Reliability Engineering and System Safety 272 (2026) 112574

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Human-Hardware-in-the-Loop simulations for systemic resilience assessment in cyber-socio-technical systems

Francesco Simone [a,*], Marco Bortolini [b], Giovanni Mazzuto [c], Giulio Di Gravio [a], Riccardo Patriarca [a]

[a] Sapienza University of Rome, Department of Mechanical and Aerospace Engineering, Via Eudossiana 18, 00184 Rome (RM), Italy
[b] Alma Mater Studiorum – University of Bologna, Department of Industrial Engineering, Viale del Risorgimento 2, 40136, Bologna (BO), Italy
[c] Università Politecnica delle Marche, Department of Industrial Engineering and Mathematical Sciences, Via Brecce Bianche, 60100, Ancona (AN), Italy

ARTICLE INFO

Keywords:
Resilience engineering
Operations management
Cyber-attacks
Cyber-physical systems
Motion capture
Complexity

ABSTRACT

Modern industrial systems require updated safety management approaches, as the tight interplay between cyber-physical, human, and organizational factors has driven their processes toward increasing complexity. In addition to dealing with known risks, managing system resilience acquires great value to address complex behaviours pragmatically. This manuscript starts from the System-Theoretic Accident Model and Processes (STAMP) as a modelling initiative for such complexity. The STAMP can be natively integrated with simulation-based approaches, which however fail to realistically represent human behaviours and their influence on the system performance. To overcome this limitation, this research proposes a Human-Hardware-in-the-Loop (HHIL) modelling and simulation framework aimed at supporting a more realistic and comprehensive assessment of systemic resilience. The approach is tested on an experimental oil and gas plant experiencing cyber-attacks, where two personas of operators (experts and novices) work. The obtained results provide a means to quantitatively assess how variations in operators' behaviours impact the overall system performance, offering insights into how resilience should be understood and implemented in complex socio-technical systems at large.

1. Introduction

Managing safety in industrial systems has become an increasingly challenging task over time. If on one hand, a growing awareness on safety-related themes has provided researchers and practitioners with a wide range of tools and methodologies to identify and to address operational risks; on the other, the continuous technological advancement has made industrial systems progressively more complex. Over the years, industrial processes have undergone a significant transformation. They evolved from predominantly manual, craft-like operations, where workers used machines as tools, to highly automated configurations in which intelligent machines can now perform more and more complicated tasks. This shift has marked a blurred transition from a simple coexistence between humans and machines to scenarios of close collaboration, where both entities work together to enable the system to achieve its goals [1]. In addition, the relevance of organizational factors related to such operational processes has been acknowledged. As such, over the last two decades, scholars increasingly emphasized the importance of adopting a systemic perspective for managing safety, suggesting the integration of human, technical, and organizational dimensions alike [2,3]. As a result, safety management has become closely tied to the concept of socio-technical systems (STSs) [4], i.e., systems made of people, machines, and organizations, all tightly and dynamically interacting to ensure that the system fulfils its intended purpose.

The mentioned evolution of industrial systems pointed towards a growing digitalization of STSs. The "technical" dimension evolved into two interconnected facets: (i) a physical part, made up of tangible equipment, and (ii) a digital part, involving all the data and the information exchanged within the cyberspace. This distinctions becomes tangible as cyber-physical systems (CPSs) [5] have become increasingly prevalent in industrial environments: while their adoption enhances performance by optimizing processes and enabling better coordination among activities, they also introduce a range of new risks. Notably, digital failures can now result in tangible, real-world physical consequences [6]. On this premises, the concept of STS has been recently expanded to highlight the additional complexities introduced by the

* Corresponding author.
E-mail address: francesco.simone@uniroma1.it (F. Simone).

https://doi.org/10.1016/j.ress.2026.112574
Received 8 September 2025; Received in revised form 26 January 2026; Accepted 8 March 2026
Available online 10 March 2026

adoption of CPSs: Patriarca et al. [7] introduced the notion of cyber-socio-technical systems (CSTS), where the "cyber" prefix is meant to emphasize how the novel dual nature of the technical dimension increases the complexity related to the "social" dimension of STSs: the social components must now interact not only with physical technical elements but also with their cyber counterparts.

