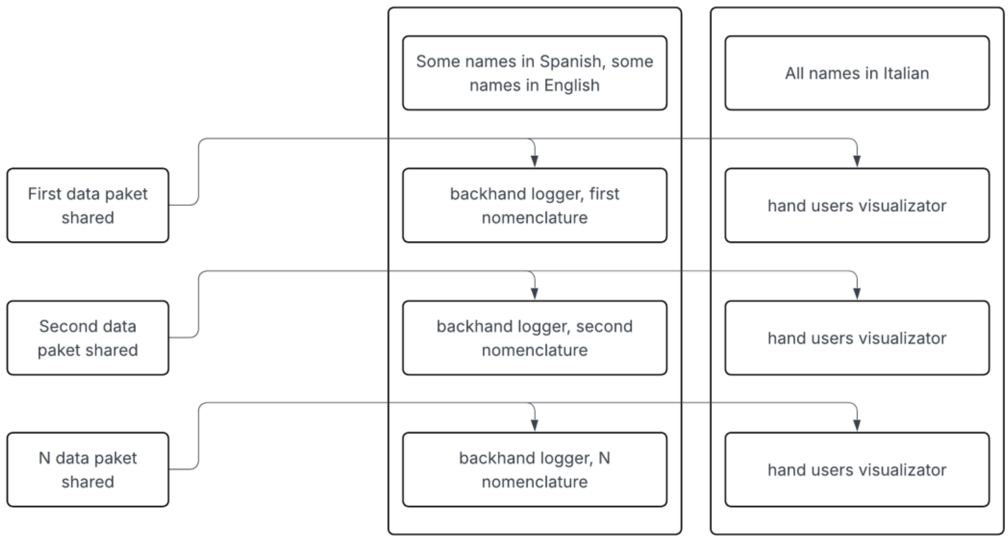

Figure 15 variable nomenclature origin scheme

Unfortunately, no temporal overlap is available between the data packets, making a direct numerical comparison of the variables impossible.

The first issue to be addressed was the unification of the language used in the variable labels. Depending on the system level from which the data were extracted, labels were expressed in Spanish, English, or Italian. This situation arises because the machine acquires data using Spanish labels, the management system stores them using English labels, and the processing and visualization system saves them using Italian labels.

To resolve this issue, the Google translation system was employed. A dedicated algorithm samples all the words contained in the labels and sends them to the Google Translation API, which automatically detects the source language and provides the corresponding translation into English. Clearly, if a label is already recognized as English, no translation is performed.

At this point, each data packet added to the timeline contains labels expressed in a single common language. However, this does not guarantee semantic consistency across datasets, as different translations of the same variable may not be identical, or nomenclature variations may exist between packets even within the same language. To address this issue, once a reference nomenclature had been defined, a comparison algorithm was developed.

This algorithm compares each variable label against all reference labels and assigns a coherence score based on their similarity (Figure 16). After evaluating all labels, each variable is associated with the reference label that exhibits the highest correlation, and a translation table is generated accordingly. The table is saved in CSV format to allow for manual inspection and validation in case of issues. When the correlation index falls below a predefined threshold, an alert is generated and manual association is required.

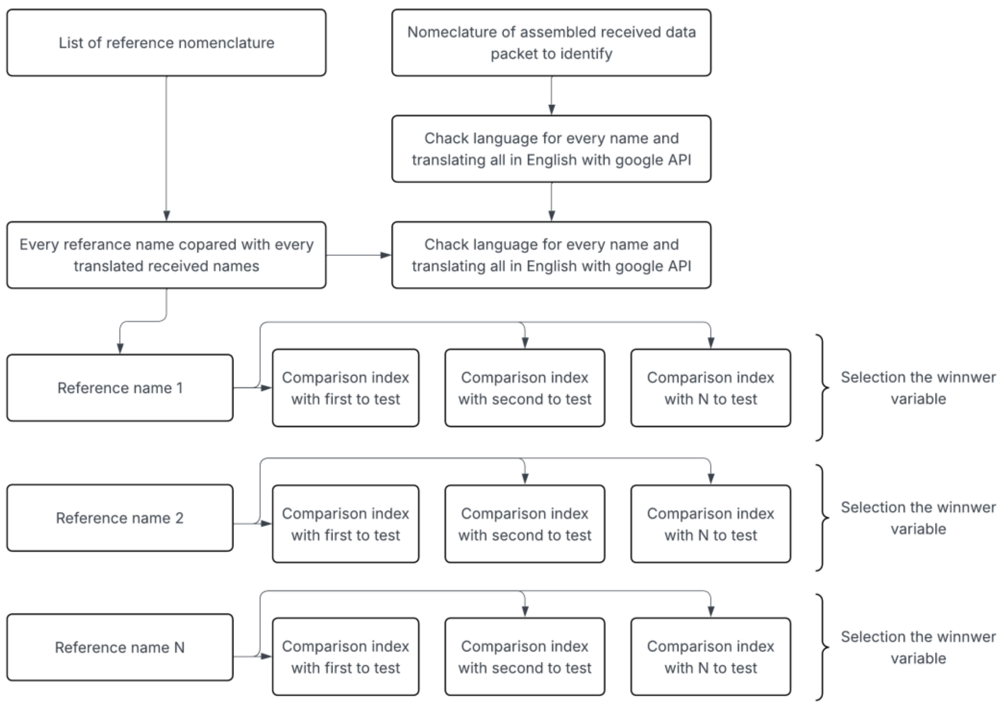

Figure 16 translating a comparing algorithm

The correlation index is computed based on the number of words that are considered synonyms, normalized by the average number of words contained in the two variable labels. Terms shorter than three characters are excluded from the analysis. Two words are considered synonyms if they are identical or if they are classified as synonyms according to a manually constructed vocabulary, derived from recurring occurrences or alerts generated during the translation process.

The correlation index is defined as follows:

$$\text{CorIDX}(A, B) = \frac{\text{Number of synonyms}}{\text{mean}(\text{words in } A, \text{words in } B)}$$

The word count in labels A and B excludes all terms shorter than three characters, in order to remove articles, prepositions, or other non-informative tokens.

Once a translation table is available for each data packet, all datasets can be aligned to a common reference nomenclature and subsequently concatenated sequentially along the time axis.

5.2 Index and Spike Correction

The quality of the acquired data plays a fundamental role in ensuring the reliability of subsequent analyses and the correct functioning of data processing algorithms. In real-world contexts, however, signals may exhibit anomalies due to acquisition errors, communication issues, or unforeseen operating conditions. Among the most frequently observed problems are data gaps, sudden jumps (spikes), and inconsistencies in the progression indices of process variables [22].

To address these issues, specific correction and validation algorithms were implemented.

The first cleaning algorithm focuses on the detection and correction of data gaps. Data gaps manifest as missing time segments or invalid samples within an otherwise continuous signal. The developed algorithms analyse the temporal coherence of the time series, identifying discontinuities in time progression or null/undefined values where valid measurements are expected.
Once a gap is identified, the missing data are reconstructed using linear interpolation between the last valid value preceding the gap and the first valid value following it. This approach preserves the overall trend of the signal, avoids the introduction of artificial discontinuities, and maintains low computational complexity.

The second algorithm compensates for anomalous jumps in index variables. A particularly critical case concerns progressive index variables, namely quantities that are expected to increase monotonically (e.g., sample counters, cycle indices, or phase identifiers). In the presence of system or communication errors, these variables may exhibit sudden resets to zero or unjustified decrements.
The correction algorithms compare each new sample with the previous value, verifying the monotonicity of the index. When a reset to zero or a jump incompatible with the expected progression logic is detected, the anomalous value is discarded and replaced with a coherent value computed according to the expected sequence. In this way, error propagation into subsequent processing stages is prevented.

The third algorithm is dedicated to recognizing the actual start of operational phases. Beyond numerical data correction, algorithms were also implemented for logical validation of the process state, with particular attention to identifying the effective initiation of an operational phase. It is possible for the signal or variable indicating the start of a phase to be triggered without the phase being actually executed. This condition may occur, for example, in the case of premature process abortion or synchronization errors.

The system is able to detect such situations by analysing the evolution of physical variables associated with the phase (e.g., temperature, pressure, speed, or current). If, following a phase-start signal, the physical variables remain substantially constant or do not exhibit the characteristic variations associated with that phase, the algorithm concludes that the phase was not effectively executed. In this case, the event is marked as invalid and can be excluded from subsequent analyses or corrected at the logical state level.

A first practical example of this approach concerns the recognition of washing phases that were actually performed. It was frequently observed that a washing phase could be logically triggered but, for various reasons, not physically executed. In this case, a dedicated routine verifies, through the analysis of the electrical power absorbed by the washing robot, whether the expected descent motions were effectively carried out (Figure 17).

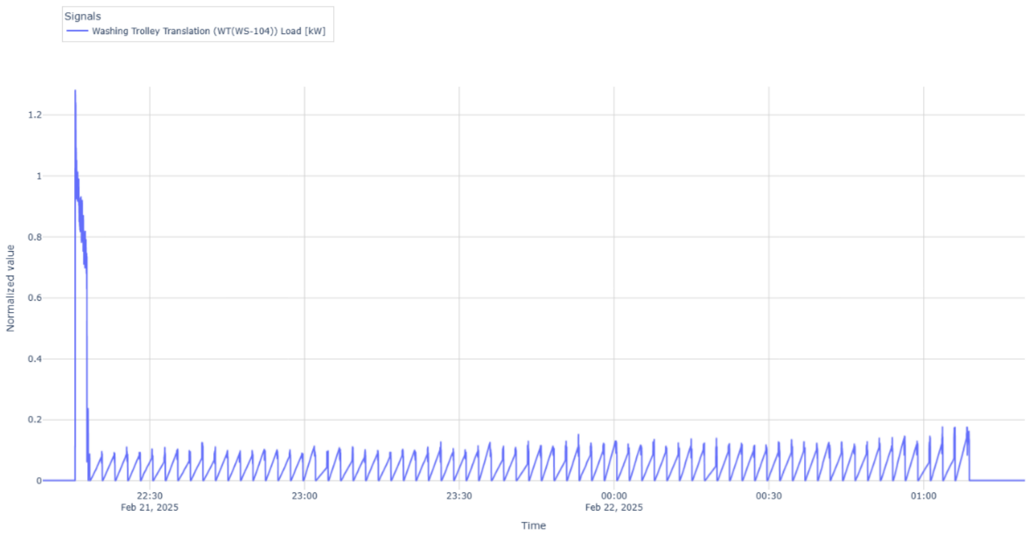

Figure 17 Washing system trolley electric power

From the previous plot, it is possible to clearly identify the number of descent and ascent movements performed and to count them; this number must correspond to the number of filtration chambers that have been washed.

5.3 Cycle Identification and Alignment

In this phase, the temporal position of the cycles is determined and the cycles themselves are extracted from the time history of the process. This step represents the first fundamental stage in the construction of a historical representation of the machine in the cycle domain. The objective is to progressively shift the analysis from the continuous time domain to the discrete cycle domain, transferring all process parameters deemed significant.

A correct mapping between the time domain and the cycle domain allows for a drastic reduction in dataset dimensionality, while preserving the majority of the information that is relevant for analysis and modelling purposes (Figure 18).

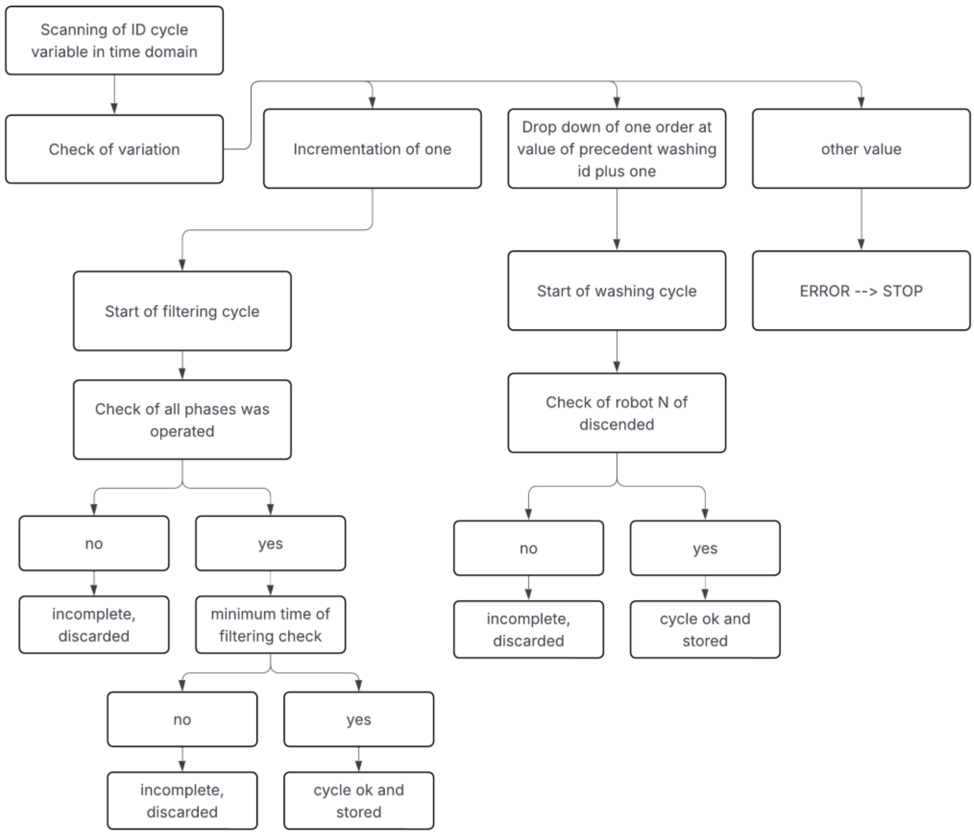

Figure 1817 Cycle identification algorithm

The cycle detection algorithm operates by scanning the time domain and verifying, at each time instant, the occurrence of the conditions required to trigger a new cycle. The first step consists of scanning the time series in search of variations in the cycle index. This index is a key process variable that must mandatorily be incremented by one at the beginning of each new cycle. Moreover, the index is a "self-descriptive" variable, as it allows discrimination between cycle types (filtration or washing) by enforcing a separation of at least one order of magnitude between the two sets of values [23].

Specifically, if the new index value is increased by a single unit with respect to the previous one, the start of a new filtration cycle is identified. Conversely, if the index undergoes a significant decrease (of at least one order of magnitude), while still remaining greater than one with respect to the last valid washing index, the event is interpreted as the beginning of a new washing cycle. Excluding possible erroneous triggers, which are intercepted and removed through the signal cleaning and conditioning algorithms, this methodology proves sufficiently robust for identifying the cycle start instant.

The subsequent steps of the algorithm are dedicated to verifying cycle completeness and identifying the corresponding end instant. For filtration cycles, these operations are performed through an analysis of the operational phases: for a cycle to be considered valid, at least all the fundamental phases prescribed by the process must have been executed, and the overall cycle duration must exceed a minimum threshold, experimentally determined and associated with the machine filling phase. If the cycle satisfies these requirements, the end of the opening phase is identified and the main cycle information, such as a unique identifier, start date and time, and end date and time, is stored in a dedicated data structure (e.g., a dictionary).

Once cycles have been identified and validated, a temporal shifting of the values related to material discharge becomes necessary. In particular, the mass and moisture content of the solid phase are not measured within the cycle to which they belong, but rather in the subsequent cycle. This misalignment is due to the placement of the sensors along the downstream conveyor belts, which detect the discharged material only after the machine has already started a new filtration cycle. To compensate for this effect, the mass and moisture discharge data are shifted backward by one cycle, thus correctly realigning them with the cycle that actually produced the material.

It is important to note that, following this transformation, these quantities are no longer suitable for use in time-dependent deterministic models. They nevertheless retain full validity as cycle-level aggregated variables and can therefore be employed for statistical, diagnostic, or data-driven analyses in the cycle domain.

With regard to washing cycles, the same verification algorithm described above is applied. If the cycle satisfies the validation criteria, it is annotated in the same manner as filtration cycles, recording the cycle identifier and the corresponding start and end times. In this way, a coherent and complete representation of the machine's operational history in the cycle domain is obtained, which constitutes the foundation for the subsequent phases of process analysis and modelling.

At this stage, the temporal position of the cycles is determined and they are extracted from the time history of the process. This step represents the first essential operation for constructing a historical representation of the machine in the cycle domain. The underlying idea is to progressively transfer the analysis into the cycle domain by carrying over all significant process parameters. An accurate mapping between the time domain and the cycle domain enables analyses to be performed on reduced datasets with minimal loss of relevant information.

5.4 Data Package Generation

Once the temporal positioning of all cycles has been uniquely identified, the system proceeds with a structured data storage phase based on data packets. This operation represents a key step in the overall architecture, as it enables the organization of information in a manner consistent with the cyclic nature of the analysed process.

In particular, each identified cycle corresponds to a single dedicated file, which is stored independently of the others. Each file is indexed using the cycle start date and time, thus ensuring a clear, unique, and easily traceable temporal reference. This time-based indexing is fundamental both for manual data inspection and for automated processing or cross-referencing with other information sources.

Within each file, all relevant information associated with the cycle is stored, including both:

- **raw temporal data**, namely the complete time series of the physical quantities acquired throughout the entire duration of the cycle;
- **aggregated parameters**, which provide a concise summary of the cycle behaviour and are computed and used by the analysis algorithms described in subsequent chapters.

This dual storage strategy makes it possible to preserve the information in its original form while simultaneously providing a compact and readily usable representation for statistical analysis, modelling, or machine learning techniques.

The adopted architecture therefore enables cycle-level data packaging, making storage space management extremely flexible. Files can be easily moved, copied, archived, or transferred without loss of logical coherence, as each packet contains all the information required for its interpretation.

An additional advantage of this structure is the ability to operate directly on individual cycles without the need to reprocess the entire dataset. Having separate files for each cycle allows targeted interventions on single dataset elements, such as modification, updating, or complete exclusion [24].

This characteristic is particularly useful during exploratory data analysis and data validation phases. Indeed, more in-depth analyses may reveal the need for localized corrections, for example due to anomalies not detected during the initial processing stage, acquisition errors, or non-representative operating conditions. In such cases, interventions can be limited exclusively to the affected files, avoiding the need to regenerate the entire data archive.

In conclusion, cycle-level packet-based storage constitutes an architectural choice that significantly improves system maintainability, scalability, and robustness, facilitating both subsequent automated analyses and manual review and correction operations.

6 Deterministic Physical-Based Methods

This chapter is devoted to the description and application of deterministic, physics-based methods for the analysis and interpretation of the machine's behaviour. These approaches are grounded in the explicit modelling of the physical phenomena governing the process, through mathematical relationships derived from conservation principles, constitutive laws, and engineering knowledge of the system. The primary objective is to provide an interpretable and coherent representation of the machine operation, capable of directly linking measured quantities to the underlying physical dynamics.

Deterministic physics-based models offer the advantage of high interpretability and enable causal analysis of process variables. Unlike purely data-driven approaches, they allow the incorporation of physical constraints, operational limits, and *a priori* knowledge, making them particularly effective in industrial contexts where data availability is limited or affected by noise and temporal misalignments. Moreover, such models constitute a fundamental tool for anomaly diagnosis, as they enable comparison between the actual machine behaviour and the expected behaviour derived from known physical laws.

In the specific context of this work, deterministic methods are applied in the cycle domain, leveraging the reconstructed operational history described in the previous chapters. This choice allows a reduction in problem complexity and enables the analysis to focus on physically meaningful indicators associated with each operating cycle. The developed models aim to describe the relationships among operating parameters, machine states, and process outputs, with particular attention to the effects of loading conditions, operational phases, and environmental variables.

6.1 Aggregated Filtration Parameters

The research aimed at obtaining aggregated parameters for the characterization of the filtration process was structured with the objective of indirectly estimating the state of the filter cloths, which represent a key element for the correct operation of the plant. Knowledge of the cloth condition is in fact essential not only for monitoring filtration efficiency, but also for planning and optimizing washing phases, which have a significant impact on operational continuity, energy consumption, and the service life of the filtering medium. From this perspective, the analysis seeks to identify synthetic and easily interpretable quantities that can serve as indicators of the system state under different operating conditions.

The adopted approach is initially based on a reference analytical model commonly used in filtration theory, which describes the relationship between process variables and the hydraulic resistances present in the system. In particular, the model considers the following fundamental parameters [25], [26]:

$$\frac{\mathrm{d}t}{\mathrm{d}V} = \alpha \mu c \frac{V}{A^2 \Delta p} + \frac{\mu R}{A \Delta p}$$

- **α**: specific resistance of the deposited solid phase (cake),
- **μ**: dynamic viscosity of the filtered fluid,
- **R**: hydraulic resistance of the filtration membrane,
- **c**: ratio between the amount of deposited solid and the volume of filtrate produced.

Within this formulation, the parameters of greatest interest are the specific resistance of the solid, **α**, and the membrane resistance, **R**, as they are directly correlated with the cleanliness state and fouling degree of the filter cloths. When properly estimated, these quantities provide relevant information on the progressive degradation of the filtration medium and on the temporal evolution of the process, making them particularly suitable for monitoring and control applications.

Although the theoretical modelling framework allows, at least in principle, the derivation of accurate and physically meaningful results, its direct application in an industrial context presents several critical challenges. The main difficulties arise from the complexity of the real system and from the limited accessibility of physical quantities within the hydraulic zone where filtration actually occurs. Unlike laboratory-scale tests, which are characterized by controlled and reproducible conditions, industrial filtration processes are strongly affected by non-ideal operating variables and transient phenomena, particularly during the start-up and shutdown phases of the filtration cycle.

- In this context, obtaining reliable measurements of several key quantities becomes particularly challenging, including:
- the pressure drop **Δp** in the vicinity of the filtration medium, which is often not directly measurable or is affected by parasitic contributions due to distributed pressure losses,
- the accurate determination of the parameter **c**, related to the effective amount of deposited solid, which is difficult to estimate in real time in an industrial plant,
- the identification, within the filtration cycle, of an operating condition that can be assimilated to the steady-state regime typically assumed in theoretical models and laboratory tests.

In light of these limitations, the direct application of the complete theoretical model proves to be poorly robust and highly sensitive to measurement uncertainties. For this reason, in order to retain the theoretical reference framework while making it more consistent with the actual operating conditions of the machinery, an alternative approach was adopted based on the separate evaluation of the parameters of interest under different operating conditions of the system.

Specifically, the hydraulic resistance of the membrane, **R**, is estimated during the filling phase operating at constant flow rate, a condition under which the contribution of the cake is negligible or at least limited. Conversely, the specific resistance of the solid, **α**, is evaluated during the constant-pressure filtration phase, in which the temporal evolution of the flow rate is predominantly governed by the accumulation of solids on the filter cloth. This strategy allows the contributions of the two main resistances to be isolated under conditions in which each of them is dominant, thereby reducing the impact of experimental uncertainties and improving the representativeness of the estimated parameters with respect to the actual behaviour of the industrial system.

6.2 Pressure Drop Evaluation Δp

To overcome this limitation, a one-dimensional calculation strategy was adopted with the aim of reconstructing the correct pressure drop acting on the filtration cake. Specifically, a 1D model was developed in the Modelica environment, with the objective of accurately mapping the pressure losses associated with the feed and discharge lines of the machine under different operating conditions (Figure 19). Within this modelling framework, the filter press plate pack is represented as a single lumped chamber, characterized by an equivalent mean pressure [27].