1.1. The System-Theoretic Accident Model and Processes (STAMP) and its related techniques

In this complex scenario, among the various systemic models and methods, the System-Theoretic Accident Model and Processes (STAMP) and its related techniques [3] stood out. The STAMP has been extensively employed not only in scientific research but also in numerous industrial case studies, demonstrating both a strong theoretical foundations, and the practical effectiveness of the resulting analyses [8].

From a theoretical standpoint, the STAMP combines principles from both systems theory and control theory to offer a novel perspective on systems safety. Specifically, the adoption of a system thinking viewpoint embedded in STAMP incorporates the effects on, and the mutual dependencies among the system's elements. Meanwhile, drawing from control theory principles, the STAMP views the system as a collection of interrelated elements that must be maintained in a dynamic equilibrium through feedback and control loops. According to the STAMP, whatever safety issue arises when such control mechanisms result inadequate or fail. The whole rationale behind the STAMP and its related techniques builds on this premises: the existence of unsafe (or unsecure) control actions (UCAs). Indeed, analysing a STAMP model turns out in identifying which control actions may result (or resulted) inadequate and how, to eventually define the safety constraints to ensure this will not happen (again). The System-Theoretic Process Analysis (STPA) and the System-Theoretic Process Analysis for Security (STPA-Sec) formalize this process in a proactive manner, while the Causal Analysis based on Systems Theory (CAST) applies a similar approach, but retrospectively.

1.1.1. STAMP-based quantification approaches proposed in literature

Patriarca et al. [8] provided a comprehensive review of the usage of STAMP and its related techniques. However, the field of application of this system-theoretic modelling, is rapidly evolving, with an acknowledgement over the past few years (i.e., 2022, onwards) of the benefits deriving from enriching typical STAMP-based qualitative findings with quantitative reasoning. For instance, Ebrahimi et al. [9] integrated STAMP with fuzzy logic to quantify the interrelationships among control levels, eventually producing interpretable "cause-effect" weights that extend the STAMP beyond its qualitative bounds. However, the quantitative results remained largely dependent on expert judgements and static assessments of interdependencies, without fully freeing the analysis from subjectivity. Similarly, Sun et al. [10] demonstrated that combining the STAMP with a cascading failure propagation model could enable dynamic risk profiling in real time. The approach was effectively used to quantify how faults propagate across a system, yet it required assumptions about the propagation mechanisms to consider. Additionally, the STAMP and the STPA have been combined with event-tree analysis [11], and Bayesian networks [12] to estimate the probabilities of accidents occurrences, leveraging probabilistic inference to produce interpretable risk scores. Also in this case, the validity of the results was sensitive to the quality of expert-elicited inputs, and it did not fully capture the system dynamics over time. A similar example can be found in the research by Liao et al. [13], where STPA was paired with Bayesian networks and Success Likelihood Index Method (SLIM) to model human error likelihood, but, again, results relied on expert-estimated probabilities and validation data. In contrast, approaches that couple STAMP-related techniques with simulations show particular advantages when it comes to the objectivity and scalability of results, permitting to iterate and test over large sets of parameters settings. Examples are the STPA-SPN framework by Bensaci et al. [14] which used Petri nets and Monte Carlo simulation to directly estimate collision frequencies in multi-robot systems, or the STPA-Sec/S method by Simone et al. [15] which enabled the resilience assessment through direct modelling of system dynamics. These simulation-based methods produced empirical distributions of outcomes under varied scenarios and operational conditions, capturing cascading effects, and time dependencies that static inference approaches cannot. While expert judgment remains essential in defining the safety control structure (SCS) and the UCAs, pairing these models with simulations allows for far richer quantitative insights. Indeed, the reviewed literature indicates that while probabilistic and static modelling techniques offer valuable results, integrating STAMP with simulation techniques provides a particularly promising path forward [16].

On these premises, this paper builds up on simulations approaches presenting a method to enhance the quantitative analysis of STAMP results, while preserving its systemic take on CSTSs.

1.2. Objective of this research

Although simulations represent a compelling solution to quantify and explore the results of a STAMP analysis, especially in complex and dynamic environments, two main challenges must be acknowledged [17]. A first difficulty in simulating CSTSs lies in the inherent simplification that modelling entails. Models are, by design, abstractions of reality and they cannot capture the full complexity of real-world systems. This limitation may be particularly problematic when dealing with CSTSs and cyber failures that can propagate non-linearly, and in context-dependent ways, often with emergent and unpredictable outcomes, eventually affecting the accuracy and interpretability of obtained results. A second challenge relates to the fact that in CSTSs the presence of human agents has an active role in the system. However, unlike technical components', the human behaviour is highly variable and influenced by cognitive, social, and contextual factors that are difficult to formalize into a model.

Building on these premises, this paper addresses the following research question: *how to quantitatively assess the results of a STAMP-based analysis for a CSTS leveraging Human-Hardware-in-the-Loop (HHIL) simulations?*

With HHIL, we here refer to a simulation approach that integrates both cyber-physical components of the system (i.e., hardware) and real humans interacting with those components. By including human agents directly in the simulation loop, this approach aims to overcome the aforementioned limitations, especially in those scenarios where human decision-making is tightly intertwined with the management of cyber failures. Indeed, this research primarily focuses on CSTSs' cyber disruptions, and the proposed method draws inspiration from – and extends – the STPA-Sec/S approach introduced by Simone et al. [15]. The method is instantiated and tested through a case study involving an experimental plant that reproduces an oil and gas facility. The experimental setup allowed for a controlled observation of the system behaviours and human responses in the presence of simulated cyber-attacks, permitting to argue the effectiveness of the HHIL approach in quantifying the STPA-Sec/S results.

2. Method

This section outlines the approach used in this study. It begins with a brief overview of the STPA-Sec/S method, which serves as the foundation for this research. Subsequently, the necessary adaptations to the simulation model are discussed in order to enable the HHIL simulations.

2.1. System-Theoretic Process Analysis for Security with Simulations (STPA-Sec/S)

STPA-Sec/S is a methodological advancement of the traditional STPA approach and its cybersecurity-oriented extension, i.e., STPA-Sec.

The method integrates system-theoretic reasoning with simulation-based modelling to overcome the largely qualitative nature of existing STPA-Sec applications. While STPA-Sec effectively identifies unsafe or unsecure control actions within a SCS, it typically offers limited insight into how these vulnerabilities evolve dynamically during cyber disruptions. Traditional STPA-Sec studies often incorporate wargaming to provide a practical perspective on the results of the analysis. However, even in this case, there is not a direct quantification linkage available. This aspect becomes particularly important when characterizing the cyber-resilience performance of a CSTS, i.e., its ability to anticipate, withstand, recover from, and adapt to cyber disruptions [18]. STPA-Sec/S addresses this gap by enabling the quantitative and structured exploration of system behaviour under cyber-attacks through simulations.

While the interested reader is invited to analyse the manuscript by Simone et al., [15], it is here provided a synthesis of the core methodological steps of STPA-Sec/S.

The process begins by defining a system intended mission, along with its boundaries, losses, hazards, and constraints, which collectively establish the scope of the analysis. A hierarchical SCS is then modelled to capture the interactions among controllers, actuators, sensors, and controlled processes. Within this structure, potential UCAs are subsequently identified. Specifically, a control action may turn inadequate and identify an UCA when: (i) not providing a feedback or a control action leads to an hazard; (ii) providing a feedback or a control action leads to an hazard; (iii) providing a feedback or a control action too early, too late, in wrong order, or with an inappropriate application leads to an hazard; (iv) a feedback or a control action provided are stopped too late, or too soon, leading to an hazard. As a result, the identified UCAs highlight functional roles that may serve as vectors for cyber-induced failures. The SCS serves as a blueprint for constructing a digital model in a simulation environment, where key system elements are represented through state variables capable of reproducing the dynamics relevant to the scope defined in the first step. Appropriate resilience metrics (ranging from physical or cyber variables to service-level or mission-specific indicators) are then engineered to quantify the system's response to disruptions. These metrics are defined by the analyst, but their identification is guided by the systemic perspective established in the first step. In practice, this means selecting measurable and accessible process parameters, or appropriate combinations of them, that best reflect the system's ability to withstand, respond to, and recover from cyber disturbances. Attack scenarios to be simulated are derived by translating the identified UCAs into perturbations introduced during simulation runs. Accordingly, comparing system performance under these scenarios with those at nominal behaviour, the STPA-Sec/S permits the computation of resilience metrics enabling the assessment of the system resilience to specific cyber-attacks. The simulation runs can extend beyond simple deterministic input-output computations, instead leveraging systematically stochastic simulations to gain a systemic understanding of the CSTS complexity.