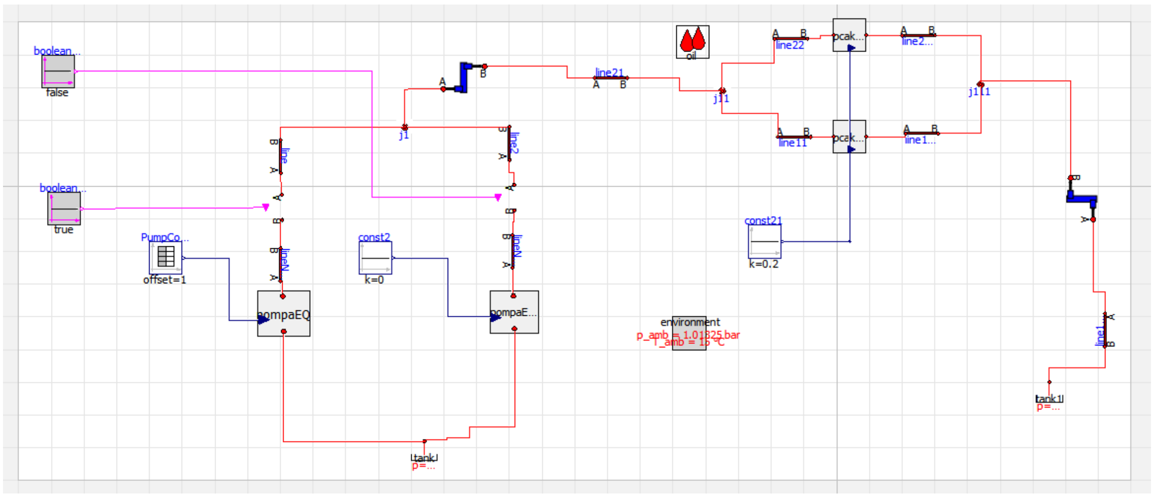

Figure 19 1D Hydraulic filter model

This simplification entails the loss of information regarding the pressure gradient within the filter; however, this aspect is not critical for the purposes of the present analysis. The objective of the model is not a detailed description of the local phenomena occurring inside the plate pack, but rather the isolation of the global pressure loss contribution attributable to the filtration cake. In this context, the lumped-parameter chamber assumption represents an acceptable compromise between accuracy and computational simplicity.

The feed and discharge lines were modelled according to the scheme shown in the reference figure, including both distributed pressure losses along the pipelines and localized losses due to bends, valves, and contractions. Based on this model, flushing simulations were performed under operating conditions representative of the actual plant operation, in order to reproduce the hydraulic behaviour of the system in the absence of the filtration cake.

Subsequently, the results obtained from the simulations were used to fit the characteristics of the two lines by means of a third-order polynomial model as a function of the flow rate. This approach made it possible to identify a set of coefficients representative of the hydraulic behaviour of the plant, to be employed as equivalent parameters within the overall physical modelling of the filtration process.

Once the intrinsic pressure losses of the plant have been characterized, it becomes possible to remove their influence from the experimental measurements of the overall pressure. In this way, the contribution of the feed and discharge lines is subtracted from the measured pressure drop, allowing the isolation of the sole Δp effectively acting on the filtration cake. This procedure represents a fundamental step for the correct estimation of the characteristic process parameters and ensures greater consistency between the theoretical model and the actual operating conditions of the industrial plant.

6.3 Calculation of Parameter c (Concentration)

The concentration of the solid phase in the feed line represents a parameter of primary importance for the analyses carried out on the filtration process, as it directly affects both the separation efficiency and the quality of the discharged solid product. Its accurate determination is therefore essential to ensure the reliability of experimental evaluations and comparisons among different operating conditions [28].

The assessment of the solid concentration is performed through analyses based on the densities of the different phases involved. In particular, once the density of clean water (ρ_acq) is known, the density of the feed fluid is experimentally measured (ρ_mis), and the density of the dry solid phase is estimated (ρ_sol), it is possible to derive the dimensionless parameter **c**, defined as the Solid/Filtrate ratio, through the following relation:

$$c = \frac{\rho_{mis} - \rho_{acq}}{\rho_{sol} - \rho_{mis}}$$

This formulation makes it possible to indirectly estimate the concentration of the solid phase from relatively simple density measurements, avoiding direct sampling of suspended solids, which would be more complex and less repeatable.

The density of the "dry" solid (ρ_sol) is not measured directly, but is instead estimated based on the weighing of the material discharged from the filtration machine. The obtained value is subsequently corrected by accounting for the residual moisture content, which is evaluated using a hygrometer. This correction is required because the discharged solid is not completely free of water, and its apparent mass includes a fraction of entrapped liquid.

In order to verify the consistency of the calculated parameter **c** and to assess the reliability of the assumed dry solid density, a cross-check can be performed by re-evaluating the solid density through a mass and volume balance. Specifically, the dry solid density can be recalculated by considering the ratio between the mass of the discharge, corrected for the contribution of entrapped water, and the effective volume of the solid component retained in the filter:

$$\rho_{sol} = \frac{M_{pes} - (V_{max} - V_{filt})\,\rho_{acq}}{V_{filt}\,c}$$

where:

- M_{pes}represents the total measured mass of the discharge,
- $V_{max}$is the maximum volume occupied by the cake,
- $V_{filt}$is the volume attributable to the solid component only,
- $(V_{max} - V_{filt})$represents the volume of water entrapped within the cake.

The comparison between the initially estimated solid density and the value recalculated using this relation allows the determination of an average density error. For the month of October 2024, this average error was found to be equal to 2.4%. This deviation is considered acceptable within the context of the analyses performed, although it is plausibly influenced to a significant extent by the measurement error associated with the hygrometer, which represents one of the most critical elements of the entire estimation procedure.

The cake density, defined as the density of the solid product effectively discharged from the filtration machine, is calculated as the ratio between the maximum volume occupied by the cake and the measured mass of the discharge:

$$\rho_{cka} = \frac{V_{max}}{M_{pes}}$$

This parameter provides a global measure of the compactness of the discharged material and implicitly includes both the solid fraction and the residual water entrapped within the cake structure. The comparison between the cake density and the dry solid density is particularly useful for assessing the water content at discharge.

In particular, a high ratio between the cake density and the dry solid density indicates a greater presence of residual water, whereas values closer to the dry density are indicative of a more dewatered cake and, therefore, of higher filtration process efficiency. The combined analysis of these parameters thus enables a quantitative evaluation of machine performance, providing useful insights both for the optimization of operating conditions and for the comparison among different process configurations.

6.4 Evaluation of Non-Directly Measurable Parameters

In the context of industrial filter presses, the determination of certain characteristic process parameters is complex, as they are not directly measurable during normal plant operation. Among these parameters are the cake resistivity **α** and the filter cloth resistance **R**, which are fundamental both for the correct interpretation of electrical measurements and for the characterization of the overall behaviour of the filtration system.

The estimation of these parameters requires the adoption of simplifying assumptions that allow the real process to be reduced to conditions comparable to those assumed in theoretical models and laboratory tests. In particular, it is necessary to identify, within the filtration cycle, phases characterized by constant flow rate and constant pressure and, within these phases, an operating condition that can be assimilated to a steady-state regime, in which the main process variables can be considered constant or slowly varying over time. Only under such conditions is it possible to apply established calculation relationships and obtain representative values of the parameters of interest.

During this phase of the cycle, the dynamic contribution associated with the actual spatial distribution of the cake is neglected, and transient variations of the plant are discarded, allowing the system to be modelled using equivalent quantities. This makes it possible to separate the contribution of the filter cloth from that of the filtered material and to proceed with the calculation of the resistivity **α** and the resistance **R** based on globally measured electrical quantities.

The adopted approach thus makes it possible to overcome the impossibility of directly measuring the parameters under consideration, providing a methodology that is consistent with the assumptions of theoretical models and sufficiently robust for application in an industrial environment.

6.4.1 Constant Flow Rate Condition – Analysis of R

The operating condition at maximum flow rate was identified as particularly suitable for the determination of the state parameter of the filter cloths, represented by the equivalent hydraulic resistance **R**. Under this operating condition, the filter is assumed to be empty at the initial time and to remain in the filling phase throughout the entire analysed time interval. This choice makes it possible to simplify the system behaviour, reducing the influence of complex transient phenomena and facilitating the interpretation of the process data.

During this phase of the filtration cycle, it is assumed that the pressure drop acting on the system is determined exclusively by the piezometric pressure, which is a direct function of the filter filling level.

$$p = \rho g h$$

$$p_{mean} = \frac{1}{h}\int_0^{h_{max}} \rho g h \, dh$$
$$p_{mean} = \frac{\rho g h_{max}}{2}$$
$$h_{max} = \frac{V_{in}\, c - V_{out}}{N_t\, w\, b}$$

where:

- $h_{max}$is the maximum height of the free water surface with respect to the solid,
- $V_{in}$is the volume of slurry entering the filter,
- $V_{out}$is the volume of water exiting the filter,
- $N_t$is the number of filtration chambers,
- $w$is the chamber width,
- $b$is the chamber thickness.

Other contributions to pressure loss, such as those associated with cake formation or with localized hydraulic resistances of the plant, are therefore neglected, as they are secondary with respect to the objective of the analysis. Based on these assumptions, a simplified reference model was defined, schematically represented in Figure 20.

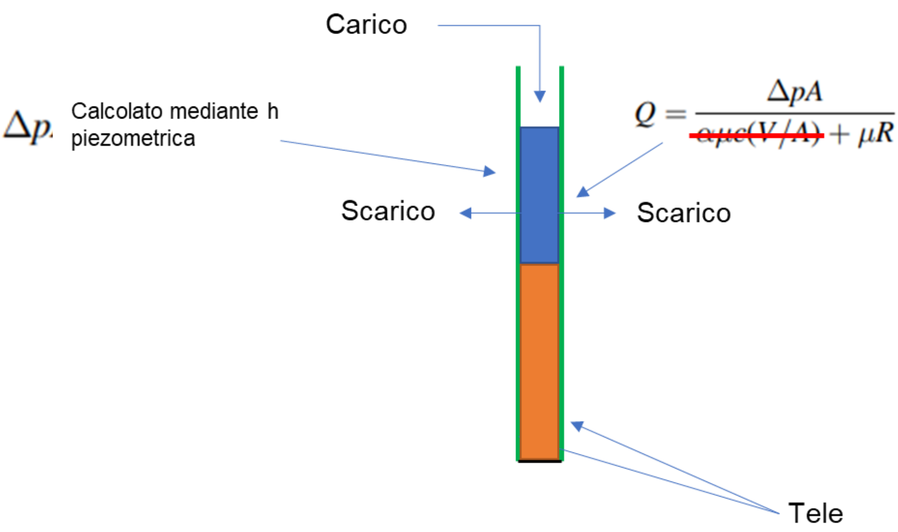

Figure 2018 Filter filling model scheme

$$A = \frac{V_{in} * c - V_{out}}{w} * 2$$

The flow area calculated in this manner makes it possible to isolate only the fraction of the filter cloths that effectively contributes to the hydraulic resistance.

The adopted model is considered valid under compliance with the following fundamental assumptions:

- stratified and impermeable solid phase,

- empty discharge line,
- presence of air in the feed line (empty filter during the filling phase).

The first assumption states that, during the initial phase of the cycle, the solid phase is predominantly deposited at the bottom of the filter, forming a compact layer that offers a hydraulic resistance toward the discharge significantly higher than that of the filter cloth alone. Under these conditions, the solid can be considered impermeable with respect to the "clean" cloth, allowing the observed flow to be attributed exclusively to passage through the cloth itself. This simplification is reasonable during the initial filling phase, when the contribution of the cake to the flow is negligible or, in any case, stable.

The second assumption considers that the discharge line does not operate at maximum flow rate due to the presence of air within it. Consequently, the hydraulic resistance associated with the discharge line can be regarded as negligible compared to that introduced by the filter cloth. This assumption allows distributed and localized pressure losses along the discharge line to be excluded from the model, focusing the analysis solely on the contribution of the cloth.

The third assumption requires that the filter is initially empty and that the analysed process takes place entirely during the filling phase. The presence of air in the feed line ensures that the fluid level increases progressively over time, making the piezometric pressure a well-defined function of the filling level. This allows the measured quantities to be directly correlated with the hydraulic behaviour of the filter cloth, ensuring consistency between the model and the actual operating conditions of the filter press.

Taken together, these assumptions make it possible to obtain a model that is sufficiently simple yet representative, suitable for estimating the hydraulic resistance **R** from globally measured quantities and consistent with the objective of characterizing the condition of the filter cloths under industrial operating conditions.

Following the procedure described in the previous sections, it is therefore possible to reconstruct the temporal evolution of the apparent hydraulic resistance of the filter cloth during the filter filling phase. This reconstruction enables the analysis of the cloth behaviour throughout the filtration cycle and the identification of any variations attributable to the operational state of the system. By way of example, Figure 21 and 22 shows the trend of the apparent resistance **R** for a single filtration cycle.

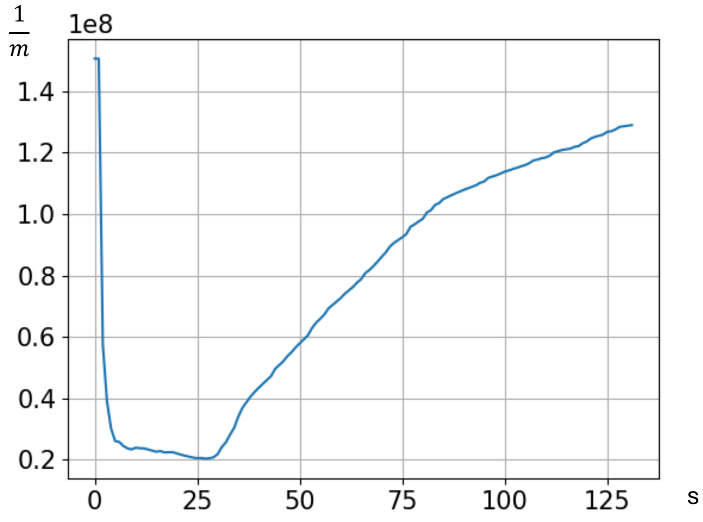

Figure 21 Hydraulic apparent resistance during filling

The trend of **R** obtained through this extraction procedure exhibits good repeatability across consecutive cycles, highlighting an overall stable behaviour under the considered operating conditions. Figure 3 shows the superposition of multiple filtration cycles, from which it can be observed that the apparent resistance maintains a fairly constant trend, thus confirming both the validity of the adopted assumptions and the robustness of the parameter extraction method.

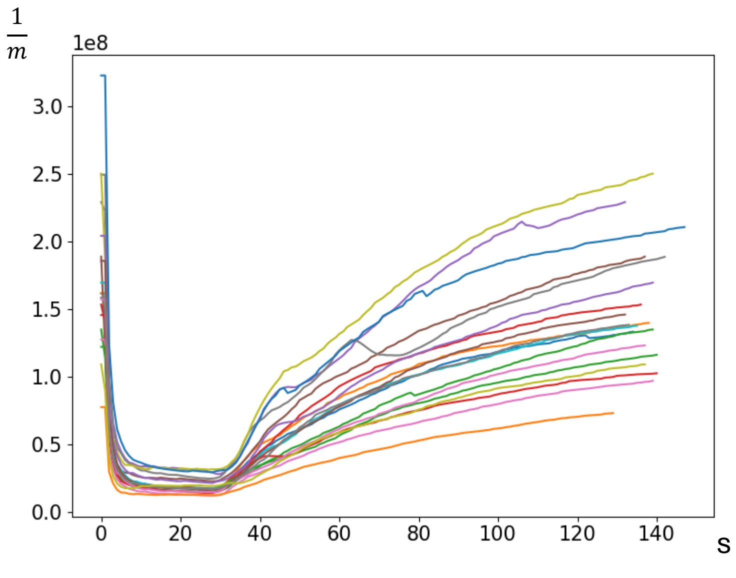

Figure 22 Hydraulic apparent resistance during filling for more cycles overlapped

By taking the first reported curve as a reference, three distinct operating regions can be identified, each characterized by different operating conditions and by a different level of data reliability. This subdivision is illustrated in Figure 23.

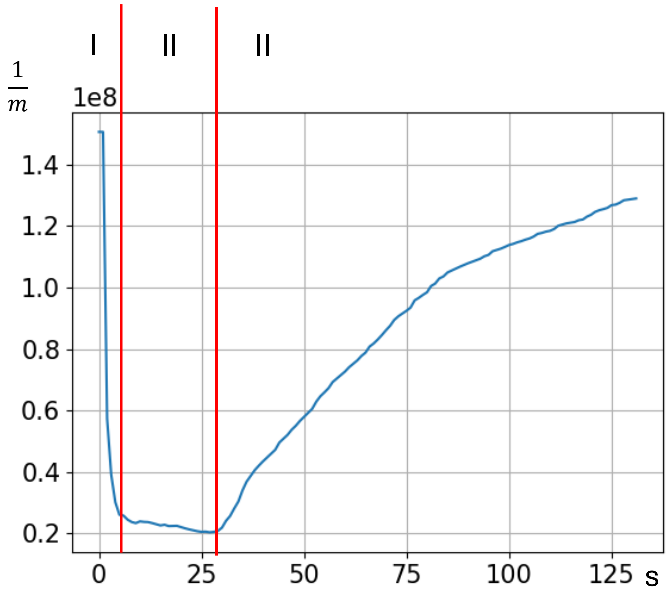

Figure 23 Working region individuation

The first operating region includes the start-up phase of the cycle and is strongly influenced by system inertia. During this phase, the measured filtrate flow rate appears to still be low, while the piezometric pressure difference begins to increase significantly. This discrepancy can be attributed to measurement transients and to the initial dynamics of the plant, which render the data poorly representative of the actual filtration conditions. For this reason, this region is excluded from the evaluation of the resistance R.

The second operating region is characterized by operating conditions that are as close as possible to a steady-state regime. In this phase, the main process variables are relatively stable and consistent with the assumptions of the adopted model. Consequently, this region represents the interval of greatest interest for the required evaluations and is used for the determination of a representative value of the filter cloth resistance.

The third operating region is instead likely affected by the loss of the impermeability assumption of the solid phase. As the solid component inside the filter increases, the apparent filtration area tends to decrease and, consequently, the fraction of fluid drained through the cake increases. This phenomenon alters the relative contribution of the filter cloth, making the assumption of flow occurring exclusively through the cloth no longer valid. This region is therefore also excluded from subsequent evaluations.

Once the region of interest (Figure 24) has been identified, it becomes necessary to adopt a data aggregation strategy related to the second operating region. In a first step, the value of R was determined by computing the average of the values within the identified interval. Although simple, this approach makes it possible to obtain a representative value for the cycle and is adequate for the objectives of the preliminary analysis. The possibility of adopting alternative aggregation strategies in the future remains open, depending on further analytical requirements or on a higher level of model refinement.

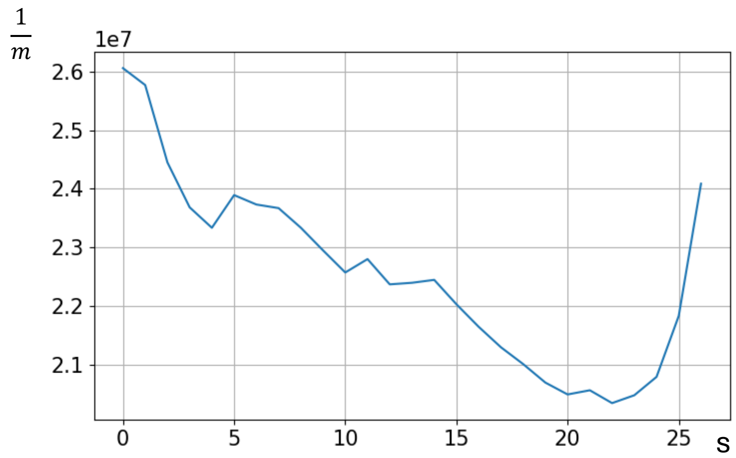

Figure 24 Zoom of interest region of example curve

Once a resistance value has been identified for each cycle, its temporal evolution over the entire observation period can be analysed. Figure 25 shows the historical trend of the parameter **R**, obtained from the filtration cycles considered to be valid.

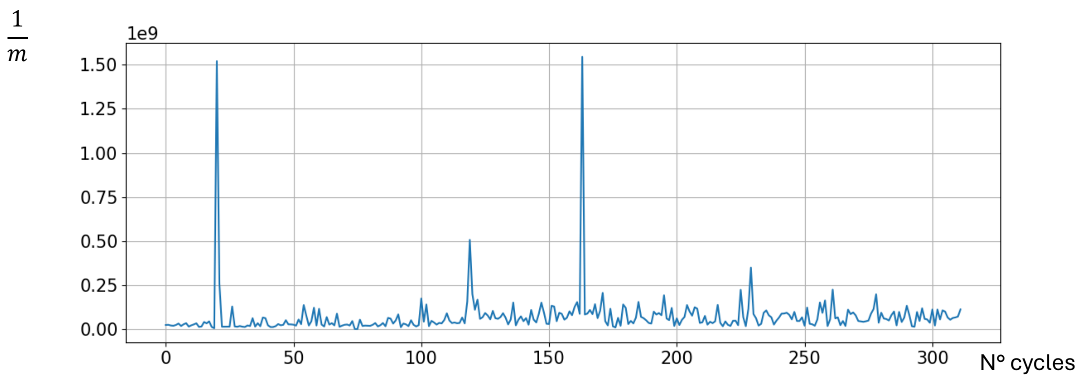

Figure 25 Computed resistivity for all cycles analysed

Despite the application of data cleaning procedures, some cycles still exhibit the presence of anomalous values (spikes), which are likely attributable to measurement disturbances or to operating conditions that are not fully consistent with the assumptions of the model. In parallel with the pointwise analysis, a global assessment of the number of effectively usable cycles is therefore also carried out, in order to ensure the reliability of the conclusions.

For improved readability of the overall trend, Figure 26 finally reports the same plot with an amplified scale, from which the spikes have been excluded.

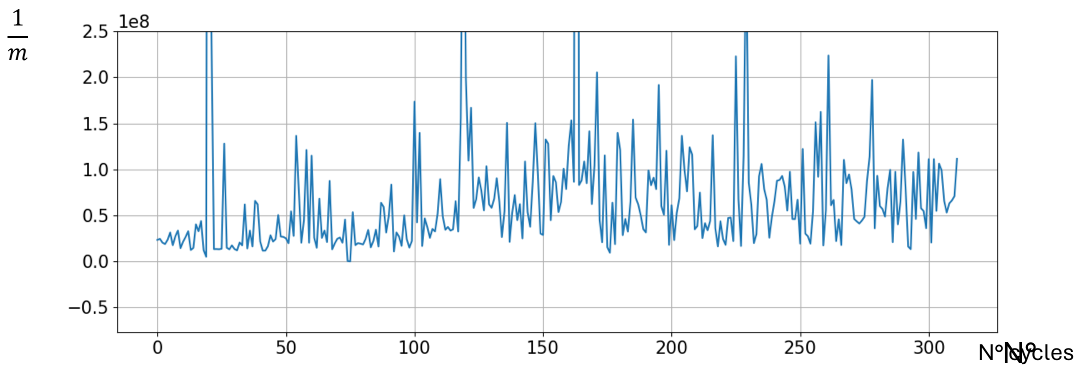

Figure 26 Computed resistivity for all cycles analysed zoomed

6.4.2 Constant Pressure Condition – Analysis of α

The operating condition at maximum pressure is exploited for the analysis of the solid resistivity parameter **α**, as it represents a phase of the filtration cycle that is more directly comparable with the experimental setups typically adopted in the literature. During this phase, the system operates under more stable conditions and with a more consolidated cake structure, making it possible to apply well-established theoretical models for the description of the phenomenon.

The adopted approach is based on a model proposed in the literature, which describes the relationship between filtration flow rate and pressure drop, and is subsequently extended by introducing a dependence on a solid packing factor and on the feed pressure. The assumptions introduced lead to a semi-empirical modelling framework which, while simplifying the real behaviour of the system, allows the derivation of a lumped parameter that is sufficiently representative for the purposes of process analysis and monitoring.

The reference model, derived from the literature, is expressed by the following relationship:

$$Q = \frac{\Delta p \, A}{\alpha_1 \, \mu \, c_w \, w + \mu R}$$

where **Q** represents the filtration flow rate, **Δp** the applied pressure drop, **A** the filtration area (in this case the entire surface of the filter cloths), **μ** the fluid viscosity, **c_w** the solid concentration, **w** the chamber height, and **R** the hydraulic resistance of the filter cloth [29].