2.2. Toward Human-Hardware-in-the-Loop (HHIL) simulations

To increase the realism of the simulation executed in the final phase of STPA-Sec/S, this research integrates HHIL logics. The HHIL simulation architecture is directly grounded in the system's SCS, as defined by the STPA-Sec/S. However, in this extended use, the SCS supports both the identification of key state variables (i.e., those affected by control actions or shared through feedback), and it serves to map the different modes of interaction between human and technological agents to be captured in the HHIL simulation model. In this context, three distinct types of human-technology interaction can be identified from a simple human-automation SCS schema (cf. Fig. 1):

Automated process control (human monitoring). In this configuration, the human operator is excluded from directly controlling the physical process, and they can only monitor the process parameters. This occurs whenever the operator chooses to not to exercise their control capabilities, but instead delegating all the process control to the automated systems;
Human indirect process control (human empowering). In this case, the human operator influences the process by modifying the logic of the automated control system. Although they are actively involved, the operator's actions are still mediated by the automation, and they do not directly interact with the physical process;
Human direct process control (human exclusivity). Here, the human operator exerts manual control over the process by directly interacting with physical actuators and reading measurements from the equipment sensors, without relying on any automation intermediaries.

The three modes of interaction can occur multiple times and in varying sequences throughout the process operations. Thus, the objective is to develop a simulation model architecture that can dynamically adapt to capture each of these interaction modes whenever they arise [19]. Based on this premise, the HHIL simulation model incorporates all three modes through an architectural design such as the one illustrated in Fig. 2. The architecture includes a physical process (i.e., "Physical process" block, cf. Fig. 2) which must be tangibly reproduced in the real world, for example through a testbed or a mock-up simulator. The existence of a real physical process is critical to enabling the interactions with the human operator (i.e., "Human operator" block, cf. Fig. 2). Since the system under study is a CSTS, some level of automation must also be incorporated. Accordingly, the architecture also considers sensors (i.e.,

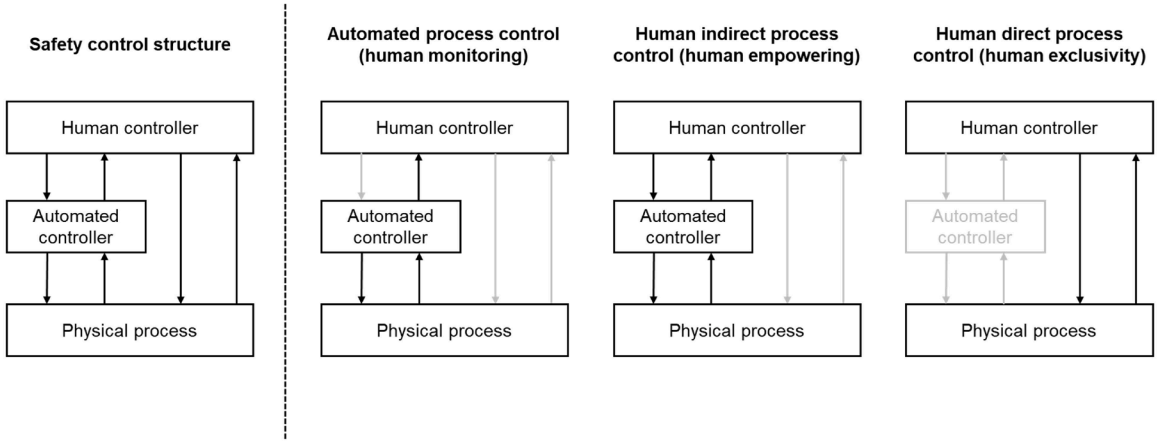

Fig. 1. An exemplary SCS of a CSTS (left), and the three types of human-technology interactions that may occur during the system operations (right).

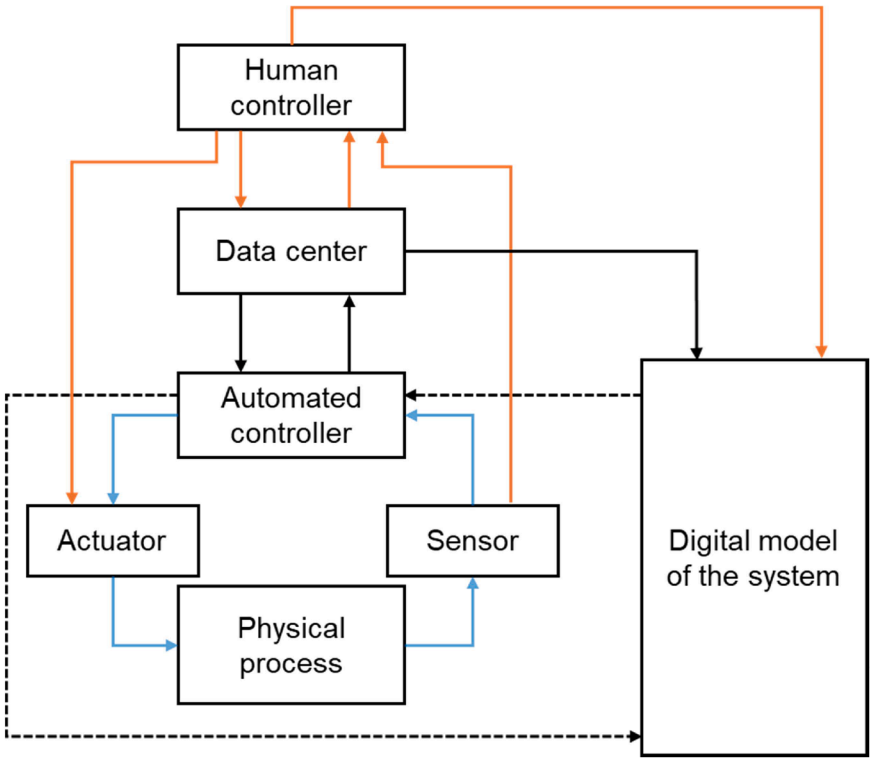

Fig. 2. HHIL simulation model architecture. Solid black arrows represent interaction links between technological elements that are always active. Dashed black arrows indicate interaction links between technological elements that are permanently available but are used only during the simulation of destructive scenarios. Solid orange arrows represent interaction links between human operators and technological elements that are always active. Solid light blue arrows denote interaction links between technological components that must not be active during destructive scenarios. Adapted from [19]..

"Sensor" block, cf. Fig. 2) and actuators (i.e., "Actuator" block, cf. Fig. 2). The sensors collect and transmit process data, while the actuators act on the system to modify its physical parameters. Indeed, the data acquired by the sensors are processed by an automated controller (i.e., "Automated controller" block, cf. Fig. 2), which issues control commands through the actuators. This automated controller – or more typically, a set of controllers – operates under the supervision of a higher-level control system (e.g., a SCADA system) that aggregates information from multiple sources within the process, then formulating appropriate control strategies (i.e., "Data center" block, cf. Fig. 2). Finally, the model architecture includes a digital counterpart of the physical process (i.e., "Digital model of the system" block, cf. Fig. 2), which is meant to enable the simulation of scenarios that cannot be safely or practically reproduced in the physical setup, such as destructive or hazardous events.