The model is subsequently extended by introducing a formulation for the parameter $\boldsymbol{\alpha_1}$ that accounts for both the filling degree of the filtration chamber and the dependence on the operating pressure:

$$\alpha_1 = \alpha \, \frac{V_{solido}}{V_{max}} \left(\frac{p}{p_{ref}} \right)^{0.99}$$

where $\mathbf{V_{max}}$ represents the maximum usable volume of the filtration chamber, $\mathbf{V_{solid}}$ the volume of solid present within the chamber, $\mathbf{p_{ref}}$ a normalization factor assumed equal to atmospheric pressure, and **p** the feed pressure.

The exponential dependence on pressure is directly derived from the literature model. Within the pressure range analysed, this dependence is sufficiently close to a linear trend; however, the adoption of an exponent equal to 0.99 was chosen in order to ensure improved numerical stability of the calculation, without significantly altering the physical meaning of the model.

Similarly to the procedure adopted for the analysis of the resistance **R**, by inverting the corresponding formulation it is possible to reconstruct the evolution of the parameter **α** within a single filtration cycle. Figure 27 shows an example of the temporal trend of **α** during a filtration cycle.

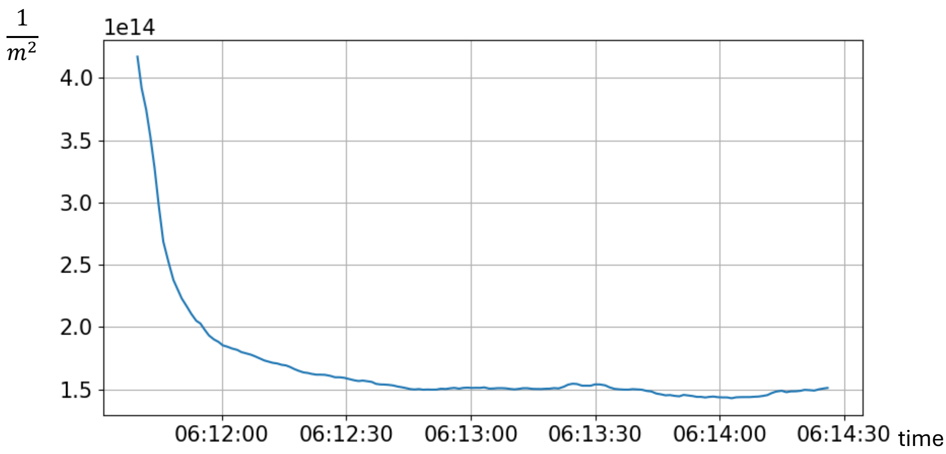

Figure 19 Solid resistivity behaviour for example cycle

Also in this case, the solution obtained for **α** exhibits good geometric repeatability across the different analysed cycles. Figure 28 shows the superposition of multiple cycles, from which an overall consistency in the trend of the parameter can be observed.

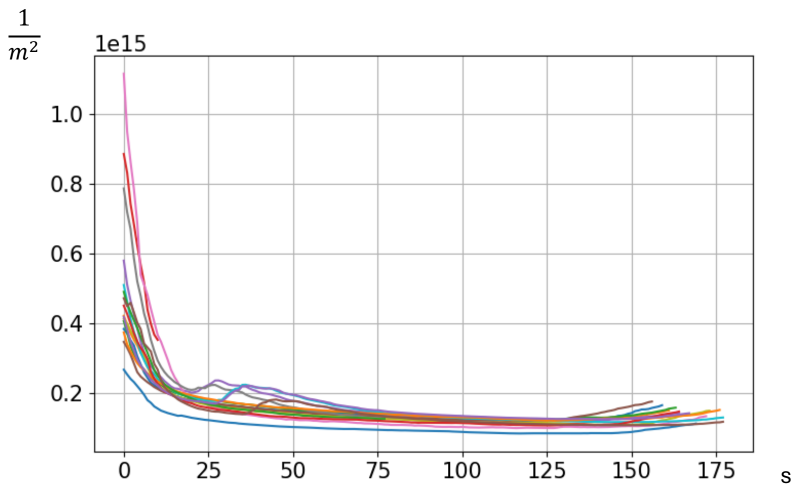

Figure 28 Solid resistivity behaviour for more cycles overlapped

As already observed for the resistance **R**, the trend of **α** is also affected by the presence of transient phenomena, particularly during the initial phase of pressing. This phase is characterized by a change in the machine operating state and by a rapid variation of the operating conditions, which makes the data poorly representative of the assumptions underlying the adopted model. For this reason, the initial portion of the cycle is excluded from the analysis (Figure 29).

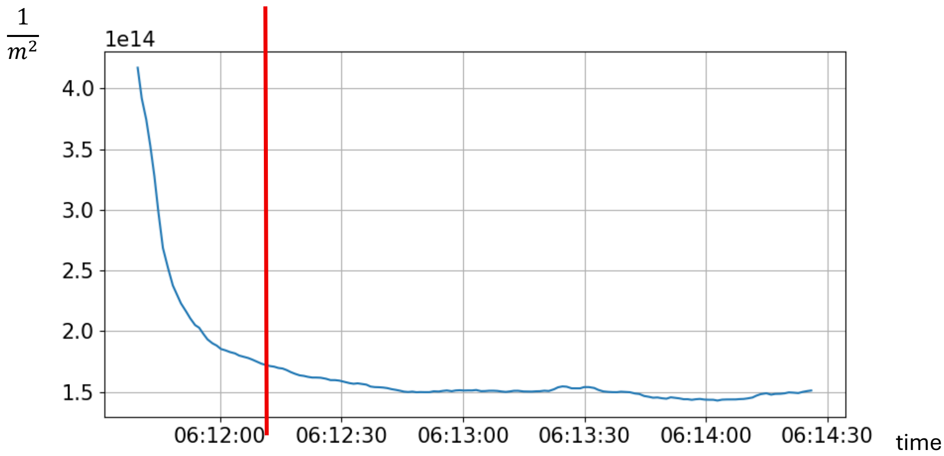

Figure 29 Solid resistivity working region identifying

Once the region of interest has been identified, as shown in Figure 30, the data corresponding to this interval are aggregated.

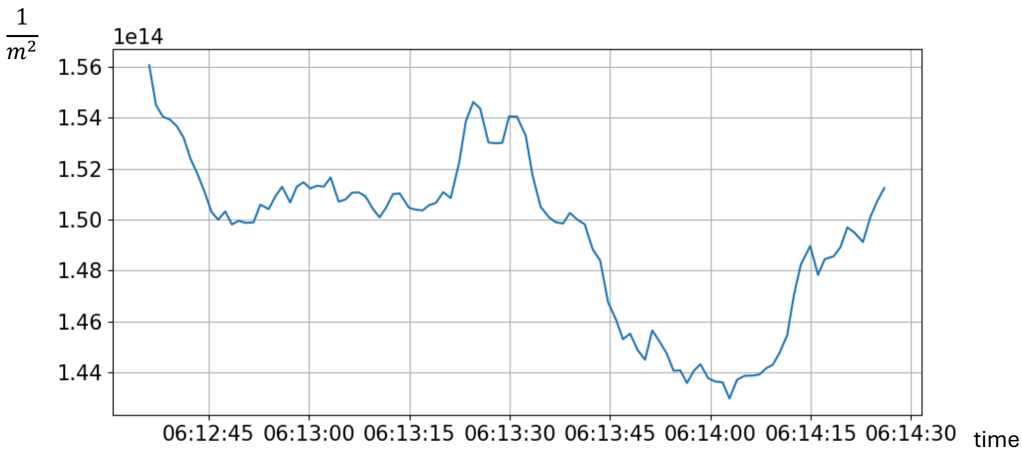

Figure 20 Solid resistivity behaviour for example cycle zoomed

Similarly to the approach adopted in the previous analysis, a data aggregation strategy based on the average of the **α** values within the selected region was applied in order to obtain a representative value for each cycle. Also in this case, it is envisaged that, in a subsequent phase, the analysis will be extended to the entire curve segment, while keeping open the possibility of modifying the aggregation strategy according to analytical requirements and potential future developments of the algorithm.

Figure 31 shows the historical trend of the parameter **α** over the analysed cycles.

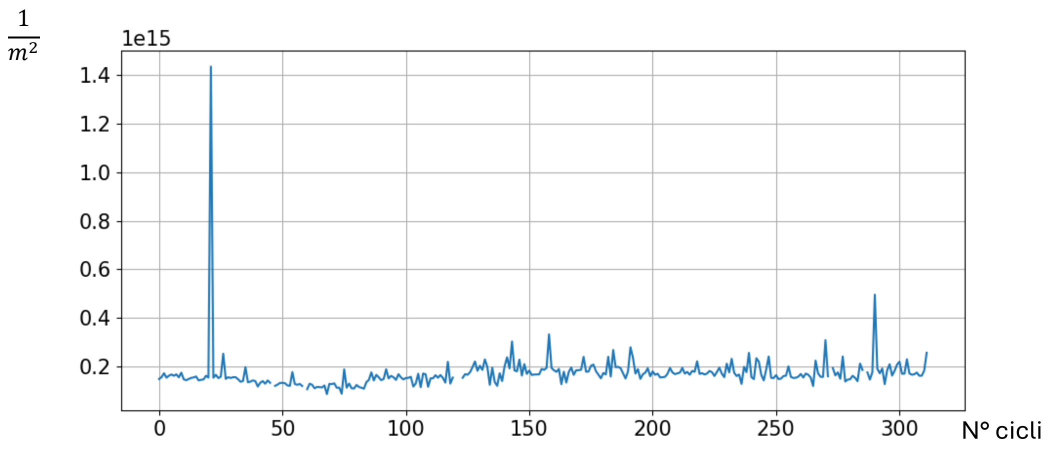

Figure 31 Computed solid resistivity for all cycles analysed

Also for **α**, the presence of some out-of-trend values and of cycles for which a valid result cannot be obtained is observed, likely due to non-compliant operating conditions or measurement disturbances. These situations are nevertheless monitored and are included in the algorithm efficiency parameter, allowing an overall assessment of the reliability of the method.

Finally, Figure 32 reports the same trend with an amplified scale, from which the anomalous values have been excluded in order to improve the readability of the overall trend.

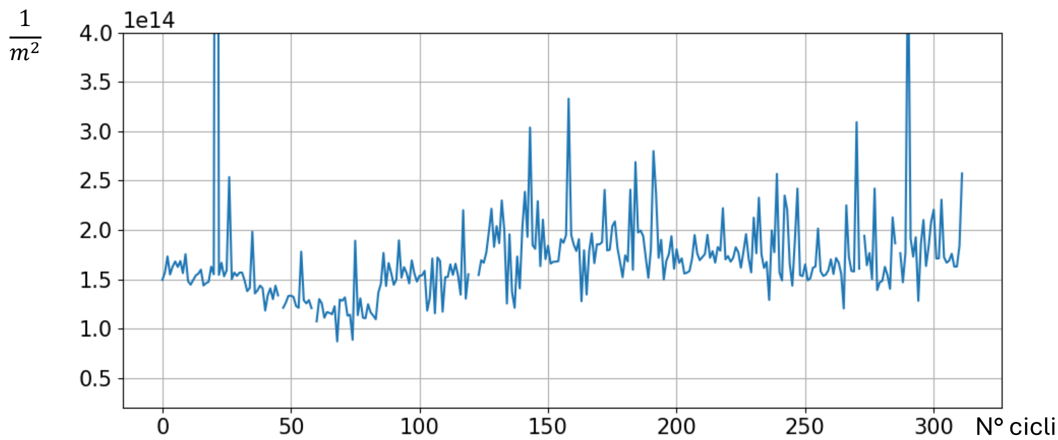

Figure 32 Computed solid resistivity for all cycles analysed zoomed

7. Cycle-Based Aggregation

Data acquired from the operation of a filter press in the mining sector, typically more extensively instrumented than those used in other fields, such as land remediation, generate a large amount of information in the time domain. In order to make these data usable for process state analysis, performance monitoring, and potential developments of diagnostic or optimization algorithms, it is necessary to define a structured aggregation strategy capable of synthesizing the information without losing the link with the underlying physical phenomena. This need becomes even more critical when numerical models or machine learning techniques are applied, as dimensionality reduction while preserving the same information content is essential to keep training times within acceptable limits.

The adopted aggregation scheme is based on multiple methodologies, distinguishing among statistical aggregation, physics-based aggregation, and numerical aggregation. This approach ensures methodological flexibility, maximizing the likelihood of capturing relevant features that are useful for subsequent analyses [30].

Physics-based aggregation

A distinctive aspect of the adopted methodology is the introduction of aggregation driven by process physics, referred to as *physics aggregation*. In this case, aggregation is not limited to a purely numerical synthesis of the data but is instead based on the direct application of analytical models derived from filtration theory.

In particular, this category includes the computation of characteristic parameters of the filtered solid, such as the resistivity $\boldsymbol{\alpha}$, and the estimation of the theoretical resistance of the filter cloth $\mathbf{R}$. These parameters are obtained from measured quantities by applying the corresponding model relationships and are subsequently aggregated over appropriately selected time windows, such as operating regions close to steady-state conditions.

This type of aggregation makes it possible to obtain lumped parameters with a direct physical meaning, enabling comparisons among different cycles and the assessment of the temporal evolution of the system state. Moreover, the use of physics-based models reduces the risk of purely statistical interpretations that lack a physical basis.

These aggregations directly coincide with the outputs of the physical formulations described in the previous chapters.

Basic numerical aggregation

A first class of aggregation concerns the application of simple numerical operators, such as:

- mean value,
- minimum value,
- maximum value,
- time integral of flow rates,
- first oscillatory components in terms of the first three dominant frequencies.

These tools allow an immediate synthesis of the behaviour of measured or derived quantities within a filtration cycle, providing first-level indicators that are easy to interpret.

The mean value yields a representative measure of the system behaviour within a specific time window, while minimum and maximum values make it possible to identify boundary conditions or potential anomalies. The integral, on the other hand, is particularly useful for evaluating cumulative quantities, such as filtered volume or process-related energy, and for comparing cycles characterized by different durations.

This type of aggregation constitutes the common basis for all subsequent processing steps and represents a first level of complexity reduction of the raw data.

Statistical numerical aggregation

A second level of aggregation is represented by the statistical aggregation of cycle data, aimed at summarizing the entire temporal evolution of a filtration cycle into a reduced set of meaningful parameters. In this approach, the cycle is considered the fundamental unit of analysis, in accordance with the inherently cyclic nature of the filtration process and with the need to compare operating conditions repeated over time.

The application of statistical methodologies for data aggregation is intended to identify latent information within the acquired time series that does not directly emerge from the observation of individual process variables. The underlying assumption of this approach is that system behaviour is influenced by physical characteristics of the process that are not directly measurable, and whose manifestation occurs indirectly through the temporal evolution of the observed variables. Such characteristics may be difficult to interpret through purely deterministic analyses or through evaluations based on individual measurement instants.

In this context, statistical aggregation makes it possible to capture recurring patterns, non-obvious relationships, and subtle variations in cycle behaviour, providing a compact yet informative representation of process dynamics. The parameters obtained through this level of aggregation therefore act as synthetic indicators of system operation, potentially sensitive to changes in operating state or to progressive degradation phenomena.

This category includes all outputs derived from the clustering models analysed in the subsequent section, which is dedicated to the application of neural models. These outputs represent an advanced form of statistical aggregation, capable of classifying cycles based on their overall dynamic characteristics and of highlighting similarities and differences among operating conditions that may appear analogous. This approach therefore enriches traditional analysis by providing a complementary interpretation of the data, oriented toward the identification of emerging behaviours and a deeper understanding of the filtration process.

Output structure and data organization

The output of the entire aggregation process is structured so as to associate each operating cycle with a results dictionary containing both the raw data table and the set of aggregated parameters. This organization makes it possible to maintain a direct link between the original data and the derived quantities, ensuring consistency and traceability of the information.

The preservation of raw data allows for detailed analyses or subsequent revisions of the aggregation criteria without the need to rerun previous routines. At the same time, the availability of aggregated parameters provides a concise representation of each cycle, which can be readily used for comparisons among different cycles and for analysing the temporal evolution of the process.

This output structure is therefore flexible and functional, enabling operations at multiple levels of detail and facilitating the integration of results into subsequent stages of system analysis and monitoring.

It should be noted that this structure is specifically tailored to the current methodological research objectives related to the addressed problem. In the event of deployment in an industrial environment, data storage and management processes would need to be revised with a focus on computational efficiency and optimized resource usage.

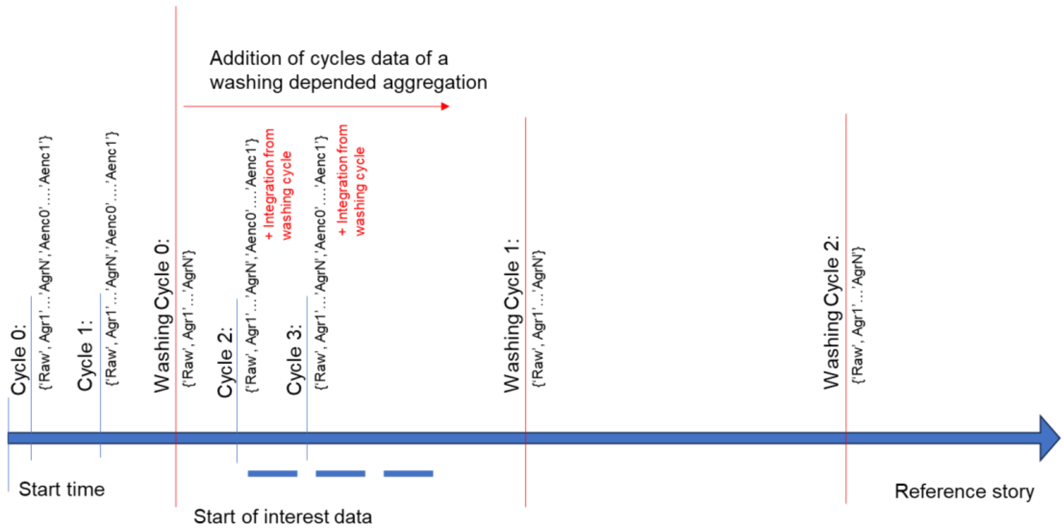

Figure 21 Cycle data storage logic scheme

The choice of adopting a dictionary-based structure makes it possible to integrate information in an orderly manner within the previously constructed data packages, allowing the different types of information to be organized according to a coherent and easily extensible structure. This approach enables the progressive addition of new aggregated parameters without altering the overall architecture of the data management system.

At the end of this processing phase, the files containing information related to individual cycles are enriched with all the computed aggregates while preserving the indexing defined in the initial stages of the analysis. This aspect ensures continuity between raw and processed data, facilitating comparisons among cycles and allowing each aggregated parameter to be traced back to its originating cycle.

The adopted structure is therefore both orderly and flexible, suitable for managing a large number of cycles and functional for subsequent phases of process analysis, visualization, and monitoring.

7.1 Statistical Numerical Aggregation

Numerical aggregation aims to identify and describe intrinsic characteristics of the process that are not directly observable or quantifiable through traditional physical or model-based analyses. In many real-world contexts, the phenomena under investigation exhibit high complexity and a strong variability component, which makes it difficult to isolate individual parameters that are physically interpretable and stable over time.

The underlying idea is that, through appropriate data aggregation techniques, it is possible to define synthetic conditions or global indicators capable of capturing the overall evolution of the system. These aggregated conditions are not intended to provide a detailed description of each individual underlying physical mechanism, but rather to represent the emergent behaviour resulting from the interaction of multiple correlated factors.

This approach is particularly effective in the case of apparently highly stochastic or noisy phenomena, where local variability masks long-term trends and renders traditional physical modelling approaches less effective. Numerical aggregation instead makes it possible to reduce problem complexity, highlighting latent structures, statistical regularities, and evolutionary patterns that are not immediately visible in the raw data.

In this sense, numerical aggregation represents a fundamental tool for the analysis of complex systems, as it enables the derivation of a compact yet informative representation of the process, which is useful both for monitoring purposes and for understanding and predicting its behaviour over time.

In this work, numerical aggregation is implemented through an autoencoder-based system, designed to learn a compact and informative representation of the data from high-dimensional signals. The autoencoder operates by progressively reducing the complexity of the input information, constructing a latent encoding that synthesizes the most relevant characteristics of the observed process. This aggregated representation is not defined a priori on the basis of physical considerations, but is instead learned directly from the data, allowing complex and nonlinear relationships to be captured that are difficult to model using traditional approaches.

Within the autoencoder architecture, the following models were employed:

- one-dimensional convolutional neural networks (1D CNNs);
- feedforward (fully connected) neural networks.

8. Neural Networks

Artificial neural networks are computational models abstractly inspired by the functioning of the biological nervous system. They are designed to learn complex and nonlinear relationships from data, making them particularly suitable for solving problems in which it is difficult to define explicit rules or deterministic mathematical models. Owing to this data-driven learning capability, neural networks are widely applied in numerous fields, including pattern recognition, signal analysis, classification, regression, and the extraction of latent features from complex datasets.

From a structural perspective, a neural network is composed of elementary units called artificial neurons, organized into layers. Each neuron receives a set of input signals, combines them through adaptive weights, and produces an output via a nonlinear activation function. The most common neural network architectures are organized into an input layer, one or more hidden layers, and an output layer, forming structures known as feedforward networks. The presence of multiple layers enables the model to learn hierarchical data representations, in which more abstract features emerge at deeper levels of the network.

The operation of a neural network is based on two main phases: forward propagation and training. During forward propagation, the input data pass through the network layer by layer, undergoing linear and nonlinear transformations until an output is produced. Training consists in optimizing the network weights so as to minimize a cost function that measures the error between the model output and the desired target. This process is generally carried out using gradient descent algorithms and backpropagation, which enable efficient weight updates even in deep network architectures.

One of the fundamental aspects of neural networks is their generalization capability, that is, their ability to provide accurate outputs on data that were not seen during the training phase. To assess this capability, it is common practice to divide the available dataset into distinct training, validation, and test sets. This partitioning allows model performance to be monitored during training and evaluated on independent data, thereby reducing the risk of overfitting.

The evaluation of neural network outputs strongly depends on the type of problem being addressed. In classification tasks, performance is typically assessed using metrics such as accuracy, precision, recall, and confusion matrices. In regression problems, which encompass the majority of the models adopted in this study, error-based indicators such as mean squared error, mean absolute error, or the coefficient of determination are commonly employed. In unsupervised models, such as autoencoders, evaluation is primarily performed through the analysis of reconstruction error and through the interpretation of the learned latent representation.

In addition to quantitative metrics, it is often beneficial to complement the evaluation with a qualitative assessment of the outputs, particularly in cases where the model produces signals, images, or time series. Visual comparison between outputs and reference data makes it possible to identify patterns, distortions, or systematic behaviours that may not be clearly captured by numerical metrics alone. In applied contexts, this qualitative analysis represents a fundamental tool for understanding model behaviour and for assessing its reliability.

In summary, artificial neural networks constitute a powerful tool for data modelling and analysis, owing to their ability to learn complex relationships directly from observations. Their operation relies on an iterative learning process that optimizes the internal parameters of the model, while output evaluation requires a combined use of quantitative metrics and qualitative analyses. This combination enables the development of effective, robust models applicable to a wide range of real-world problems.

8.1 Autoencoders

Autoencoders are a class of machine learning models belonging to the unsupervised learning paradigm, designed with the objective of learning compact, informative, and structured representations of input data [31]. Unlike supervised approaches, in which learning is guided by the presence of explicit labels, autoencoders use the data themselves as a reference, training the model to reproduce the input at the output. This mechanism enables the system to learn intrinsic regularities in the data without relying on external information or on prior assumptions regarding the nature of the observed process.

The principle underlying autoencoders consists in forcing the model to compress the information contained in the original data into an internal representation of reduced dimensionality (Figure 34). This compression is achieved in a controlled manner through an architectural structure that limits the model's ability to directly memorize the input. In this way, the autoencoder is encouraged to capture the most relevant and stable features of the data, while disregarding noise and less informative local fluctuations. This approach is particularly effective when dealing with high-dimensional and complex data, such as time-series signals, industrial process data, or measurements originating from nonlinear dynamic systems [32].

From an architectural standpoint, an autoencoder is generally composed of two main modules: an encoder, which transforms the input data into a compact latent representation, and a decoder, which uses this representation to reconstruct an estimate of the original signal. Model training is carried out by minimizing a loss function that quantifies the discrepancy between the input and its reconstruction, typically using metrics such as mean squared error or other cost functions appropriate to the data type. Although the explicit objective is reconstruction, the true outcome of interest lies in the quality of the learned latent representation.

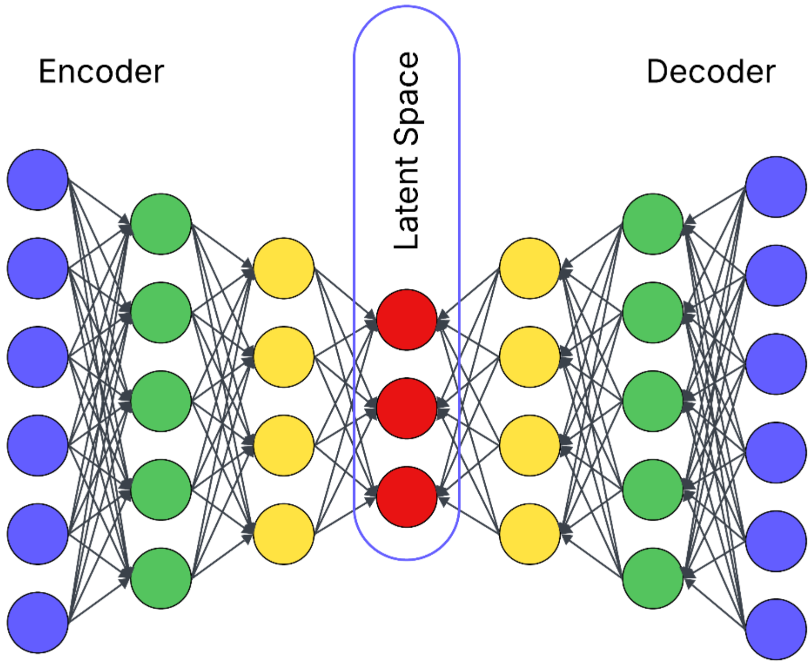

Figure 34 Autoencoder example logic

The latent representation, often referred to as the *bottleneck*, constitutes the conceptual core of an autoencoder. It represents a form of automatic numerical aggregation, as it synthesizes the information contained in the input data into a reduced set of abstract variables. Unlike manually engineered features, which require in-depth domain knowledge, latent variables emerge directly from the learning process and reflect the statistical and dynamical structures present in the data. This makes autoencoders particularly suitable for the exploratory analysis of complex phenomena for which a complete or unambiguous physical modelling is not available.

Throughout the literature, numerous autoencoder variants have been developed, each designed to address specific challenges or to enhance particular model properties. Deep autoencoders extend the basic architecture by introducing multiple hidden layers, enabling the learning of hierarchical representations in which deeper levels capture increasingly abstract features. Sparse autoencoders impose sparsity constraints on the latent representation, promoting solutions in which only a fraction of the units are active for each input, thereby improving interpretability and encoding robustness. Denoising autoencoders, by contrast, are trained to reconstruct the original signal from an intentionally corrupted version of the input, encouraging the model to learn stable structures and reducing sensitivity to noise.

A particularly relevant role is played by convolutional autoencoders, in which fully connected layers are partially or entirely replaced by convolutional layers. This architectural choice allows the exploitation of information locality and weight sharing, making it especially effective for structured data. In the case of time series, one-dimensional convolutions enable the identification of local patterns along the temporal axis, such as transitions, oscillations, or recurring events, while preserving temporal coherence. The use of pooling operations and resolution reduction further enables a progressive abstraction of information.

Autoencoders are also widely applied in the context of time-series analysis and dynamic signal processing, where they can be used for dimensionality reduction, latent feature extraction, noise filtering, and anomaly detection. The ability to operate directly on raw data makes it possible to avoid overly restrictive preprocessing steps, leaving the model to autonomously learn the relevant structures. The resulting latent representations can subsequently be used as inputs for clustering algorithms, condition monitoring systems, or higher-level predictive models.

A crucial aspect in autoencoder design concerns the balance between compression capability and reconstruction quality. An excessively large latent representation may lead the model to a trivial identity mapping, thereby reducing the informational value of the learned aggregation. Conversely, excessive compression may cause significant information loss, compromising the model’s ability to adequately represent the observed phenomenon. The choice of model architecture, number of layers, and bottleneck dimensionality must therefore be made in consideration of the specific characteristics of the data and the objectives of the analysis.

In conclusion, autoencoders represent an extremely flexible and powerful tool for the unsupervised analysis of complex data. Their ability to learn compact, nonlinear, and data-driven representations makes them particularly suitable for the numerical aggregation of high-dimensional signals and for describing the global behaviour of complex dynamic systems. In the following chapters, the specific network architectures employed in this work will be examined in detail, with particular attention devoted to one-dimensional convolutional networks and their role within the autoencoder framework.

8.2 Feedforward Neural Networks

Feedforward neural networks, also known as Multi-Layer Perceptrons (MLPs) or fully connected networks, represent the most classical and fundamental form of artificial neural networks. They constitute the conceptual starting point for many more advanced architectures and are widely used in numerous application domains, both in supervised and unsupervised settings. Their simple and general structure makes them a flexible tool for approximating complex and nonlinear functions [33].

A feedforward network is characterized by a unidirectional flow of information, which proceeds from the input to the output without the presence of cycles or recurrent connections. This property distinguishes feedforward networks from recurrent neural networks and simplifies both their theoretical analysis and training process. In this architecture, neurons are organized into successive layers: an input layer, one or more hidden layers, and an output layer.

Formally, let us consider an input vector

$$\mathbf{x} \in \mathbb{R}^n$$

representing a sample of the observed data. Each layer of the network implements an affine transformation followed by a nonlinear activation function. For a generic layer l, the operation can be expressed as:

$$\mathbf{h}^{(l)} = \phi\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)}+\mathbf{b}^{(l)}\right)$$

where:

- $\mathbf{W}^{(l)}$is the weight matrix of layer l,
- $\mathbf{b}^{(l)}$is the bias vector,
- $\phi(\cdot)$is a nonlinear activation function,
- $\mathbf{h}^{(l-1)}$represents the output of the previous layer, with $\mathbf{h}^{(0)} = \mathbf{x}$.

The activation function plays a crucial role, as it introduces nonlinearity into the model. Without it, a feedforward network composed of multiple layers would reduce to a single linear transformation. Among the most commonly used activation functions is the Rectified Linear Unit (ReLU):

$$\phi(z) = \max\,(0, z)$$

as well as the sigmoid function and the hyperbolic tangent. The choice of activation function significantly affects both the expressive power of the network and the stability of the training process.

From a theoretical standpoint, feedforward networks possess a highly relevant property known as the universal approximation theorem. This result states that a feedforward network with at least one hidden layer and a sufficient number of neurons is capable of approximating any continuous function defined on a compact set with arbitrarily small error. This theorem provides a solid theoretical foundation for the use of feedforward networks as general-purpose models for representing complex nonlinear relationships.

Training a feedforward network consists in estimating the set of parameters $\{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\}$that minimize a cost function L, defined according to the task at hand. In supervised settings, the cost function measures the discrepancy between the network output $\hat{\mathbf{y}}$and the target value $\mathbf{y}$, whereas in unsupervised contexts it may be defined in terms of reconstruction error or internal model consistency. Optimization is generally performed using gradient descent algorithms and backpropagation.

Backpropagation enables the efficient computation of the gradient of the cost function with respect to each network parameter by exploiting the chain rule. In particular, for each layer an error signal is propagated backward, allowing the weights to be updated according to the rule:

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial L}{\partial \mathbf{W}^{(l)}}$$

where ηdenotes the learning rate. This mechanism makes it possible to train deep networks, provided that appropriate strategies are adopted to mitigate issues such as vanishing gradients.

A relevant aspect of feedforward networks is their ability to perform global information combination. Unlike convolutional networks, which operate on local regions of the input, fully connected networks establish connections between each neuron and all units in the previous layer. This characteristic enables the modelling of global dependencies and complex interactions among variables, but it also entails a significant increase in the number of parameters and in the risk of overfitting, especially in the presence of high-dimensional inputs.

For this reason, in modern architectures feedforward networks are often used in combination with other network types. In particular, they are widely employed as integration and compression layers downstream of convolutional or recurrent modules. In such contexts, feedforward networks are responsible for aggregating locally extracted features and projecting them into a more compact and informative representation space.

Within the autoencoder framework, feedforward networks play a fundamental role in both the compression and reconstruction phases of information processing. After pattern extraction through convolutional or recurrent layers, fully connected layers enable the combination of learned features and their progressive reduction until a low-dimensional latent representation is obtained. Similarly, during the decoding phase, feedforward networks allow this representation to be expanded and transformed back into a form compatible with the reconstruction of the original input.

From an interpretative perspective, the representation learned by a feedforward network can be viewed as a nonlinear transformation of the data space, in which the original variables are combined to highlight latent structures and high-level relationships. Although such representations are not directly interpretable in physical terms, they are extremely useful for classification, clustering, regression, and exploratory analysis tasks.

In summary, feedforward networks constitute an essential element of the neural network ecosystem. Their structural simplicity, combined with substantial expressive capability, makes them a versatile tool for learning complex nonlinear relationships. In the context of this work, they represent a key component within more articulated architectures, contributing significantly to numerical aggregation and to the synthesis of information extracted from process data.

8.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) represent one of the most relevant architectures in the field of deep learning, particularly for the analysis of structured data characterized by strong local correlations. Initially introduced in the context of computer vision, convolutional networks have progressively established themselves in many other application domains, such as signal processing, time-series analysis, and the treatment of data originating from complex dynamical systems. Their success is closely linked to their ability to automatically extract meaningful features from data, reducing the need for manual feature engineering.

The fundamental principle underlying convolutional networks consists in the application of local filters that systematically slide over the input, enabling the identification of recurring patterns and local structures. Unlike fully connected networks, in which each neuron is connected to all units in the previous layer, CNNs introduce local connectivity and weight sharing, drastically reducing the number of parameters to be learned. This property makes convolutional networks more efficient, more robust to noise, and better scalable than dense networks of comparable depth [34].

From an architectural perspective, a convolutional network is generally composed of a sequence of convolutional layers, possibly interleaved with dimensionality reduction layers (pooling), followed by one or more fully connected layers in the final stages of the model. Convolutional layers apply a set of kernels or filters to the input, producing activation maps that highlight the presence of specific local patterns. Each filter is designed to respond to particular signal characteristics, such as edges, transitions, oscillations, or localized events, depending on the nature of the analysed data.

From a formal standpoint, let us consider a one-dimensional input signal

$$\mathbf{x} = [x_1, x_2, \dots, x_T]$$

of length T. A one-dimensional convolution operation applies a filter (or kernel)

$$\mathbf{w} = [w_1, w_2, \dots, w_K]$$

of size K, producing an activation map $\mathbf{y}$defined as:

$$y_t = \sum_{k=1}^{K} w_k \ x_{t+k-1} + b$$

where $b$denotes the bias term.

A nonlinear activation function is then typically applied to each output element, most commonly the Rectified Linear Unit (ReLU).

A central aspect of convolutional networks is weight sharing, whereby the same filter is applied to different regions of the input. This characteristic enables the model to recognize the same pattern regardless of its position, introducing a form of spatial or temporal invariance. In the context of signals and time series, this property translates into the ability to detect recurring events or behaviours along the temporal axis, even in the presence of shifts or local variations.

Pooling layers serve the purpose of reducing information resolution by aggregating local activations into more compact representations. Operations such as max pooling or average pooling reduce data dimensionality, lowering computational cost and increasing model robustness with respect to local variations or noise. This progressive reduction in resolution further promotes increasingly abstract representations of information, whereby deeper network layers capture higher-order features.

In the case of one-dimensional max pooling with window size P, the operation can be defined as:

$$z_j = \max_{t \in P_j} y_t$$

The combined use of convolutional and pooling layers allows CNNs to construct feature hierarchies, in which early layers learn simple and local patterns, while deeper layers integrate such information into more complex and global structures. This hierarchical organization makes convolutional networks particularly suitable for the analysis of complex phenomena, in which global behaviour emerges from the interaction of multiple local components.

Although convolutional networks were originally developed for two-dimensional image processing, the concept of convolution naturally extends to data of different dimensionalities. In the case of one-dimensional convolutions, filters operate along a single axis, typically the temporal one. This makes 1D CNNs especially well suited for the analysis of time series, industrial signals, sensor data, and multivariate sequences, where the temporal dimension plays a central role.

One-dimensional convolutional networks enable the automatic extraction of local temporal patterns, such as transients, cycles, oscillations, or anomalies, while preserving the sequential order of the data. Unlike approaches based on global statistical features, the convolutional paradigm maintains a fine-grained description of signal dynamics, leaving it to the model to identify the relevant structures. This is particularly advantageous in industrial or experimental contexts, where process dynamics may vary over time and are not easily describable through simplified physical models.

A further advantage of convolutional networks lies in their architectural flexibility. The number of filters, their size, network depth, and pooling strategies can be tailored to the characteristics of the data and the objectives of the analysis. This flexibility makes it possible to balance model expressiveness with the need to avoid overfitting, which is especially critical when dealing with limited or noisy datasets.

Within the context of unsupervised learning, convolutional networks find a natural application in autoencoder architectures. In this case, convolutional layers are used to replace or complement fully connected networks, allowing the model to learn latent representations that explicitly account for the local structure of the data. The use of one-dimensional convolutions in autoencoders applied to time series enables more informative numerical aggregations, as they are based on the automatic extraction of meaningful temporal patterns.

In particular, the use of 1D convolutional networks in the encoding phase allows for the progressive reduction of the temporal resolution of the signal while preserving essential dynamic information, whereas the decoding phase employs inverse operations to reconstruct the original signal. This integration of convolutional architectures within the autoencoder framework represents an effective approach for the synthesis and analysis of complex signals, as will be discussed in detail in the chapter dedicated to convolutional autoencoders applied to process data.

8.4 Autoencoders for Aggregation

The autoencoder developed in this work is designed to perform an automatic numerical aggregation of the RAW time series associated with each machine cycle, preserving their complete dynamics while progressively reducing their dimensionality. The model is conceived as a one-dimensional convolutional autoencoder, in which 1D convolutional networks and feedforward networks are integrated within a single end-to-end architecture [35].

The primary objective of the model is not the fine reconstruction of the signal itself, but rather the learning of a compact latent representation that synthesizes the global behaviour of the cycle, filtering out noise and local variability. Reconstruction of the original signal is used solely as a training constraint, ensuring that the latent representation retains the essential information contained in the input data.

Input data

The model receives as input a multivariate time series represented as a matrix:

$$\mathbf{X} \in \mathbb{R}^{T \times C}$$

where Tdenotes the temporal length of the cycle and $C$the number of channels or measured variables. Each sample corresponds to a single machine cycle, provided to the model in its RAW form, without manual feature extraction or preliminary reductions based on physical assumptions.

This choice allows the model to directly learn relevant temporal structures and correlations from the data, avoiding the introduction of biases associated with manual feature engineering.

Convolutional encoding block

The first part of the autoencoder is dedicated to the automatic extraction of local temporal patterns through a sequence of one-dimensional convolutional layers. These layers apply 1D filters along the temporal axis, enabling the identification of transitions, oscillations, and recurring structures within the signal.

The convolutional encoding is organized into multiple successive levels, in which:

- the number of filters progressively decreases,
- temporal resolution is reduced through pooling operations,
- nonlinearities allow the capture of complex and nonlinear relationships.

This phase transforms the original high-dimensional signal into a more compact representation, while preserving temporally structured and physically relevant information.

Compression through feedforward networks

At the end of the convolutional block, the extracted features are flattened and passed to a sequence of fully connected layers. This section of the autoencoder plays a crucial role in the numerical aggregation process, as it enables the global combination of local information extracted by the convolutional layers.

The feedforward layers progressively reduce the dimensionality of the representation until a low-dimensional latent space is reached. This latent space represents the final outcome of the numerical aggregation of the cycle:

$$\mathbf{z} = f_{\text{enc}}(\mathbf{X}), \mathbf{z} \in \mathbb{R}^d$$

with $d \ll T \cdot C$.

The reduced dimensionality of the latent vector forces the model to select and synthesize only the most relevant information, making $\mathbf{z}$an abstract and compact description of the process state during the considered cycle.

Phase	Layer	Layer Type	Parameter	Output shape
Input	Input	Input	Serie temporale RAW, $T = 200$, $C = 130$	(200, 130)
Encoder	Conv1	Convolution 1D	64 filters, kernel = 3, ReLU, padding = same	(200, 64)
Encoder	Pool1	MaxPooling 1D	pool size = 2, padding = same	(100, 64)
Encoder	Conv2	Convolution 1D	32 filters, kernel = 3, ReLU, padding = same	(100, 32)
Encoder	Pool2	MaxPooling 1D	pool size = 2, padding = same	(50, 32)
Encoder	Conv3	Convolution 1D	16 filters, kernel = 3, ReLU, padding = same	(50, 16)
Encoder	Flatten	Flatten	,	(800)
Encoder	Dense1	Fully Connected	128 neuron, ReLU	(128)
Encoder	Dense2	Fully Connected	64 neuron, ReLU	(64)
Bottleneck	Dense3	Fully Connected	32 neuron, ReLU	(32)
Bottleneck	Latent	Fully Connected	**16 neuron**, ReLU	**(16)**
Decoder	Dense4	Fully Connected	32 neuron, ReLU	(32)
Decoder	Dense5	Fully Connected	64 neuron, ReLU	(64)
Decoder	Dense6	Fully Connected	128 neuron, ReLU	(128)
Decoder	Dense7	Fully Connected	800 neuron, ReLU	(800)
Decoder	Reshape	Reshape	,	(50, 16)
Decoder	Conv4	Convolution 1D	16 filters, kernel = 3, ReLU, padding = same	(50, 16)
Decoder	Up1	UpSampling 1D	factor = 2	(100, 16)

Decoder	Conv5	Convolution 1D	32 filtri, kernel = 3, ReLU, padding = same	(100, 32)
Decoder	Up2	UpSampling 1D	factor = 2	(200, 32)
Decoder	Conv6	Convolution 1D	64 filters, kernel = 3, ReLU, padding = same	(200, 64)
Output	Output	Convolution 1D	130 filters, kernel = 3, **Sigmoid**, padding = same	(200, 130)

Table 3 Autoencoder aggregator structure

Latent space as numerical aggregation

The latent space constitutes the central element of the architecture. Each machine cycle is represented by a low-dimensional vector that incorporates temporal, dynamical, and structural information from the original signal. This representation is not designed to be directly interpretable in physical terms, but rather to serve as a basis for subsequent analyses, such as clustering, condition monitoring, or higher-level modelling.

From a methodological perspective, the latent space can be interpreted as a nonlinear mapping of the cycle space:

$$\mathbf{X} \rightarrow \mathbf{Z}$$

in which cycles exhibiting similar behaviour tend to be represented by nearby latent vectors.

Decoding and reconstruction block

The second part of the autoencoder is devoted to reconstructing the original signal from the latent representation. The decoder is structured symmetrically with respect to the encoder, using a sequence of feedforward layers followed by reshaping, upsampling operations, and one-dimensional convolutions.

The decoder implements a mapping:

$$\hat{\mathbf{X}} = f_{\text{dec}}(\mathbf{z})$$

which produces an estimate of the original signal. The reconstruction quality is evaluated through a loss function based on the reconstruction error, typically defined as:

$$\mathcal{L} = \| \mathbf{X} - \hat{\mathbf{X}} \|_2^2$$

It is important to emphasize that the decoder is not used during the analysis phase, but exclusively during training, as a regularization mechanism that constrains the latent space to preserve the essential information of the signal.

Architectural considerations

The combination of one-dimensional convolutional networks and feedforward networks makes it possible to exploit the strengths of both architectures: on the one hand, the local extraction of temporal patterns; on the other hand, the capability for global feature combination. This integration makes the autoencoder particularly suitable for the analysis of complex process signals, characterized by nonlinear dynamics and variability that are difficult to describe using traditional physical models.

Overall, the proposed autoencoder represents a data-driven numerical aggregation system, capable of transforming high-dimensional RAW time series into a compact, informative representation suitable for subsequent analyses. In the following chapters, the obtained results and the applications of the latent representation in the context of process monitoring and analysis will be discussed in detail.

Autoencoder training principle

The training of the autoencoder follows the typical unsupervised learning paradigm, in which the input signal coincides with the desired output signal. Specifically, for each machine cycle provided to the model, the input RAW time series is used as the reference for output reconstruction. No external labels or additional targets are therefore required: the model is trained to reproduce its input as accurately as possible.

Formally, given an input signal $\mathbf{X}$, the autoencoder learns a composite function:

$$\hat{\mathbf{X}} = f_{\text{dec}}(f_{\text{enc}}(\mathbf{X}))$$

and parameter optimization is carried out by minimizing a cost function that measures the discrepancy between $\mathbf{X}$and $\hat{\mathbf{X}}$. In this work, the loss function is defined in terms of reconstruction error, typically expressed as:

$$\mathcal{L} = \| \mathbf{X} - \hat{\mathbf{X}} \|_2^2$$

This approach forces the model to compress the signal information into a reduced-dimensional latent representation that still allows for accurate reconstruction of the input. Since the autoencoder cannot simply memorize the original signal due to the constraint imposed by the latent space, it is encouraged to learn an encoding that captures the most structurally relevant characteristics of the cycle.

It is important to note that reconstructing the original signal does not represent the final objective of the analysis, but rather serves exclusively as the mechanism through which the latent representation is learned. The direct comparison between input and output ensures that the latent space preserves the essential information of the signal, acting as a constraint that regularizes the learning process.

In this context, the autoencoder uses the input signal as a form of self-supervision, with reconstruction error as the sole optimization criterion. The resulting latent representation can therefore be interpreted as a coherent numerical aggregation of the cycle, learned directly from the data and independent of physical assumptions or explicit process models.

8.4.1 Autoencoder Results on Time Series

Figure 34 shows a representative example of a signal reconstructed in the time domain using the proposed autoencoder. In the plot, the blue curve represents the original signal, while the red curve indicates the signal reconstructed by the model from the latent representation. Visual comparison highlights the model's good ability to reproduce the overall signal trend, preserving the main dynamic characteristics and the relevant temporal structures.

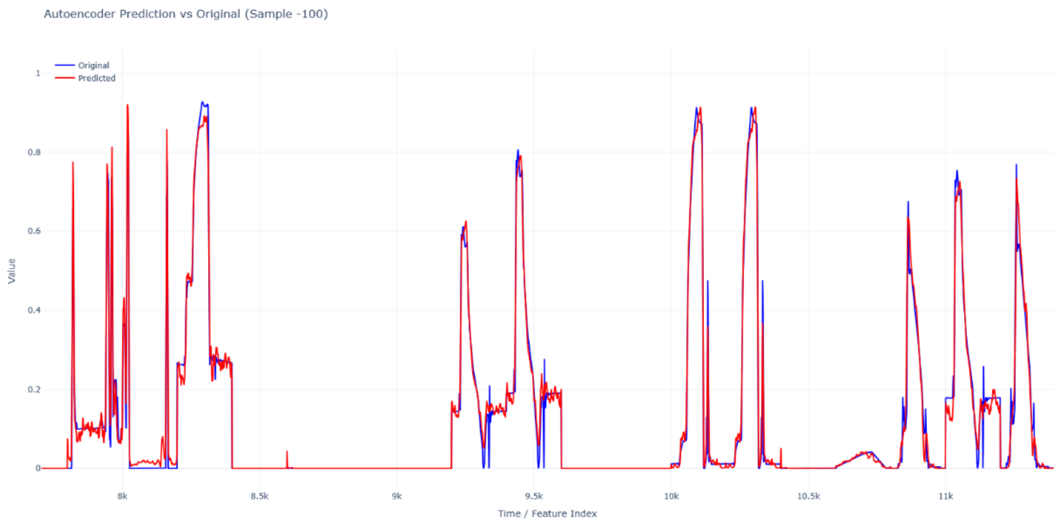

Figure 35 Autoencoder reconstruction output

The quantitative analysis of reconstruction (Figure 35) performance shows that, over the entire testing domain, the average reconstruction error is approximately 2.5%. This value indicates that the autoencoder is able to effectively compress the information contained in the RAW time series and to reconstruct it with limited loss, thereby confirming the quality of the learned latent representation. It is important to emphasize that this result does not stem from a simple memorization of the signal, but rather from the learning of a compact encoding that captures the structurally most relevant components of the process.