To support the simulation of all the previously discussed human-technology interaction modes (illustrated in Fig. 1), the HHIL model architecture must enable specific interfaces between its elements. These shall allow the model to dynamically represent various operational configurations depending on the scenario being simulated. In particular, Fig. 2 outlines the required interaction links. Starting from the physical layer, continuous data exchange must be ensured between the physical process, sensors, actuators, and the automated controller (i.e., solid light blue arrows, cf. Fig. 2). However, these interactions with the real, tangible process must remain configurable in the model: they must be available, but selectively disabled in simulations involving potentially destructive scenarios. In such cases, the digital model of the system substitutes the real physical process, and the automated controller interacts with this digital representation instead, creating a Hardware-in-the-Loop configuration (i.e., dashed black arrows, cf. Fig. 2). Conversely, other interfaces between technological components must remain active at all times (i.e., solid black arrows, cf. Fig. 2). Specifically, the automated controller must continuously communicate with the data centre to ensure up-to-date process information, whether the data originate from the physical system or the digital model. The data centre, in turn, exchanges this information with the digital model to maintain consistency and enable accurate process simulation when needed. The digital model must also incorporate information related to the human controller. Human-related data can be collected via (e.g.) motion capture systems, cameras, or wearable sensors, and they should be fed into the digital model in a Human-in-the-Loop fashion to enable the simulation of human behaviour and its effects on the system performance. To this end, the human controller must be able to interact with the various technological elements in the system, including: sensors (e.g., by directly reading data to monitor process conditions), actuators (e.g., by manually adjusting process parameters), and the automation system (e.g., by modifying control logic via the data centre). All these human-technology interfaces are represented by solid orange arrows in Fig. 2.

3. Case study

This section presents the case study used to test the HHIL modelling and simulation approach proposed in this paper. The analysed process is firstly introduced, followed by the application of the STPA-Sec/S methodological steps. This latter has been intentionally centred on the development and use of the HHIL model, rather than providing a comprehensive overview of the entire methodology application. Consequently, several details have been intentionally omitted to maintain the focus on the role and the implementation of HHIL simulations within the broader approach.

3.1. Process overview

The case study used to instantiate the proposed HHIL extension of STPA-Sec/S is an experimental mock-up plant located in the Department

of Industrial Engineering and Mathematical Sciences at the Università Politecnica delle Marche in Ancona, Italy [20]. The plant is designed to emulate the artificial extraction of natural gas from a depleted well by exploiting the pressure from an adjacent, active, oil well. This approach is commonly adopted in the energy sector to reduce the cost and the complexity of installing mechanical pumps at reservoir depths. In accordance with the "do no significant harm" principles [21], the plant substitutes crude oil and natural gas with room-temperature water and air, while preserving the structural and functional behaviour of an industrial two-phase flow extraction system. The core component of the process is a gas-liquid ejector, which uses a high-pressure water stream to create a vacuum, allowing ambient air to be drawn into the system. This mimics the operation of real-world ejectors used in the oil and gas industry for the transport of a two-phase gas-liquid mixtures. The process begins with a pump drawing water from an open tank, whose pressure is regulated by a solenoid valve (i.e., AV1), before being delivered to the ejector. Once the high-pressure water enters the ejector, it accelerates, converting pressure into kinetic energy to generate a vacuum. This latter draws in air via an inlet pipeline dedicated to air, eventually forming a two-phase air-water mixture at the ejector outlet. The mixture flows into a vertical tank which serves as a phase separator, isolating the gas and the liquid components. The tank outlet are managed through two additional solenoid valves (i.e., AV2 and AV3) which regulate the exit of the water and the air, respectively, allowing for the active control over the liquid level and the tank internal pressure. To automatically monitor and control the plant, a Revolution Pi (RevPi) Core 3 industrial controller is employed. The RevPi collects real-time data from the plant's sensors (see Table 1 for the complete list of sensors and actuators equipped on the plant), processes it via a PID logic, and commands the actuators (i.e., AV1, AV2, AV3) accordingly. Data and commands are transmitted using a Message Queuing Telemetry Transport (MQTT) publish-subscribe communication architecture, which connects the RevPi to a human-machine interface implemented on both a desktop and a mobile app. The user interface is designed for both real-time supervision of the plant operations, and for enabling the operators to intervene in process control. It provides both live and historical data visualizations for key performance indicators, while offering manual control functionalities for adjusting parameters and tuning the control logics. In this context, the operator plays a critical role in ensuring the whole system's functioning, particularly during anomalous or unexpected conditions that go beyond their passive monitoring. If necessary, the operator can physically inspect the plant to gather additional information or perform direct manual interventions, thereby ensuring that the ejector all the other components continue to operate within their intended performance specifications.