Model training was performed using 80% of the available cycles, selected as the training set, while the remaining 20% were kept separate and used exclusively as a test set in order to evaluate the generalization capability of the autoencoder on data not seen during the learning phase. This data split ensures that the observed performance is representative of the model's behaviour on new cycles and is not influenced by overfitting phenomena.

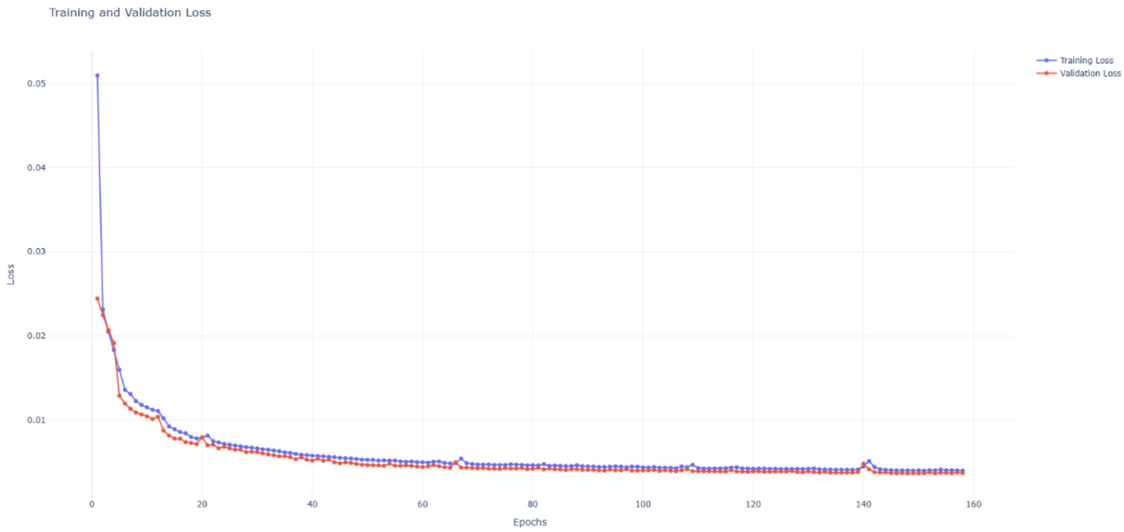

Figure 36 Autoencoder training and testing losses

The validation-loss and training-loss curves confirm the stability of the model, as both exhibit an asymptotic decrease toward zero (Figure 36). This behaviour indicates a well-converged training process and the absence of significant divergence between training and validation performance[36].

The sample shown in the figure represents a segment of the overall temporal history, obtained by consecutively assembling all input variables. This representation enables a compact visualization of the multivariate behaviour of the signal and highlights the model's ability to simultaneously reconstruct the different components of the process. Although the plot shows only a limited portion of the total signal, it is indicative of the overall performance of the autoencoder across the entire analysed temporal domain.

Overall, the reconstruction results confirm the effectiveness of the autoencoder as a numerical aggregation tool, capable of synthesizing complex temporal signals while maintaining a high level of fidelity with respect to the original data. In the following sections, the informational content of the latent representation and its use for process analysis and monitoring will be examined.

9 Performance Index Definition

From the perspective of any optimization activity, it is necessary to identify a performance index of interest. Such an index is a numerical parameter that must reflect the industrial requirements of the analysed process. In the specific case of the filter press, this parameter must represent the productivity of the machine while simultaneously accounting for possible variations in the quality of the output product [37].

The performance parameter was therefore defined according to the following formulation:

$$KPI = \left((slurry\ input\ flow)^{ex\_s} * (humididty\ of\ out)^{-hy\_s}\right)^{\frac{1}{(ex_s+hy_s)}}$$

This definition of the performance index (KPI) makes it possible to track both the incoming slurry flow rate, and thus the amount of processed material, and the moisture content of the discharged product. Operating conditions characterized by a high slurry flow rate and low product moisture result in a high value of the index, whereas a low flow rate combined with high moisture content leads to a reduced KPI value.

The two exponents appearing in the formulation allow the relative weight of the two contributions to be calibrated. These exponents represent adjustable parameters, selected on the basis of industrial considerations related to machine performance evaluation. In other words, they depend on the relative importance assigned to productivity versus final product quality.

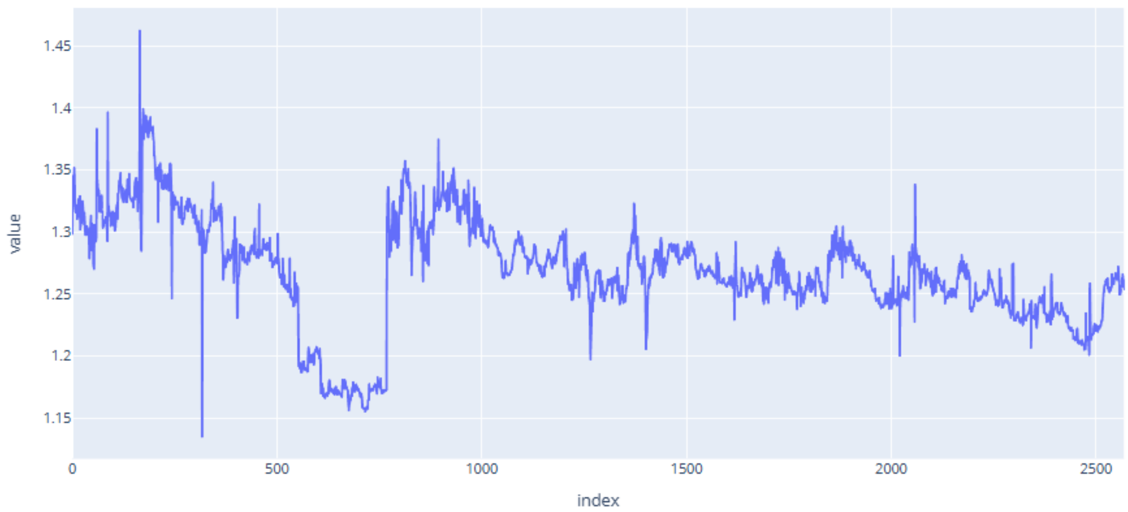

Figure 37 Computed KPI

Figure 37 shows the trend of the KPI for the analysed sample of cycles, considering both weighting coefficients set equal to 1.

Once the KPI has been defined in these terms, it is appropriate to introduce a clarification regarding the manner in which the solid throughput is defined. This quantity, in fact, represents a component of the KPI that does not depend exclusively on the instantaneous state of the process within the single analysed cycle. For all practical purposes, the analysis is conducted in the cycle domain, and the evaluation of machine productivity cannot be limited solely to the amount of material processed within an individual operating cycle.

Productivity is strongly influenced by the temporal evolution of the process and by the conditions occurring during both operational and auxiliary phases, as well as by phenomena that develop over a time scale longer than the duration of a single cycle. For this reason, a correct evaluation of the solid throughput must necessarily take into account a historical neighbourhood composed of multiple consecutive cycles, in order to identify possible evolutionary trends in machine behaviour.

In order to construct a robust indicator of productivity, or more generally of the processed material flow rate, the numerical derivative of the cumulative processed material curve over time was therefore employed. This curve represents the overall evolution of the total amount of solid processed by the machine, while its derivative provides an estimate of the material throughput, similarly to what would be obtained in a continuous machine, that is, the effective throughput of the system. This approach makes it possible to obtain an index that describes the machine behaviour in the temporal vicinity of the considered cycle.

Finally, it is essential to clearly distinguish this solid throughput index from the filtration efficiency index, which is computed over the duration of a single cycle only. The latter is useful for evaluating the specific performance of the filtration phase, but it does not account for auxiliary process phases nor for temporal dynamics developing across successive cycles. Consequently, while the filtration efficiency index provides localized information on cycle-level performance, the solid throughput index offers a more global and dynamic view of machine productivity.

In summary, the distinction between these two indicators is essential for a correct interpretation of system performance: the former is oriented toward the evaluation of individual operating phases, whereas the latter enables the analysis of the overall process evolution and the long-term behaviour of the machine.

10 KPI Predictor–Optimizer

Once an appropriate performance index (KPI) has been defined, it is possible to proceed with two subsequent analysis stages. The first stage consists in the development of a feedforward model aimed at estimating the KPI value associated with each machine operating cycle based on the characteristic parameters of the cycle itself. This model enables the description of the potentially nonlinear relationship between process variables and the performance index, providing a quantitative evaluation of performance for each operating cycle.

The second stage involves the development of a predictive model based on Recurrent Neural Networks (RNNs), applied to cycle parameters that have been appropriately aggregated in the time domain. This approach makes it possible to model the dynamic evolution of the process and to capture temporal dependencies between consecutive cycles, thereby enabling the prediction of the future behaviour of process parameters and, consequently, of the expected KPI value.

This methodology allows the future evolution of the performance index to be estimated in advance and the temporal occurrence of the next machine washing event to be evaluated. By comparing the predicted KPI values with an industrially defined acceptability threshold, it becomes possible to proactively schedule washing interventions, avoiding both excessive performance degradation and premature interventions, with a resulting improvement in overall operational efficiency.

10.1 Recurrent Neural Networks (RNN)

In the field of machine learning, Recurrent Neural Networks (RNNs) represent a class of models particularly suited for the analysis and modelling of sequential and temporal data. Unlike feedforward neural networks, in which information processing is static and devoid of temporal dependencies, RNNs introduce recurrent connections that allow the maintenance of an internal representation of the system state over time [38].

This characteristic makes RNNs well suited for describing dynamic processes, in which the future evolution of the system depends not only on the current input but also on past history. Such models find application in numerous domains, including time-series analysis, industrial signal processing, process variable forecasting, and, more generally, the modelling of systems characterized by temporal dependencies.

Formally, given a sequential input $\{\mathbf{x}_t\}_{t=1}^{T}$, an RNN computes a sequence of hidden states $\{\mathbf{h}_t\}$according to the relation:

$$\mathbf{h}_t = f(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h)$$

where $\mathbf{W}_{xh}$and $\mathbf{W}_{hh}$are the weight matrices associated with the input and the previous hidden state, respectively, and $f(\cdot)$is a nonlinear activation function. The network output can be expressed as:

$$\mathbf{y}_t = g\left(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y\right)$$

Training of these models is typically performed using backpropagation through time (BPTT). However, traditional RNNs suffer from well-known limitations related to gradient instability, which hinder the learning of long-term temporal dependencies.

Over time, several recurrent architectures have been proposed to overcome the limitations of classical RNNs and to improve their ability to represent complex temporal dynamics.

Traditional RNNs, often referred to as *vanilla RNNs*, exhibit a simple structure but are poorly suited for long sequences due to the vanishing or exploding gradient problem. To address these issues, architectures incorporating mechanisms for controlling information flow have been introduced.

Among these, Gated Recurrent Units (GRUs) introduce update and reset gates that enable more efficient management of internal memory, while maintaining relatively low computational complexity.

Long Short-Term Memory (LSTM) networks, which are discussed in greater detail below, represent one of the most established and widely adopted solutions, owing to their ability to preserve relevant information over extended temporal horizons [39].

Long Short-Term Memory architecture

Long Short-Term Memory networks were introduced with the objective of structurally addressing the long-term memory problem in recurrent networks. The distinctive feature of this architecture is the presence of a *cell state*, which enables the propagation of information along the temporal sequence with minimal distortion.

Each LSTM unit consists of:

- a cell state $\mathbf{c}_t$,
- a hidden state $\mathbf{h}_t$,
- three control mechanisms, known as gates: the input gate, the forget gate, and the output gate.

These gates dynamically regulate which information should be stored, forgotten, or exposed at the output, allowing the network to adapt to the dynamics of the input signal.

The operation of an LSTM cell can be formally described by a set of equations governing the evolution of the internal state. Given the input $\mathbf{x}_t$and the previous hidden state $\mathbf{h}_{t-1}$, the gates are defined as follows.

The forget gate determines how much past information is retained:

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f\right)$$

The input gate controls the incorporation of new information into the cell state:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i)$$

The candidate cell state is computed as:

$$\tilde{\mathbf{c}}_t = \tanh\,(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c)$$

The cell state is then updated according to:

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t$$

Finally, the output gate regulates the generation of the hidden state:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o)$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$$

Thanks to this structure, gradients can propagate along the cell state without rapidly vanishing, enabling the learning of long-term temporal dependencies.

Training of LSTM networks is performed using BPTT, similarly to classical RNNs, but benefits from improved numerical stability. Nevertheless, the architectural complexity entails a large number of parameters, leading to increased computational cost.

To mitigate overfitting and improve model generalization, techniques such as regularization, dropout applied to recurrent connections, and early stopping are commonly employed. Additionally, input sequence normalization and segmentation are crucial in real-world applications.

Recurrent neural networks constitute a key tool for modelling dynamic systems. In particular, Long Short-Term Memory networks represent an effective and well-established solution for the analysis of complex temporal sequences, thanks to their ability to handle long-term dependencies and adapt to nonlinear dynamics.

For these reasons, LSTM networks are particularly well suited for industrial and process-oriented applications, such as those addressed in this thesis, where understanding the temporal evolution of observed variables plays a central role.

10.2 Forecaster

The objective of the developed predictive model is to estimate the future evolution of the aggregated process parameters and, consequently, of the performance indicator (KPI). In this analysis phase, the forecasting task was carried out exclusively on the so-called *statistical* parameters, namely the latent variables extracted from the autoencoder. These latent features represent a compact yet highly informative synthesis of the operating cycle behaviour, allowing the model to work on a reduced representation of the data while preserving the most relevant dynamics.

The model is designed to operate on a historical temporal window of data, composed of a fixed number of consecutive samples, in order to predict the value of the immediately following sample. This strategy enables the network to capture short-term and local temporal dependencies that characterize the sequence of aggregated parameters. By iteratively repeating this procedure, shifting the temporal window forward by one sample at a time, the model can generate a multi-step prediction trajectory, describing the future evolution of the system over an extended time horizon. This iterative forecasting mechanism makes it possible to simulate the expected trend of the process and provides an early indication of potential deviations or performance degradation.

From an architectural standpoint, the neural network combines two complementary modelling approaches in a synergistic way: Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The LSTM modules are employed to learn temporal dependencies and identify recurrent patterns at different time scales, leveraging their intrinsic capability to retain and update memory of relevant information over time. In parallel, the convolutional layers perform a filtering operation on the input sequences, proving particularly effective in extracting structured patterns and repetitive dynamics across wider temporal spans, while also improving robustness to noise and local variability.

These two processing branches operate simultaneously on different temporal resolutions, enabling the model to integrate both short- and medium-term dynamic information. The contributions from the convolutional and recurrent components are subsequently fused and propagated through a feed-forward layer, which generates as output the predicted vector of aggregated process parameters for the next time step. This fusion mechanism enhances the representational capacity of the network by combining feature extraction capabilities with sequential modelling, ultimately improving prediction consistency and temporal stability.

This hybrid approach allows the model to exploit the strengths of both CNN and LSTM architectures, increasing its ability to generalize across different operating conditions and to produce reliable and coherent forecasts over time. Finally, the predicted aggregated parameters are used as input for the KPI estimation model, enabling an anticipatory evaluation of the system's performance evolution. This predictive framework supports proactive decision-making, allowing timely operational interventions, improved process monitoring, and more effective predictive maintenance strategies.

10.2.1 Hybrid Forecasting Model

The objective of the developed predictive model is to estimate the future evolution of the aggregated process parameters and, consequently, of the performance indicator (KPI). In this analysis phase, the prediction is performed exclusively on "statistical" parameters, namely on the latent variables derived from the autoencoder, which provide a compact and informative synthesis of the behaviour of each operating cycle [40].

The model is designed to operate on a historical temporal window of data, consisting of a fixed number of consecutive samples, in order to predict the value of the immediately following sample. This approach enables the capture of local and short-term temporal dependencies present in the sequence of aggregated parameters. By iteratively repeating this procedure and shifting the temporal window forward by one sample at a time, the model is able to generate a predictive trajectory that describes the future evolution of the system over an extended time horizon.

The neural network architecture combines two different modelling techniques in a synergistic manner: Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The LSTM components are employed to model temporal dependencies and to identify recurring patterns across time scales, exploiting their ability to retain relevant information over time. In parallel, the convolutional layers perform filtering of the input sequences, proving particularly effective in extracting repetitive patterns over broader temporal scales.

The two processing paths operate in parallel on different temporal scales, allowing the model to integrate dynamic information over both short- and medium-term horizons. The contributions of the convolutional and recurrent components are subsequently combined and fed into a feedforward layer, whose role is to produce as output the vector of aggregated parameters predicted at the next time step.

This hybrid approach makes it possible to leverage the strengths of both architectures, enhancing the model's generalization capability and enabling stable and coherent predictions over time. The predicted aggregated parameters are finally used as input for the KPI estimation model, allowing the anticipated evaluation of system performance evolution and supporting predictive, data-driven operational decisions.

10.2.2 Predictor Results

The defined model enables the processing of the temporal history of the aggregated parameters and the generation of simulated prediction curves over the time horizon of interest. In a first phase, for each available aggregated cycle, a historical data window is constructed and used as input to the predictive model. Starting from this window, the model iteratively estimates the evolution of the aggregated parameters, producing predictions for a fixed number of subsequent cycles. In the present case, the prediction horizon was set to 20 cycles, under the realistic assumption that no information about future cycles is available Figure 38.

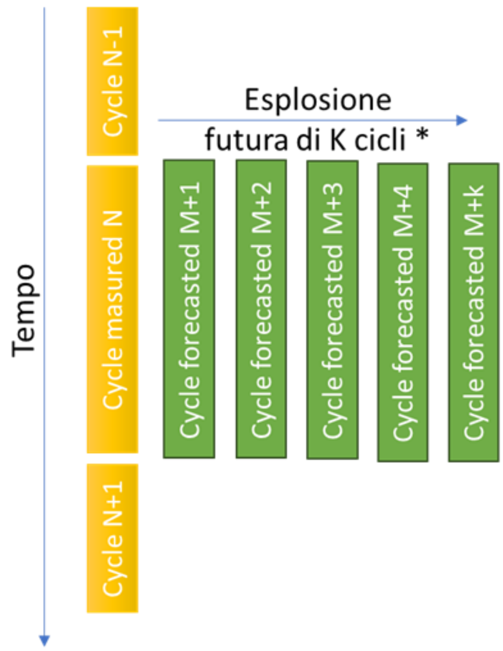

Figure 22 Forecaster scheme

In a second phase, the obtained predictions are propagated forward in time by exclusively using the values estimated by the model to update the input window, thus simulating an autonomous prediction condition (open-loop). This approach makes it possible to assess the stability and coherence of the predicted trajectories over the medium term, as well as the model's ability to capture the dynamic evolution of the process in the absence of real future measurements.

Finally, in a third validation phase, the generated prediction curves are compared with the corresponding variables actually acquired by the system in subsequent cycles. This comparison allows the accuracy of the predictive model to be evaluated by analysing the deviation between predicted and measured values, and to quantify the reliability of the predictions as a function of the considered prediction horizon.

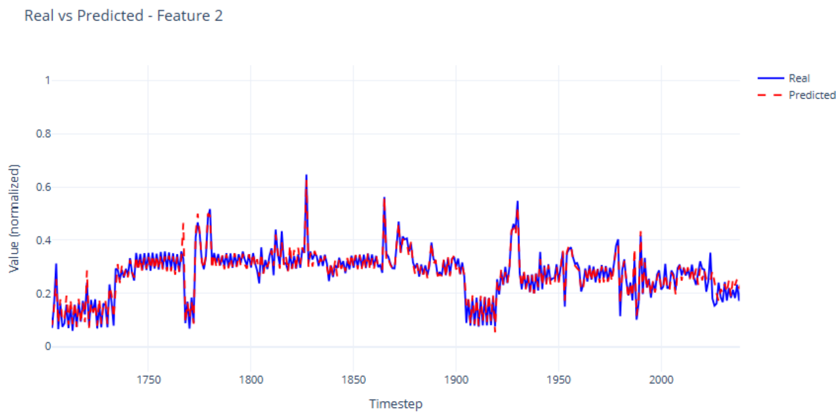

Figure 39 Aggregate forecaster on training data

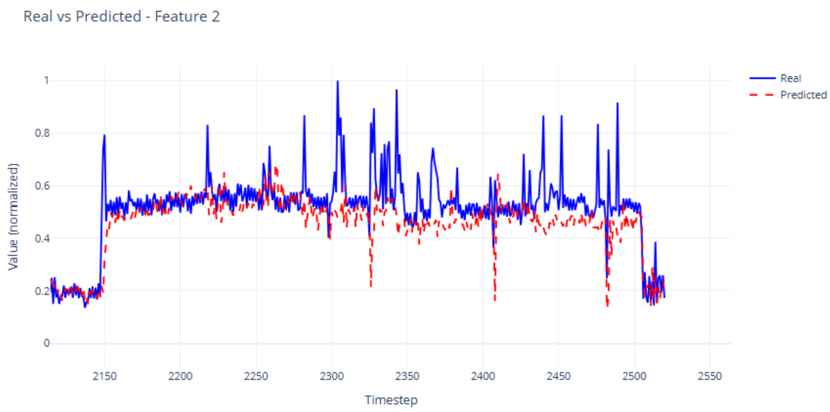

Figure 40 Aggregate forecaster on testing data

As can be observed in Figures 39 and 40, referring to the comparison between real and predicted values of a feature selected for illustrative purposes, a high degree of agreement is observed in the training data domain, whereas this agreement is significantly reduced in the testing data domain. This behaviour is further confirmed by the trend of the loss function reported in Figure 41, which shows a progressive reduction of the loss on the training set accompanied by a stabilization on the validation set.

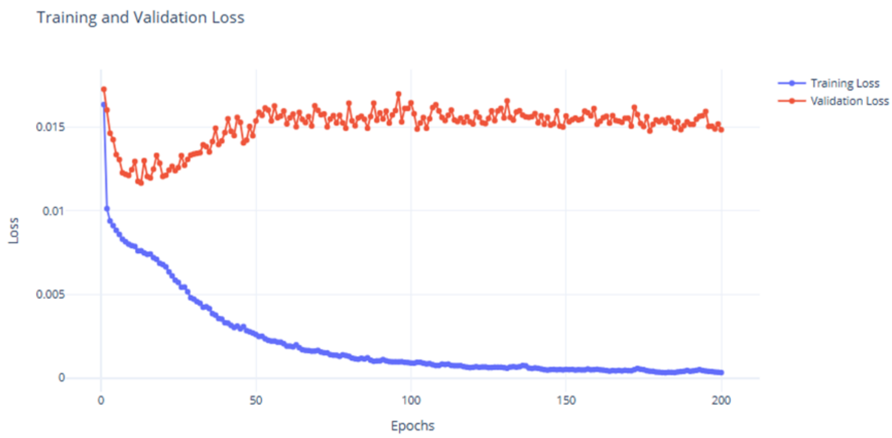

Figure 41 losses for forecaster model

The obtained result clearly highlights the presence of an overfitting phenomenon, which is typical of complex neural models trained on datasets of limited size. Under such conditions, the neural network proves particularly effective at memorizing historical patterns observed during the training phase, but exhibits a reduced generalization capability when applied to previously unseen data. In other words, the model accurately learns the training domain, but fails to effectively recognize patterns that are useful for producing reliable predictions outside that domain [41].

This evidence necessitates a critical evaluation of the applicability of the proposed system. If the model is nevertheless considered usable, it becomes essential to explicitly and rigorously define its domain of applicability, clearly specifying the operating conditions under which the predictions can be regarded as reliable.

In the specific case analysed, the modelling process is subject to a boundary condition that significantly limits both its effectiveness and its potential for improvement. This limitation is primarily attributable to the relationship between the intrinsic variability of the phenomenon to be modelled and the size of the available historical dataset. In other words, the acquired data are neither sufficiently numerous nor sufficiently diverse to provide an adequate representation of the full set of dynamic patterns that may be observed during machine operation.

A closely related issue, which essentially reflects the same concept, is that the machine operated for most of the observation period under relatively stationary conditions. In this context, the observed signal variability is predominantly associated with noise or with small-amplitude local oscillations. As a consequence, the predictive model tends to focus on estimating these fluctuations of limited operational relevance, rather than identifying structural variations or meaningful process trends.

Indeed, the overall trend of the signal is correctly captured by the model, while high-frequency local variations prove difficult to predict. From a numerical and statistical standpoint, the two conditions described, limited process variability and reduced dataset size, lead to the same outcome: an incomplete representation of the system state space. In particular, local variability that is not adequately represented in the historical data can be attributed, at least in part, to components of a purely random nature.

If such variability effectively corresponds to uncorrelated stochastic noise, there exists no finite dataset size capable of enabling the learning of a meaningful predictive structure, precisely due to its intrinsically random nature. In this sense, attempting to model or predict the purely random component of the signal is neither scientifically nor industrially relevant.

However, this consideration highlights a crucial aspect: in order to obtain predictions of genuine practical interest, the system must exhibit structured and appreciable variability. This implies the need to experimentally explore additional operating points of the machine, introducing different operating conditions that allow the dataset to be enriched and the validity domain of the model to be expanded. Only through an adequate exploration of the operating space can the model's generalization capability be improved and robust, industrially usable predictions be achieved.

10.2.3 KPI Analysis on Forecasts

Consequently, once a reliable predictive model is available, it becomes natural to extend the analysis to the evaluation of the performance index (KPI) based on the estimated data. To this end, a representative KPI surface can be constructed, in which the real and predicted values of the aggregated parameters are related. Specifically, the **Y-axis** reports the observed real values, the **X-axis** the future predictions obtained from the model, while the **Z-axis** represents the KPI value computed from these quantities.

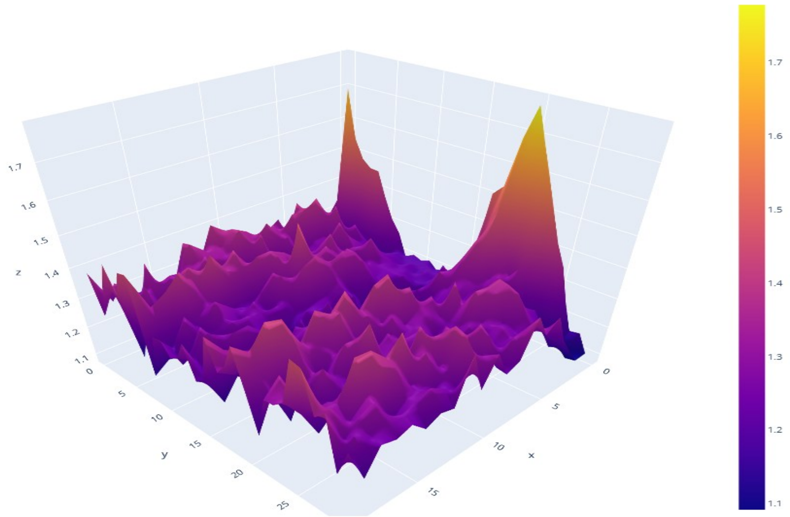

Figure 42 KPI prevision surface

The analysis of the surface shown in Figure 42 highlights the same critical issues already encountered in the evaluation of the predictive model for the aggregated parameters. Similarly to what was observed in the temporal history of the input variables, the KPI behaviour appears to be essentially stable and characterized by oscillations of limited amplitude around an almost constant mean value. This trend is evident both along the plane corresponding to the real values, identifiable as the plane $(x = 0, y, z)$, and along the direction associated with the predictions over the subsequent 20 cycles, that is, along the plane $(x, y = \text{observation cycle}, z)$.

This result confirms that, under the analysed operating conditions, the system operates in a strongly stationary regime, in which process variability is insufficient to produce significant variations in the performance index. Consequently, the KPI surface does not exhibit marked transitions or critical regions that could be effectively exploited for optimization purposes.

In light of these considerations, it becomes evident that, in order to properly and comprehensively validate the proposed analysis, it is necessary to observe the machine behaviour under operating conditions characterized by greater variability. In particular, from the perspective of optimization related to the scheduling of washing operations, it would be appropriate to introduce a controlled variation of washing times, which are currently kept constant. Such variation would allow the exploration of different system states and the generation of a more informative KPI dynamics.

This operation could also be implemented in a dynamic manner, by analysing the model output in real time and simultaneously monitoring the KPI trend computed from real data, represented by the plane $(x = 0)$, and that predicted for future cycles, represented by the plane associated with the observation-cycle direction. Such an approach would make it possible to use the KPI surface as a decision-support tool, enabling the early identification of performance degradation and the adaptive, data-driven optimization of the temporal placement of washing operations.

11 Direct Solution and Anomaly Identification

Contrary to a purely iterative approach aimed at identifying a solution to the washing problem through the optimization of a performance parameter, the objective of this work is the construction of a reference variable capable of representing the evolutionary state of the machine. Such a variable should enable the identification of anomalous or particularly relevant historical conditions, thereby allowing a deeper analysis of system behaviour over time.

The availability of a reference state parameter is particularly valuable for intercepting operating conditions that have a decisive influence on the occurrence of specific phenomena, whether these correspond to situations that should be avoided or to desirable states that should be achieved. This type of scenario is especially relevant in cases where a clear and directly measurable performance parameter to be maximized is not available, or when the primary objective is not the optimization of nominal performance, but rather the prevention of ancillary conditions that may significantly affect machine functionality, reliability, or operational continuity.

In the specific case analysed, it has been observed that the identified performance parameter remains particularly stable both in the historical analysis and in future predictions. This stability is plausibly attributable to the fact that the machine is constrained to operate within a “safe” operating region from a productivity standpoint, while the main criticalities triggering intervention do not directly affect productivity itself. In other words, the system is intentionally kept away from potentially degrading conditions, making any progressive performance deterioration scarcely observable.

Referring to a more general scenario, this situation can be likened to that of a system subjected to highly anticipatory maintenance interventions in order to prevent sudden failures. Under such conditions, the natural deterioration associated with component aging does not clearly manifest itself in observable performance parameters, since the system is restored before these effects become measurable.

In the context of high-pressure washing applied to the machine under study, critical phenomena indeed occur, such as the failure of the filter cake to detach. These events require washing interventions earlier than those that would be dictated by a simple degradation of filtration performance. As a consequence, the intervention logic currently adopted is not directly linked to a performance index, but rather to the prevention of potentially problematic operating conditions.

This scenario is of particular interest because, if it were possible to characterize and model the dependence of such anomalies (e.g., failure of cake detachment) on the operating conditions of the system, it would be possible to identify a boundary operating region beyond which the risk of anomaly becomes significant. This would enable a more rational and targeted definition of washing intervention timing, preventing the occurrence of undesired phenomena and overcoming a purely empirical approach.

In the specific case considered, however, no variable is currently available that can provide even a qualitative assessment of the functional state of the filter. The operational setup of washing procedures is in fact based exclusively on the historical experience of the machine manufacturer, without objective information regarding the validity or optimality of this choice in the specific operating context analysed. Furthermore, it is not possible to assess how conservative this approach is, or conversely, to what extent it could be improved.

To address this critical limitation, an impact analysis of historically performed washing operations on the variables acquired from the filter was therefore undertaken. The objective of this analysis is to identify correlations, patterns, or significant variations in the measured signals that may act as indirect indicators of system state. In this way, the groundwork is laid for the definition of a reference state variable capable of describing the machine's evolution over time and of supporting future strategies for monitoring and optimization of the washing process.

12 Similarity Analysis

The idea underlying this analysis concerns the possibility of inferring the condition of the filter cloths at the time a washing operation is performed by observing the impact that such an intervention has on the overall system behaviour. The working hypothesis is that, under identical operating conditions and with comparable washing system performance, the dirtier the cloth, the greater the change in machine behaviour between the cycles preceding and following the washing operation.

In this context, the washing operation can be interpreted as a controlled perturbation of the system: the effect observed downstream of the intervention provides indirect information about the initial state of the filter cloth. If this effect can be appropriately quantified, it becomes possible to define a measure of behavioural difference between cycles, which can be used as a relative indicator of washing effectiveness. Under the assumption that the washing system operates in a substantially stable and repeatable manner over time, this indicator would depend primarily on the amount of fouling present on the cloth at the time of intervention, rather than on intrinsic variations in the washing process itself.

To achieve this objective, a similarity analysis among operating cycles was performed, exploiting unsupervised clustering algorithms. The underlying idea is to compare the acquired cycles with one another by evaluating their degree of similarity based on the measured process variables, and to identify groups of cycles exhibiting highly similar behaviour. In this way, it becomes possible to observe how cycles are distributed before and after a washing event and to assess the magnitude of the change introduced by the intervention.

These approaches have already been partially introduced in previous analyses, particularly when addressing the problem of representing the multidimensional space of process variables through reduced-dimensionality structures. In that case, the problem was addressed through the use of autoencoders, with the aim of extracting a compact yet informative representation of the system state.

In the present context, however, the objective is not dimensionality reduction per se, but rather the use of unsupervised aggregation methodologies to group cycles that are highly similar in terms of their dynamic behaviour. Clustering techniques enable the identification of recurring patterns in the data without requiring labels or a priori information regarding the cleanliness state of the filter cloths.

By analysing the resulting clusters, it becomes possible to evaluate whether the cycles preceding and following a washing operation belong to the same group or to different groups, and to what extent a displacement in feature space is observed. Such displacement can be interpreted as an indirect measure of the impact of the washing operation and, consequently, of the initial condition of the filter cloth.

In the following section, several analysis methodologies and clustering techniques considered particularly suitable for this purpose are presented, together with the rationale supporting their adoption in the specific context of the filtration system under investigation.

12.1 Clustering Methodologies

12.1 .1 K-means Algorithm

K-means is one of the simplest and most widely used clustering algorithms, owing to its computational efficiency and ease of implementation. The objective of K-means is to partition a dataset into K distinct clusters, where K is a number specified a priori by the user [42], [43].

The algorithm operates by minimizing the sum of squared distances between each data point and the centroid of the cluster to which it is assigned. Formally, K-means seeks to minimize the following objective function:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} \| x - \boldsymbol{\mu}_i \|^2$$

where C_i denotes the i-th cluster and $\boldsymbol{\mu}_i$ its centroid.

The algorithm can be summarized through the following steps:

1. Random initialization of the K centroids.
2. Assignment of each data point to the nearest centroid.
3. Recalculation of the centroids as the mean of the assigned points.
4. Iteration of steps 2 and 3 until convergence.

Among the main advantages of K-means are its conceptual simplicity, fast execution even on large datasets, and the ease of interpretation of the resulting clusters. However, the algorithm also presents several limitations. First, it requires the number of clusters K to be defined in advance, which is not always known a priori and often must be estimated using empirical methods such as the *elbow method* or the *silhouette coefficient*.

Furthermore, K-means implicitly assumes that clusters are spherical and of similar size, making it less effective in the presence of clusters with irregular shapes or differing densities. The algorithm is also sensitive to outliers and to the initialization of the centroids, which may lead to suboptimal solutions.

12.1.2 HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise

HDBSCAN is a density-based clustering algorithm developed as an extension of DBSCAN (Density-Based Spatial Clustering of Applications with Noise). It combines principles of hierarchical clustering with those of density-based methods, overcoming several limitations of the original DBSCAN approach [44].

Unlike K-means, HDBSCAN does not require the number of clusters to be specified a priori. The algorithm identifies dense regions in the data space and separates them from low-density regions, which are considered noise. The main parameter of HDBSCAN is the minimum cluster size, which controls the granularity of the resulting clustering.

HDBSCAN initially constructs a hierarchical representation of the data based on a local density measure. Subsequently, it extracts an optimal flat partition from this hierarchy by selecting the most stable clusters in terms of persistence.

One of the main advantages of HDBSCAN is its ability to identify clusters of arbitrary shape and varying density, making it particularly suitable for complex, real-world datasets. In addition, the algorithm naturally handles noise and outliers, which are explicitly labelled as such.

However, HDBSCAN exhibits higher computational complexity compared to K-means and may be less intuitive to interpret, especially for non-expert users. Furthermore, although parameter selection is less critical than in DBSCAN, it can still significantly influence the clustering results.

12.1.3 Self-Organizing Maps (SOM)

Self-Organizing Maps (SOMs), introduced by Teuvo Kohonen, represent a neural-based approach to clustering and dimensionality reduction. SOMs belong to the family of unsupervised neural networks and are particularly used for the visualization of high-dimensional data [45]

A SOM consists of a two-dimensional grid of neurons, each associated with a weight vector having the same dimensionality as the input data. During the learning phase, for each input sample, the winning neuron, referred to as the Best Matching Unit (BMU), is identified as the neuron whose weight vector is most similar to the sample.

Subsequently, not only the BMU but also the neighbouring neurons in the grid are updated according to a neighbourhood function that decreases over time. This mechanism allows the map to self-organize while preserving the topological relationships among the data.

SOMs are particularly useful for exploratory data analysis, the visualization of complex patterns, and dimensionality reduction. Compared to K-means and HDBSCAN, they provide a richer and more intuitive representation of the relationships among clusters.

On the other hand, SOMs require a longer training phase and careful selection of several hyperparameters, such as grid size and learning rate. Furthermore, the interpretation of the resulting clusters may be less straightforward than with methods purely based on distance or density.

12.1.4 Comparison of Methods

In summary, K-means, HDBSCAN, and SOM represent three different approaches to clustering, each characterized by specific strengths. K-means is well suited to simple and well-structured problems, HDBSCAN excels in complex and noisy contexts, while SOMs are particularly effective for visual analysis and for achieving a global understanding of the data. The choice of the most appropriate algorithm strongly depends on the nature of the dataset and on the objectives of the analysis.

12.2 Clustering and Washing Impact Index

At this stage of the analysis, it becomes necessary to perform an unsupervised clustering of the operating cycles, without imposing a priori the number of clusters to be identified. This choice is essential in order to freely and unconstrainedly observe the evolutionary trend of the machine, allowing the data themselves to reveal the underlying structure. For this purpose, the HDBSCAN algorithm was employed, as it is particularly well suited to identifying clusters with variable density and to handling the presence of noise in the data.

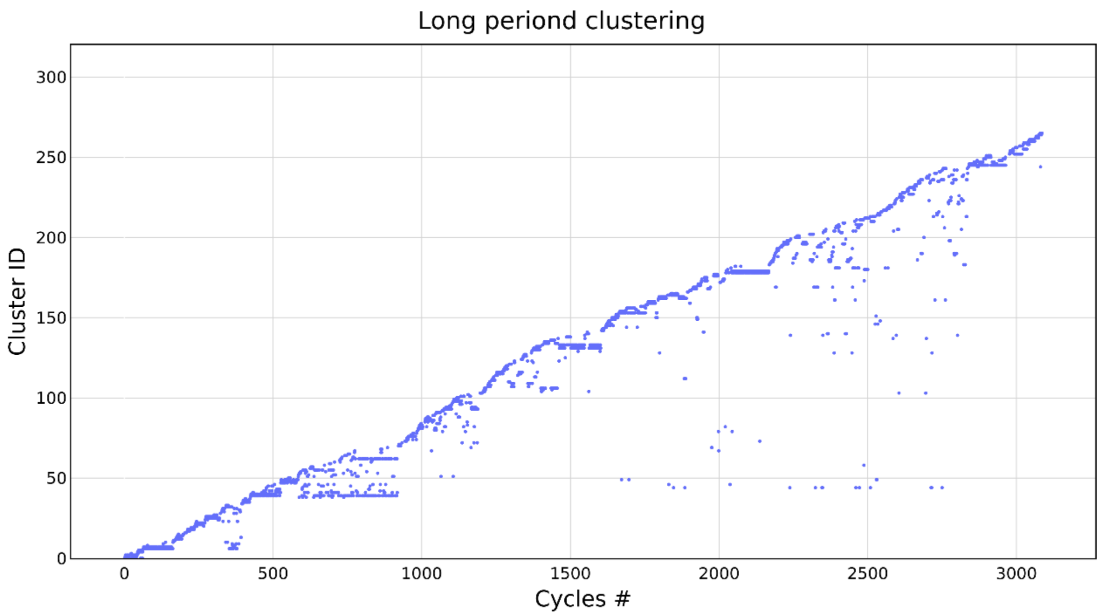

Figure 43 Long period clustering

Figure 43 reports the clustering results over the entire observation period: for each analysed cycle, the corresponding cluster assignment is indicated. It is immediately apparent that the more than 3,000 cycles considered are divided into approximately 280 distinct clusters. In order to improve the readability of the system's temporal evolution, the clusters were relabelled by assigning increasing indices based on the time of first occurrence of each cluster. As a result, clusters associated with the earliest cycles in the dataset are assigned lower indices, while those corresponding to more recent cycles are assigned higher indices.

This reorganization makes it possible to clearly highlight how the machine exhibits a temporal evolution of its behaviour, following an overall trend that appears largely linear. Although this phenomenon is clearly evident from the analysis, it is largely expected, since it is well known that environmental conditions and the characteristics of the processed material are not constant over time. Nevertheless, the question remains open as to whether, by extending the observation period, it would be possible to identify a periodicity in these characteristics or, conversely, the presence of marked discontinuities attributable to known events or to structural changes in the process.

It should be noted that the observation period available for this analysis is limited, for industrial reasons, to the interval between April and September. This temporal window is likely relatively short and, moreover, is positioned exactly between two major plant modifications. Despite these limitations, the analysis shows that, even in the presence of a clear evolutionary trend, some clusters persist over a particularly long-time span. This behaviour may be indicative of specific and recurring operating conditions, such as particularly favourable working regimes, or alternatively of systematic errors or plant configurations that reoccur over time.

The considerations developed thus far are coherent and meaningful but would ideally require validation through direct observation of the machine in the field. For example, it would be highly valuable to correlate the obtained results with reports of operational issues, maintenance interventions, or product quality assessments. However, due to logistical, geographical, and economic constraints, such a comparison is not feasible in the present case.

To address this limitation, the idea was therefore developed to compare the cluster indexing with washing interventions. The objective is to assess whether, and to what extent, washing operations produce a measurable change in system behaviour. In order to make this analysis meaningful, the first step consisted in removing the global trend of the machine, that is, filtering out the evolutionary components associated with slow and structural variations in operating conditions. The aim is to isolate only local variations, attributable to operating epochs between successive washing events.

This operation was carried out by developing a dedicated algorithm for the identification of breakpoints between different operating epochs.

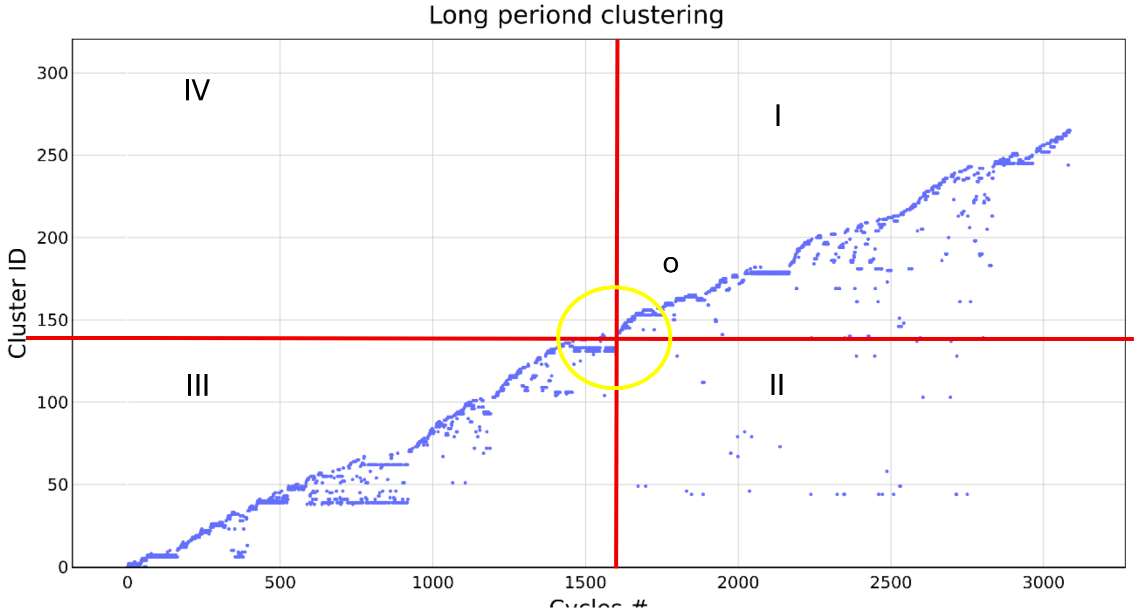

Figure 44 Long period clustering epoch splitting

Referring to Figure 44, the underlying idea of the algorithm consists in identifying points O such that, within a predefined neighborhood of analysis (represented by the yellow circle), the presence of cycles falling in the second and fourth quadrants of a reference system centered at O is minimized. The underlying hypothesis is that, within the same operating epoch, only a limited number of cycles are associated with clusters that are typical of different epochs.

Under this assumption, it becomes relatively straightforward to identify, also from a graphical perspective, distinct operating epochs of the machine, as illustrated in Figure 45.

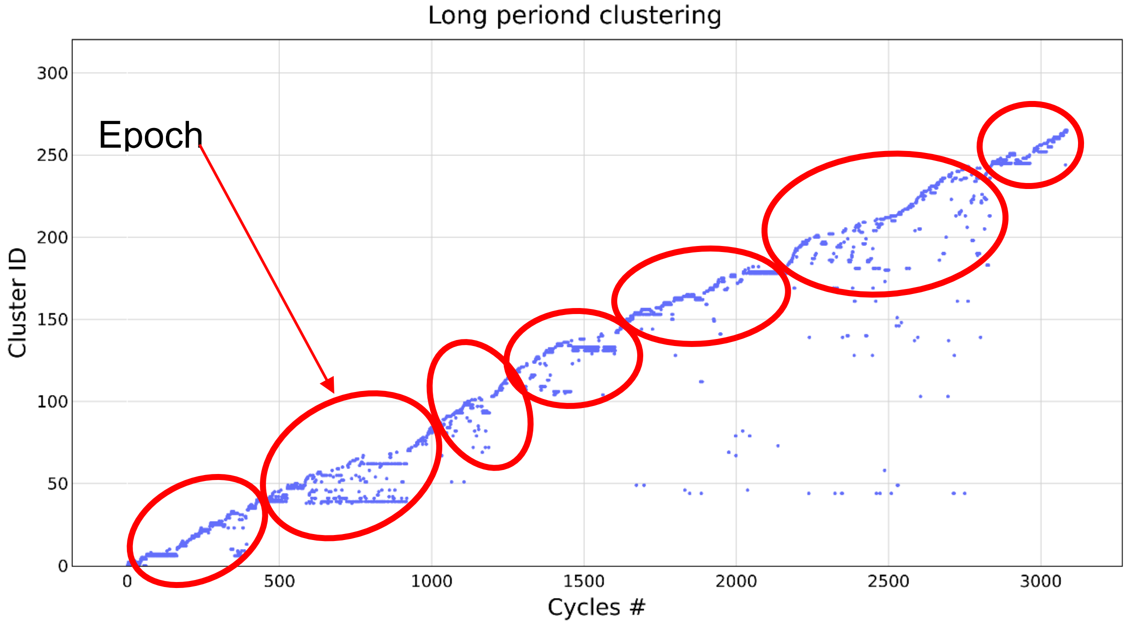

Figure 45 Long period clustering epochs identifying

Within each epoch defined in this manner, a further clustering analysis was subsequently performed, restricted exclusively to the cycles belonging to that specific epoch. In this case, a Self-Organizing Map (SOM)-based approach was adopted, using a configuration with five levels and a single feature, producing an output of size 5×1. This choice allows for an ordered and easily interpretable discretization of the system behavior within each epoch.

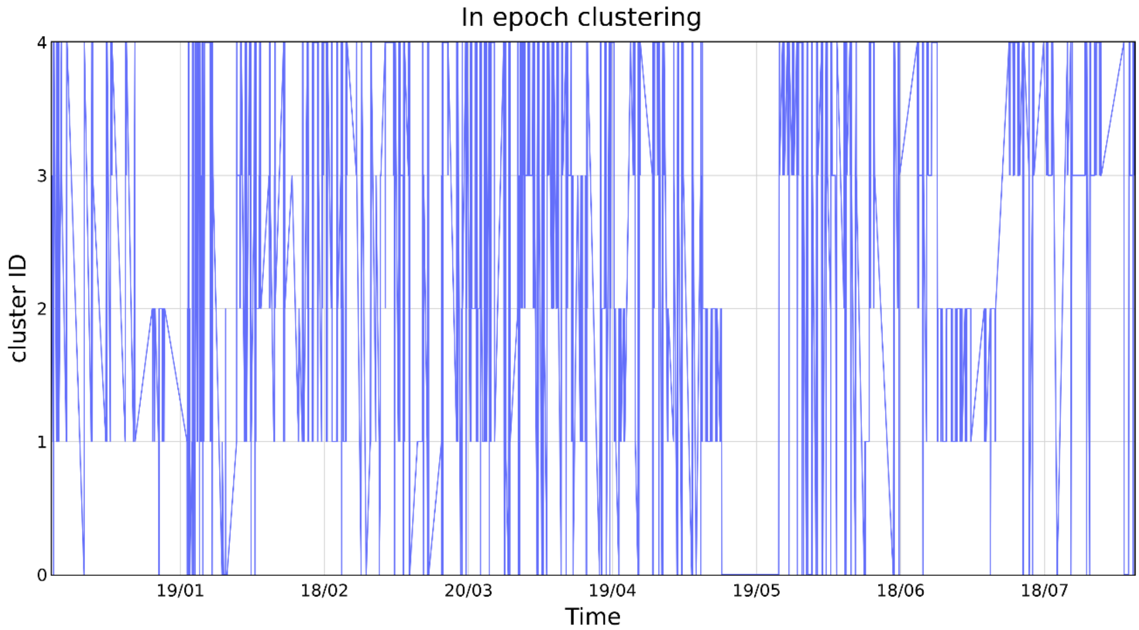

Figure 4623 clustering on epochs

Figure 46 reports the output of the five-level clustering for all epochs, represented as a function of the start time of each cycle. Based on this result, a comparison was performed between the five cycles preceding and the five cycles following each washing event, computing a dissimilarity index according to the Bray–Curtis formulation.

The Bray–Curtis index is a widely used dissimilarity measure for quantifying the degree of difference between two samples based on their quantitative components [46]. Originally introduced in the field of ecology to compare species composition across different communities, this index is now also commonly applied in multivariate data analysis and clustering techniques. Given two vectors

$$\mathbf{x} = (x_1, x_2, \dots, x_n), \mathbf{y} = (y_1, y_2, \dots, y_n),$$

the Bray–Curtis dissimilarity is defined as:

$$BC(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} | x_i - y_i |}{\sum_{i=1}^{n} (x_i + y_i)}.$$

The value of the index ranges between 0 and 1, where 0 indicates complete identity between the two samples and 1 represents maximum dissimilarity. One of the main characteristics of the Bray–Curtis index is its sensitivity to relative differences in abundances rather than absolute differences, making it particularly suitable for non-negative and often sparse data. Moreover, the index is not influenced by joint absences, that is, by components that are zero in both samples. This property distinguishes it from other distance measures and favors its use in contexts where the presence or intensity of variables is more informative than their absence.

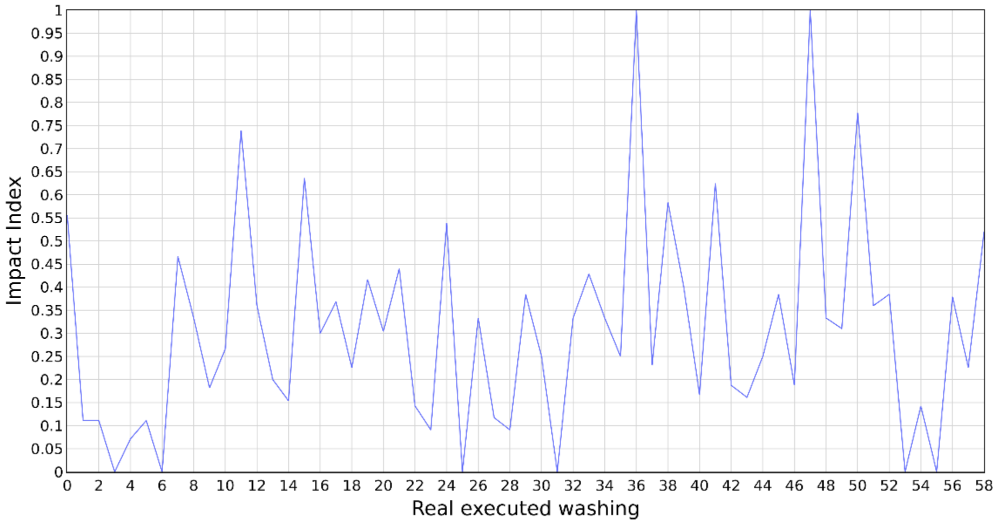

Figure 47 Washing influence factor value

Figure 47 shows the trend of the dissimilarity index for each washing event. In light of this result, it is possible to recall the effectiveness hypothesis formulated at the beginning of the chapter: a washing operation has a greater impact on machine operation when the variation in the measured characteristics before and after the intervention is larger.

Implicit in this approach is the assumption that the aggregated parameters used for clustering contain information that is directly or indirectly dependent on the state of the filter cloth at the time of washing. This assumption constitutes the conceptual basis for using the dissimilarity index as a relative indicator of cloth cleanliness and of the effectiveness of the washing intervention.

12.3 Applicability of the Incidence Index

The index reported in Figure 46 makes it possible to define a measure representing the level of filter cloth fouling, normalized with respect to the available observation period and referenced to the moment at which each washing operation is performed. This index therefore provides a relative assessment of the cloth condition within the analyzed operating context, enabling comparison among washing events that occurred under different conditions but within the same temporal window.

Under the assumption of a constant and repeatable operation of the washing system, and assuming that each intervention is capable of restoring the machine to a stable and comparable clean state regardless of the initial conditions, the evolution of the index can be approximated by a triangular curve with linear interpolation. This curve provides an intuitive representation of the evolution of the cloth condition: immediately after washing, the fouling level is minimal, and it then progressively increases over time until the next intervention.

Figure 48 shows an example of this idealized representation of the cloth condition, obtained from the washing events and the computed impact indices. This construction makes it possible to transform a series of pointwise measurements, derived from the comparison between cycles preceding and following washing, into a continuous variable that describes the evolution of the filter state over time.

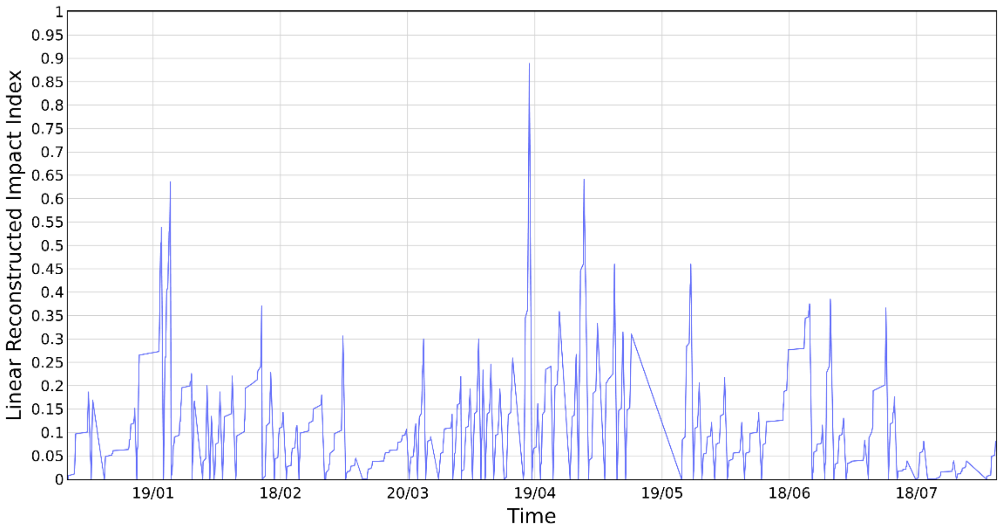

Figure 24 Dirty index interpolated over influence factor

Starting from this representation, a dedicated feedforward neural network was developed with the objective of learning the relationship between the aggregated parameters of a single cycle and the cloth state index defined by the triangular curve. By training the model on historical data, it was possible to reconstruct the temporal evolution of the index using exclusively the information available within the considered cycle.

The result of this estimation is shown by the red curve in Figure 49, which demonstrates the model's good ability to capture and track the evolution of the filter cloth condition over time.

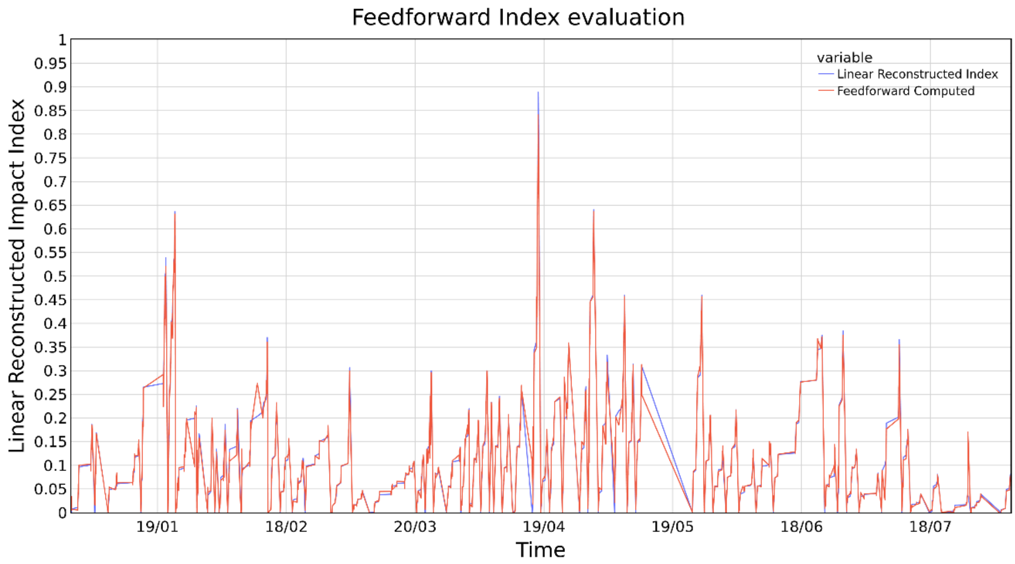

Figure 259 Dirty index interpolated over influence factor predicted over computed

The availability of a model capable of estimating this index solely from aggregated parameters is of particular importance, despite the derived nature of the index itself. Indeed, the direct computation of the reference curve requires knowledge of the cycles following a washing event, typically the five cycles immediately after washing, and therefore presupposes that the washing operation has already occurred. In other words, the *a posteriori* construction of the index assumes knowledge of the future behaviour of the machine, information that is not available in a real operational context.

The ability of the model to provide a coherent estimate of the same index based exclusively on an analysis limited to the single cycle of interest confirms the presence, within the aggregated parameters, of meaningful information related to the state of the filter cloths. This result represents an indirect yet significant indication that the system behaviour contains measurable traces of the fouling level, even in the absence of dedicated sensors or direct performance indicators of the filter.

Starting from this estimated value, it becomes possible to define an operational threshold that provides a decision-support tool capable of triggering washing operations only when they are actually necessary, thus overcoming logic based on fixed temporal thresholds or purely empirical criteria. From a prospective standpoint, this approach could contribute to a more efficient management of the washing process, reducing unnecessary interventions while simultaneously preventing the occurrence of critical operating conditions.

Finally, it should be emphasized that the analysis developed thus far requires experimental validation through direct observation of the machine behaviour. In particular, it is necessary to verify the reliability of the estimated index and to evaluate the impact of potential variations in washing management strategies based on the proposed outputs. Only through this validation phase will it be possible to confirm the actual effectiveness of the approach and its potential applicability in a real industrial context.

13 Validation and Result Analysis

The results obtained from the application of the two developed analysis methodologies make it possible to perform an in-depth performance assessment of the machine, allowing both an evaluation of the effectiveness of the washing operations and a linear, relative estimation of the filter cloth fouling level. In particular, these results provide a quantitative basis for interpreting system behaviour over time and for assessing the impact of routine maintenance operations on the operating conditions of the machine.

The analysis of machine performance trends, based on a physically interpretable index, proves particularly effective in providing results that are directly correlated with quantities having a clear physical meaning. This approach is of significant methodological value, as it allows the observed variations in the performance index to be directly linked to the underlying physical phenomena, in coherence with the assumptions formulated in the initial phase of the analysis. In this context, the primary objective is to identify and quantify the influence of washing operations on a parameter that meaningfully represents the system state.

However, the application of this methodology has highlighted a relevant limitation: the excessive stability of the analysed performance parameter. This stability strongly limits the ability of the index to provide useful short-term indications and makes it necessary to rely on a much longer operational history characterized by greater variability in operating conditions. In other words, in order to observe a significant variation in the index, and thus identify a performance degradation trend, it would be necessary to substantially extend the observation period or modify the system's operating conditions. Ideally, the machine should be operated without washing for a sufficiently long time to allow a measurable degradation in performance to emerge, a condition that is, however, difficult to achieve in industrial practice.

Precisely due to the impossibility of obtaining a physically interpretable indicator that is both sufficiently sensitive and representative of the phenomena of interest, a second methodology was developed, aimed at evaluating the global impact of washing operations on the machine. This second approach stems from the observation that, in practice, washing operations are performed well before an actual degradation of system performance occurs. This evidence suggests that washing is carried out under conditions in which the machine has not yet reached a significant degradation state, with the risk of introducing interventions that are not optimal from an operational and economic standpoint.

In light of this result, the usefulness of having direct and reliable information to guide decision-making related to washing operations becomes immediately apparent. By introducing a fouling threshold beyond which malfunctions or unacceptable operating conditions begin to manifest, it becomes possible to more precisely identify the moments when washing is truly necessary. Such a system would, on the one hand, prevent excessively frequent and potentially unnecessary washing operations, and on the other hand reduce the risk of encountering operational issues due to excessive fouling accumulation [47].

Unlike the first methodology, this second approach is not based on an explicit physical formulation of the analysed phenomenon. Consequently, it does not provide direct physical interpretability of the results, making it necessary to rely on prolonged observation of machine behaviour in order to reliably correlate the model output with real operating conditions and with the occurrence of potential critical events. While this lack of interpretability represents a significant limitation, it simultaneously opens the way to a more flexible analysis that is potentially more sensitive to global system variations.

From the perspective of a thorough validation of the proposed methodologies and future development of the work, it is therefore essential to apply the developed algorithms to datasets acquired under different operating configurations and over longer periods of operation. A comparative analysis of the outputs obtained under these different conditions would allow the robustness and generalizability of the assumptions formulated during the analysis to be assessed. The achievement of stable and coherent results would significantly strengthen the validity of the proposed methodologies and support their potential application in an industrial context.

A further necessary step consists in deliberately varying the machine configuration in a controlled manner, with particular reference to the timing and execution mode of washing operations, and observing the effects on overall system behaviour. With regard to the performance-index-based analysis, it is evident that the study should be conducted in an operating region characterized by a measurable variation of the index itself. This would theoretically require extending the intervals between successive washing operations or redefining the physical KPI as a function of the specific phenomenon to be monitored.

In industrial practice, however, both options are difficult to implement. Extending washing intervals is generally not permitted due to the risk of filter cake detachment failure, while the available data do not reveal clear physical evidence of this phenomenon that could be directly used as an indicator. Similarly, redefining an appropriate physical KPI proves challenging in the absence of reliable and measurable signals associated with the limiting phenomenon [48].

In light of these considerations, the most feasible path for validating the methodologies and further developing the modelling framework consists in deliberately shifting washing operations, with the aim of achieving the same level of impact on the system for each washing event. Under these conditions, it becomes possible to observe whether operational issues, such as cake detachment failure, arise in correspondence with these operating phases. In parallel, it is necessary to construct an observation variable capable of quantitatively representing such issues, providing a synthetic yet meaningful parameter.

Based on these observations, it will then be possible to identify robust correlations between the observation variable and the quantities measured by the system, with the objective of explicitly defining a valid KPI for monitoring the washing-related limiting phenomenon. Such a KPI could ultimately be integrated into the first methodology, allowing the physical interpretability of the performance index to be combined with greater sensitivity to real operating conditions, thereby improving the overall effectiveness of the monitoring and decision-support system.

13.1 Test of direct formulation on historical data

On the other hand, since it is currently not possible to perform a retroactive control of the machine that would allow “in-field” exploration of new high-pressure washing management configurations, it is still possible to leverage the available historical data to estimate what the influence of a different washing strategy—defined through a direct modelling approach—would have been. In other words, under the same historical production conditions (i.e., the same executed cycles), an a-posteriori dirt accumulation profile can be reconstructed and one can simulate when it would have been appropriate to trigger a wash according to an alternative logic compared to the one actually adopted.

To make this reconstruction exercise consistent and repeatable, the following working assumptions were maintained:

- The washing impact index is representative of the dirt level on the cloths (or it is proportional to the amount of material effectively removed by the wash).
- Dirt accumulates progressively cycle by cycle, meaning each cycle contributes to increasing the overall level of residue deposited on the cloths.
- Washing is considered ideally effective, in the sense that after each wash the accumulation is brought back to a reference level (ideally close to "zero" in the reconstructed index). This is a simplifying assumption useful for simulation, but it should be critically discussed during the experimental validation phase.

The first operation carried out concerns estimating the average contribution that each cycle provides to dirt accumulation on the cloths. To this end, a mean residue accumulation gradient was defined and computed between two consecutive washes by using:

- the impact value associated with the wash (which is assumed to "measure" how dirty the system was before the cleaning event), and
- the number of cycles elapsed between the current wash and the previous one.

$$\text{mean gradient} = \frac{\text{washing impact}}{\text{cycles since the last wash}}$$

This quantity can be interpreted as an estimate of the average "dirt increase per cycle" over the considered segment. Although it is a simplification (because it transforms a potentially non-linear process into an average ratio), it enables comparing different periods and verifying whether the accumulation behaviour is sufficiently regular to be described through a threshold-based logic.

This methodology highlights a remarkable stability of the computed gradients across different operating periods (Figure Figure). The observed stability is highly relevant: it suggests that, despite operational variability, cycles affect accumulation in a comparable way even across different analysis windows. In practical terms, the fouling process appears to exhibit a repeatable "structural" component, which is useful for building a control strategy.

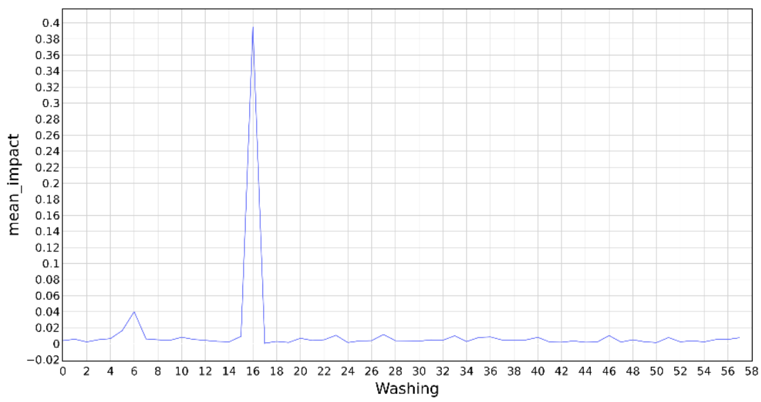

Figure 50 Cycle based mean of washing impact index

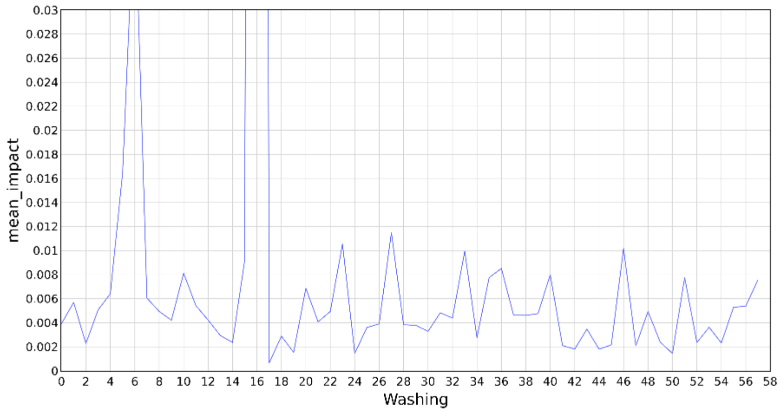

Figure 51 Cycle based mean of washing impact index zoom

The good stability of the gradient across different wash-to-wash intervals significantly increases confidence in the developed methodology. Similar values indicate that, on average, the “cycle” unit produces a similar effect on the fouling level across distinct analysis periods.

However, it is also important to highlight the intrinsic limitation of the approach: by using an average gradient between washes, information about possible non-linearities of the phenomenon is lost.

These effects are not explicitly captured by the mean gradient, but they may remain indirectly visible as fluctuations of the gradient itself across different intervals. At this stage, the linear approximation is reasonable for a first historical reconstruction, although it is not sufficient for a definitive validation.

Once the average impact per cycle (i.e., the accumulation gradient) has been estimated, it becomes possible to reconstruct a dirt index accumulated over time independently of the washes actually performed historically. In practice:

- it is assumed that at each cycle dirt increases by an amount equal to the mean gradient in the relevant reference period,
- these increments are progressively summed to obtain the accumulation trend,
- when the accumulation exceeds a defined threshold, a wash is “called” and the accumulation is brought back to the initial level (consistently with the perfect-wash assumption).

At this point, the control strategy can be formulated as a simple threshold rule: perform a wash when the reconstructed index exceeds a maximum allowable limit. This limit represents a compromise between two typically conflicting objectives:

- reducing the number of washes (less downtime, lower consumption, reduced mechanical stress),
- avoiding degraded operating conditions (risks of detachment failure, inefficiency, process defects, etc.).

In the reconstructed example (Figure 26), a threshold equal to 0.6 was imposed, interpreted as: “the reconstructed dirt/impact index should be kept below 60% of the maximum observable according to this methodology.” In other words, a precautionary setting is adopted, anticipating washing before reaching the potentially most critical operating region.

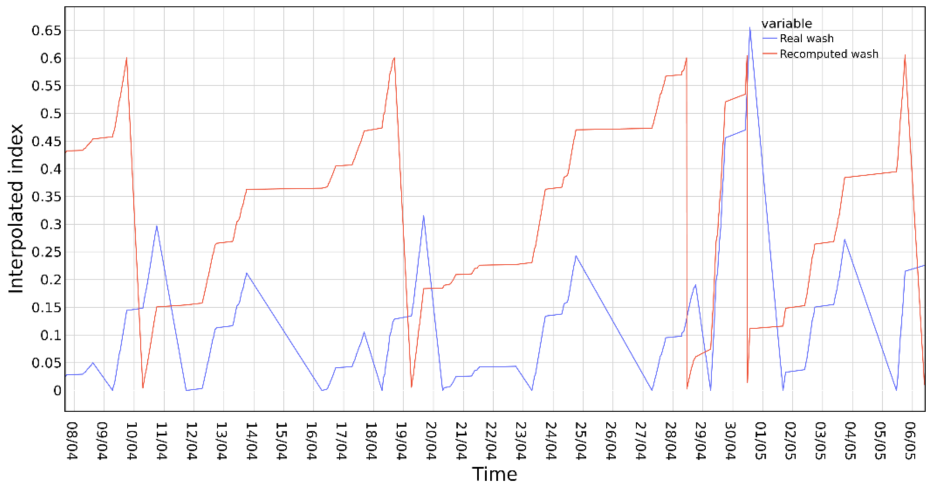

Figure 262 Historical recomputed washing example

The reconstruction made it possible to estimate that, over the analysed historical period of 3016 cycles, it would have been possible to move from 60 washes actually carried out to approximately 25 washes required according to the simulated threshold-based strategy.

Although this result depends on simplifying assumptions, it is significant because it makes explicit the potential benefit in terms of:

- reduction in the number of washes (with possible benefits in machine availability and productivity),
- reduction in washing-related consumption (water, energy, possible additives),
- reduction in wear and stress induced by the high-pressure procedure on the involved components.

At the same time, it is essential to point out that this “theoretical” benefit could be altered if the assumptions are not fully met. In particular, two aspects may bias the result:

- different performance evolution of the machine caused by a different washing pattern,
- non-perfect washing effectiveness,
- non-linear accumulation.

For this reason, in order to perform a reliable evaluation of performance improvement and the operational sustainability of the strategy, an experimental campaign on the machine is still necessary. Such a campaign should foresee that washes are driven according to the modelling output and that the effect is assessed over a sufficiently long horizon (on the order of a few thousand cycles), comparing performance, anomalies, corrective interventions, and functional indicators.

A point of particular interest concerns precisely the strategy used to define the threshold that triggers washing, currently set at 60% for this preliminary evaluation. Looking forward, the threshold should not be chosen only in a “statistical” or “precautionary” manner, but should be tied to a true operational limit beyond which functional issues emerge.

Since it has been observed that it is not possible to robustly appreciate a variation of a generic performance index as a function of washing, it becomes necessary to rely on a direct methodology, i.e., one capable of carrying functional information. A concrete example (already mentioned) is failure to detach the cloths, or the occurrence of abnormal conditions attributable to excessive material build-up.

As no “pure” physical information on the output is currently available (as would be the case when adopting purely physical models with dedicated measurements), it is necessary to experimentally identify the limit beyond which functionality problems occur. In practice, the campaign should:

- identify the critical event (e.g., cake detachment failure, discharge degradation, recurring anomalies),
- map the value (or range) of the index and/or output parameters that precede and accompany the event,
- define an operational limit and a control threshold consistent with an adequate margin.

In the specific case in which the problem is associated with cake detachment failure, the limit should be observed at which a parameter (or a set of output parameters) reliably signals the onset of the issue. This investigation would not only make it possible to prevent malfunctions, but also to maximise the allowable threshold without compromising functionality: in other words, it would be possible to further reduce the number of washes while maintaining a controlled level of risk.

In theory, one may assume that the current control strategy is conservative and that the “theoretical” maximum limit could approach 100% of the reconstructed index; however, a lower factor was used to test the effectiveness of the output and to maintain a safety margin also with a view to its use in an experimental campaign. The final goal is not to push the system to its limit, but to define a robust, repeatable, and operationally defensible control rule.

14 Conclusions

This thesis addressed the application of the Digital Twin paradigm to a complex industrial system characterized by high operational criticality, namely a filter press used for tailings treatment in the mining sector. Filter presses are key assets in mineral processing plants, as they enable the dewatering and solid-liquid separation required to reduce the volume of waste material, improve water recovery, and ensure safe handling of tailings. In the contemporary industrial scenario, mining companies face a context increasingly shaped by stringent environmental requirements, complex logistical constraints, and growing pressure to reduce energy consumption and operational costs. In such a framework, the efficient management of filtration processes represents a decisive factor not only for plant sustainability, but also for overall economic competitiveness and long-term operational resilience.

More specifically, filtration performance has a direct impact on downstream operations, on the achievable water recirculation rate, on the stability of the produced cake, and on the capacity of storage or transport infrastructures. At the same time, filter presses are affected by multiple sources of variability, including changes in feed concentration, mineralogical composition, temperature, slurry rheology, and mechanical wear of components such as plates, seals, pumps, and filter cloths. These factors contribute to making the system highly nonlinear, difficult to monitor, and sensitive to both internal and external disturbances. Consequently, the availability of methodologies capable of providing a deep understanding of the real state of the filter press, and of anticipating degradations or inefficiencies, becomes essential for reducing operational risk and improving plant productivity.

The main objective of this work was therefore to develop a methodological approach capable of improving plant productivity while simultaneously reducing energy consumption, water usage, downtime, and unnecessary maintenance interventions. The guiding idea was that these goals can be achieved through a deeper and more structured understanding of the actual state of the filtration system, obtained by integrating different modelling paradigms within a coherent Digital Twin architecture. Rather than focusing exclusively on the performance of a single predictive model, the thesis aimed at establishing a systematic workflow enabling both interpretability and scalability, with the long-term perspective of supporting industrial implementation.

The central contribution of the thesis lies in the definition and implementation of a Digital Twin not limited to a simple numerical replica of machine behaviour, but conceived as an evolutionary tool for analysis, diagnostics, and decision support. In other words, the developed Digital Twin was not designed as a static representation of an "ideal" machine, but as a dynamic framework able to evolve with the plant, adapt to changes in operating conditions, and incorporate new data sources and models over time. This interpretation is particularly relevant in industrial environments, where the system behaviour changes progressively due to wear, maintenance actions, parameter drift, and variability in raw materials.

In this work, the adopted approach integrates deterministic models based on the physics of the filtration process with statistical and machine learning methodologies. This hybrid perspective makes it possible to exploit physical interpretability and data-driven learning simultaneously. On one side, physics-based models provide an explicit representation of the causal mechanisms underlying filtration dynamics, enabling the estimation of interpretable parameters and the formulation of meaningful diagnostic insights. On the other side, machine learning models can capture complex nonlinear patterns that might be difficult to formalize through analytical equations alone, particularly when the operating conditions vary significantly, and when hidden interactions among variables generate emergent behaviour.

A first significant and enabling result of the work concerns the structuring of the entire data acquisition, cleaning, and organization pipeline. Industrial systems often generate large quantities of raw data in the form of time-series signals acquired by sensors, PLCs, and supervisory control systems. However, such data are typically affected by missing values, outliers, inconsistent sampling rates, sensor drift, and non-stationary noise. Furthermore, raw data rarely follow a structure that is immediately suitable for advanced analyses or model training. As a consequence, the transformation of raw acquired signals into a coherent and reliable dataset represented a crucial step for enabling subsequent modelling tasks.

In particular, the work addressed the segmentation and re-organization of the signals in the operating cycle domain. This decision reflects the intrinsic nature of filter press operations, which are composed of repeated cycles, each characterized by phases such as filling, pressurization, filtration, cake formation, possible washing stages, air blowing, and discharge. By restructuring the dataset at the cycle level, the work established a coherent representation of the process that preserves the physical meaning of the signals while enabling the comparison among cycles under different conditions. The definition of single-cycle data packages made it possible to drastically reduce the dimensionality of the problem while preserving the connection with the underlying physical phenomena. Additionally, the cycle-based representation enabled the extraction of meaningful cycle-level descriptors, such as durations, pressure slopes, filtration rates, and end-of-cycle conditions.

From the perspective of deterministic modelling, the work demonstrated the feasibility of estimating non-directly measurable parameters starting from global process variables measured during operation. In industrial practice, many state indicators are not directly available because their measurement would require invasive sensing, laboratory analysis, or expensive instrumentation. Within this thesis, attention was dedicated to parameters such as the hydraulic resistance of the filter cloths and the resistivity of the solid cake. These variables are strongly linked to the filtration efficiency, to the permeability of the media, and to the structural properties of the produced cake. Their evolution may reflect cloth clogging, wear, blinding phenomena, or changes in feed material characteristics.

The possibility of reconstructing their dynamics on a cycle-by-cycle basis constitutes a valuable element for the complete characterization of the machine operating condition. Indeed, the estimated parameters were found to act as indicators of the system state and its temporal evolution. By monitoring these parameters across cycles, it becomes possible to detect progressive degradation trends, identify transitions between stable and unstable regimes, and support the scheduling of maintenance activities. Additionally, parameter estimation offers a structured interpretation of performance losses: rather than observing a generic decline in throughput, it becomes feasible to attribute such decline to a specific physical cause, such as increasing cloth resistance or unfavourable changes in cake structure.

Alongside deterministic modelling, the thesis introduced advanced numerical aggregation and representation learning techniques, particularly autoencoders and convolutional neural networks. The primary motivation behind this choice was the need to handle high-dimensional cycle representations and capture complex nonlinear behaviours that cannot be fully explained by simplified physical models. While deterministic models provide interpretability, they may rely on assumptions that do not always hold in industrial scenarios, especially when multiphase phenomena, heterogeneous materials, or transient operational practices are involved. Machine learning models can complement this limitation by learning latent patterns directly from the data.

Autoencoders, in particular, were used as unsupervised learning tools for dimensionality reduction. Their ability to compress the high-dimensional cycle signals into a low-dimensional latent space provides compact representations that preserve the most relevant information. These latent features can be interpreted as a global fingerprint of each filtration cycle, capturing subtle relationships among variables such as pressure, flow rate, timing, and system response. Similarly, convolutional neural networks were leveraged to exploit local structures and temporal patterns within the cycle domain. By learning hierarchical features, CNNs can identify characteristic signatures of normal operation and deviations from expected behaviour.

The latent representations learned by these models provided a compact and informative description of the global behaviour of filtration cycles, paving the way for clustering methodologies, similarity analysis, and anomaly detection. In particular, cycle clustering can support the identification of operational regimes, distinguish between different feed conditions, and segment the process into families of similar cycles. Similarity analysis enables the retrieval of cycles comparable to a current one, which can be useful for diagnostic reasoning and decision support. Anomaly detection, on the other hand, provides the foundation for early warning systems capable of recognizing abnormal patterns that may precede failures, quality losses, or excessive energy consumption.

A further significant contribution is represented by the definition of performance indicators (KPIs) and their integration into predictive and optimization models. Industrial decision-making requires actionable metrics that translate raw signals and internal model parameters into meaningful measures of efficiency and productivity. The work therefore defined KPIs related to throughput, cycle time, energy usage, filtration effectiveness, and machine availability. These indicators act as a bridge between modelling outputs and real operational objectives, making the Digital Twin relevant for plant-level decision processes.

Moreover, the integration of KPIs into forecasting frameworks showed how it is possible to anticipate the evolution of plant performance and support more informed operational decisions. Forecasting models, including hybrid approaches that combine physical knowledge and data-driven prediction, were evaluated to estimate the future trend of performance measures and key parameters. This capability is particularly valuable in a process such as filtration, where performance degradation may develop gradually but can also exhibit sudden changes due to variability in feed. Predictive models can assist operators by providing early insights on potential performance losses, enabling proactive adjustments of control variables such as pressure profiles, cycle duration settings, or feed strategies.

In this sense, the developed Digital Twin can be regarded as a concrete enabler of dynamic optimization of process parameters. Instead of relying on fixed operational setpoints, the system can support adaptive adjustments based on the actual and predicted state of the machine. This perspective aligns with modern industrial paradigms such as predictive maintenance, adaptive control, and continuous improvement, where decisions are not solely based on experience but also supported by systematic data-driven insights.

Overall, the obtained results confirm the validity of the proposed approach and highlight its industrial potential. The integration of physics-based and data-driven models enabled the development of a Digital Twin capable of providing a deeper understanding of filtration dynamics, detecting meaningful state indicators, and supporting prediction and optimization tasks. While acknowledging the simplifications introduced and the limitations imposed by the quality of the available data, the work demonstrates how the hybrid methodology can overcome many of the limitations typical of the two approaches when considered individually. In fact, deterministic models alone may suffer from inadequate representational power under variable conditions, while purely data-driven models may lack interpretability and generalization when training data are limited or biased. Their combination therefore provides robustness, interpretability, and modelling flexibility.

The Digital Twin thus emerges not only as a tool for a posteriori analysis, but as an active component of the production system. It is capable of supporting operational decisions, guiding maintenance scheduling, enabling early detection of inefficiencies, and ultimately contributing to more sustainable and cost-effective operations. The developed framework can be extended to additional monitoring variables, connected to real-time data streams, and integrated into control room dashboards for continuous usage.

Future developments include the extension of the model to multiple machines, enabling fleet-level comparisons and cross-learning among similar filter presses operating under different conditions. Additional directions involve real-time integration with plant control systems, which would allow the Digital Twin to operate continuously as a monitoring and advisory tool. Furthermore, the refinement of optimization algorithms from a multi-objective perspective represents a key opportunity, as industrial filtration requires balancing competing objectives such as productivity, cake dryness, energy usage, cloth lifetime, and maintenance cost.

In conclusion, this thesis lays the foundations for a mature and industrially relevant use of the Digital Twin paradigm in the mining filtration sector. By combining methodological rigor, physical insight, and modern machine learning techniques, the work contributes to the development of more efficient, sustainable, and intelligent plants. The proposed approach can support mining companies in addressing emerging sustainability requirements, improving operational performance, and building a data-driven culture in process management and maintenance strategies.

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Tesi di dottorato finanziata dall'Unione europea- Next Generation EU, Missione 4, componente 2 "Dalla Ricerca all'Impresa" - Investimento 3.3 "Introduzione di dottorati innovativi che rispondono ai fabbisogni di innovazione delle imprese e promuovono l'assunzione dei ricercatori dalle imprese".

References

[1] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, “Digital twin-driven product design, manufacturing and service with big data,” *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9–12, pp. 3563–3576, Feb. 2018, doi: 10.1007/s00170-017-0233-1.

[2] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital Twin in manufacturing: A categorical literature review and classification,” *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018, doi: 10.1016/j.ifacol.2018.08.474.

[3] A. Fuller, Z. Fan, C. Day, and C. Barlow, “Digital Twin: Enabling Technologies, Challenges and Open Research,” *IEEE Access*, vol. 8, pp. 108952–108971, 2020, doi: 10.1109/ACCESS.2020.2998358.

[4] Q. Qi and F. Tao, “Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison,” *IEEE Access*, vol. 6, pp. 3585–3593, 2018, doi: 10.1109/ACCESS.2018.2793265.

[5] M. Grieves and J. Vickers, “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems,” in *Transdisciplinary Perspectives on Complex Systems*, Cham: Springer International Publishing, 2017, pp. 85–113. doi: 10.1007/978-3-319-38756-7_4.

[6] M. M. Rathore, S. A. Shah, D. Shukla, E. Bentafat, and S. Bakiras, “The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities,” *IEEE Access*, vol. 9, pp. 32030–32052, 2021, doi: 10.1109/ACCESS.2021.3060863.

[7] R. Zhang, F. Wang, J. Cai, Y. Wang, H. Guo, and J. Zheng, “Digital twin and its applications: A survey,” *The International Journal of Advanced Manufacturing Technology*, vol. 123, no. 11–12, pp. 4123–4136, Dec. 2022, doi: 10.1007/s00170-022-10445-3.

[8] J. Guo, N. Zhao, L. Sun, and S. Zhang, “Modular based flexible digital twin for factory design,” *J Ambient Intell Humaniz Comput*, vol. 10, no. 3, pp. 1189–1200, Mar. 2019, doi: 10.1007/s12652-018-0953-6.

[9] H. Tang, D. Chen, Y. Zhang, B. Xu, U. A. Bhatti, and M. Yu, “Bidirectional Interaction Techniques Based on Device Digital Twin Model,” in *2024 39th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, IEEE, Jun. 2024, pp. 132–138. doi: 10.1109/YAC63405.2024.10598558.

[10] P. Almasan *et al.*, “Digital Twin Network: Opportunities and Challenges,” Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.01144

[11] L. Liu, X. Zhang, X. Wan, S. Zhou, and Z. Gao, “Digital twin-driven surface roughness prediction and process parameter adaptive optimization,” *Advanced Engineering Informatics*, vol. 51, p. 101470, Jan. 2022, doi: 10.1016/j.aei.2021.101470.

[12] D. Van Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong, “Edge Intelligence-Based Ultra-Reliable and Low-Latency Communications for Digital Twin-

Enabled Metaverse," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1733–1737, Aug. 2022, doi: 10.1109/LWC.2022.3179207.

[13] M. Manngård, W. Lund, and J. Björkqvist, *Using Digital Twin Technology to Ensure Data Quality in Transport Systems*. 2020.

[14] R. K. Soman, K. Farghaly, G. Mills, and J. Whyte, "Digital twin construction with a focus on human twin interfaces," *Autom Constr*, vol. 170, p. 105924, Feb. 2025, doi: 10.1016/j.autcon.2024.105924.

[15] M. Krupas, E. Kajati, C. Liu, and I. Zolotova, "Towards a Human-Centric Digital Twin for Human–Machine Collaboration: A Review on Enabling Technologies and Methods," *Sensors*, vol. 24, no. 7, p. 2232, Mar. 2024, doi: 10.3390/s24072232.

[16] M. Paziraei and A. Motamedi, "A survey of visualization techniques for digital twins," Jan. 2024.

[17] A. Löcklin, T. Jung, N. Jazdi, T. Ruppert, and M. Weyrich, "Architecture of a Human-Digital Twin as Common Interface for Operator 4.0 Applications," *Procedia CIRP*, vol. 104, pp. 458–463, 2021, doi: 10.1016/j.procir.2021.11.077.

[18] R. J. . Jewell and A. B. . Fourie, *Paste and thickened tailings : a guide*. Australian Centre for Geomechanics, University of Western Australia, 2015.

[19] Ladislav. Svarovsky, *Solid-liquid separation*. Butterworth-Heinemann, 2000.

[20] K. C. O'Leary, J. D. Eisnor, and G. A. Gagnon, "Examination of plant performance and filter ripening with particle counters at full-scale water treatment plants," *Environ Technol*, vol. 24, no. 1, pp. 1–9, Jan. 2003, doi: 10.1080/09593330309385529.

[21] F. Kaswalder, D. Cavalli, and A. Paglianti, "Tailings dewatering by pressure filtration: process optimisation and design criteria," 2018, pp. 427–438. doi: 10.36487/ACG_rep/1805_35_Kaswalder.

[22] P. Agarwal and S. McNeill, "Real-Time Cleaning of Time-Series Data for a Floating System Digital Twin," in *Offshore Technology Conference*, OTC, Apr. 2019. doi: 10.4043/29642-MS.

[23] A. de Brauwere, F. De Ridder, R. Pintelon, J. Meersmans, J. Schoukens, and F. Dehairs, "Identification of a periodic time series from an environmental proxy record," *Comput Geosci*, vol. 34, no. 12, pp. 1781–1790, Dec. 2008, doi: 10.1016/j.cageo.2008.01.007.

[24] H. Wang and J.-A. Wang, "Data Processing Package for Cyclic Integrated Reversible Bending Fatigue Testing," Oak Ridge, TN (United States), Jul. 2021. doi: 10.2172/1814278.

[25] Ladislav. Svarovsky, *Solid-liquid separation*. Butterworths, 1977.

[26] B. F. Ruth, G. H. Montillon, and R. E. Montonna, "Studies in Filtration - I. Critical Analysis of Filtration Theory," *Ind Eng Chem*, vol. 25, no. 1, pp. 76–82, Jan. 1933, doi: 10.1021/ie50277a018.

[27] M. C. Oliveira and M. Iten, "Modelling of industrial water circuits with a customised Modelica library," *Appl Therm Eng*, vol. 169, p. 114840, Mar. 2020, doi: 10.1016/j.applthermaleng.2019.114840.

[28] R. Wakeman, “The influence of particle properties on filtration,” *Sep Purif Technol*, vol. 58, no. 2, pp. 234–241, Dec. 2007, doi: 10.1016/j.seppur.2007.03.018.

[29] R. Holdich, E. Tarleton, and S. Willmer, “Compressible cake filtration,” IChemE, 1994. [Online]. Available: https://dspace.lboro.ac.uk/

[30] B. S. Dobronets and O. A. Popova, “Improving reliability of aggregation, numerical simulation and analysis of complex systems by empirical data,” in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, May 2018. doi: 10.1088/1757-899X/354/1/012006.

[31] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” *Science (1979)*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[32] P. V. Ca, L. T. Edu, I. Lajoie, Y. B. Ca, and P.-A. M. Ca, “Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pascal Vincent Hugo Larochelle Yoshua Bengio Pierre-Antoine Manzagol,” 2010.

[33] G. Bebis and M. Georgiopoulos, “Feed-forward neural networks,” *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, Oct. 1994, doi: 10.1109/45.329294.

[34] P. Almasan *et al.*, “Digital Twin Network: Opportunities and Challenges,” Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.01144

[35] Y. Wang, H. Yao, and S. Zhao, “Auto-encoder based dimensionality reduction,” *Neurocomputing*, vol. 184, pp. 232–242, Apr. 2016, doi: 10.1016/j.neucom.2015.08.104.

[36] T. Takase, S. Oyama, and M. Kurihara, “Effective neural network training with adaptive learning rate based on training loss,” *Neural Networks*, vol. 101, pp. 68–78, May 2018, doi: 10.1016/j.neunet.2018.01.016.

[37] J. Palmer, “Simulation of tailings filtration performance,” 2020. doi: 10.36487/ACG_repo/2052_90.

[38] W. Xiangxue, X. Lunhui, and C. Kaixun, “Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN,” *Arab J Sci Eng*, vol. 44, no. 4, pp. 3043–3060, Apr. 2019, doi: 10.1007/s13369-018-3390-0.

[39] H. Abbasimehr, M. Shabani, and M. Yousefi, “An optimized model using LSTM network for demand forecasting,” *Comput Ind Eng*, vol. 143, p. 106435, May 2020, doi: 10.1016/j.cie.2020.106435.

[40] R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, “A hybrid CNN-LSTM model for typhoon formation forecasting,” *Geoinformatica*, vol. 23, no. 3, pp. 375–396, Jul. 2019, doi: 10.1007/s10707-019-00355-0.

[41] N. Ghosh and M. Belkin, “A Universal Trade-off Between the Model Size, Test Loss, and Training Loss of Linear Predictors,” *SIAM J Math Data Sci*, vol. 5, no. 4, pp. 977–1004, Dec. 2023, doi: 10.1137/22M1540302.

[42] R. J. G. B. Campello, D. Moulavi, and J. Sander, “Density-Based Clustering Based on Hierarchical Density Estimates,” 2013, pp. 160–172. doi: 10.1007/978-3-642-37456-2_14.

[43] P. Almasan *et al.*, "Digital Twin Network: Opportunities and Challenges," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.01144

[44] G. Stewart and M. Al-Khassaweneh, "An Implementation of the HDBSCAN* Clustering Algorithm," *Applied Sciences*, vol. 12, no. 5, p. 2405, Feb. 2022, doi: 10.3390/app12052405.

[45] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans Neural Netw*, vol. 11, no. 3, pp. 586–600, May 2000, doi: 10.1109/72.846731.

[46] E. W. Beals, "Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data," 1984, pp. 1–55. doi: 10.1016/S0065-2504(08)60168-3.

[47] A. Löcklin, M. Müller, T. Jung, N. Jazdi, D. White, and M. Weyrich, "Digital Twin for Verification and Validation of Industrial Automation Systems – a Survey," in *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2020, pp. 851–858. doi: 10.1109/ETFA46521.2020.9212051.

[48] C. Tudu, S. Sharma, and D. Kumar, "Computational Modeling and Digital Twin Technologies in Medical Device Development," *Biomedical Materials & Devices*, Jul. 2025, doi: 10.1007/s44174-025-00469-0.