3.2. Applying the STPA-Sec/S steps

The close collaboration between the human operators and the automated technological components in achieving the system's operational objectives makes the above descripted experimental plant a representative example of a CSTS to be used for testing the STPA-Sec/S HHIL approach. Key details of this implementation are reported in the following paragraphs.

Table 1
List of plant components (sensors and actuators) with corresponding description.

Component name	Description
Sensor S1	Sensor measuring the pressure of water entering the ejector
Sensor S2	Sensor measuring the flow rate of water entering the ejector
Sensor S5	Sensor measuring the pressure inside the tank
Sensor S6	Sensor measuring the level of water inside the tank
Sensor S7	Sensor measuring the flow rate of air drawn into the ejector
Actuator AV1	Valve regulating the passage of inlet water, it is used to regulate the pressure of water entering the ejector
Actuator AV2	Valve regulating the passage of outlet water, it is used to regulate the level of water inside the tank
Actuator AV3	Valve regulating the passage of outlet air, it is used to regulate the pressure inside the tank

3.2.1. Define the scope of the analysis

During the problem identification phase (cf. Section 2.1), the analysis focused on the potential cyber-induced failures the plant might face. In particular, the focus was posed on those cyber-attacks leading to the violation of critical pressure thresholds, or resulting in insufficient water or air flowrates, in turn compromising the ability to perform the extraction process. The full list of losses, their associated hazards, and the system-level constraints is considered out of scope, as acknowledged early in Section 3.

3.2.2. Model the safety control structure

Regarding the modelling of the SCS (cf. Section 2.1), the resulting SCS is shown in Fig. 3, where it is possible to notice all the key elements previously described in Section 3.1 and summarized in Table 1.

3.2.3. Identify the unsafe and unsecure control actions

The SCS in Fig. 3 was analysed to retrieve the system's UCAs. The analysis was first carried out by the authors, who manually analysed the SCS and identified all formally correct UCAs, resulting in 90 UCAs. These were then reviewed with the mock-up plant operators to determine which were practically reasonable, yielding a final set of 65 UCAs. For demonstration purposes, the analysis presented in this paper focus only on three UCAs:

UCA 1. The RevPi provides a wrong modification of position for the air actuated valve (AV3) while controlling the amount of air in tank during steady state operations;
UCA 2. The RevPi does not provide any modification in the positioning of the air actuated valve (AV1) while controlling the flow of water in pipeline (pump – ejector) during steady state operations;
UCA 3. The RevPi provides a wrong modification of position for the air actuated valve (AV2) while controlling the water level in tank during steady state operations.

Additional details with respect to the simulation scenarios developed based on these UCAs are later reported in Section 3.2.5.

3.2.4. Develop the simulation model

In relation to the HHIL model architecture shown in Fig. 2, the implementation of the blocks labelled "Physical process", "Sensor", "Actuator", "Automated controller", and "Data center", along with their respective interfaces, should be straightforward to interpret, as they directly correspond to the physical components and connections that constitute the mock-up plant described in Section 3.1. Greater attention, however, should be devoted to explaining how the remaining two blocks (i.e., the "Human controller", and the "Digital model of the system") and their interfaces with the plant's components have been integrated into the simulation model.

The integration of the human operator into the HHIL simulation model has been realized through the development of a human digital twin using the Perception Neuron Studio motion capture system by Noitom. This solution was mainly chosen for its portability and for its ability to perform high-precision motion tracking without the need for external cameras or fixed infrastructure, thus avoiding any modifications to the existing mock-up plant layout. The setup mimicked the one by Bortolini et al. [22], and it included 17 inertial measurement sensors placed on key joints of the operator's body to capture the full kinematics of their movements at up to 240 frames per second. However, for the purposes of this study, only the pelvis sensor was used to approximate the operator's centre of mass, tracking their movements within the mock-up plant environment at a resolution of 90 fps. The raw output of the motion capture system consisted of a time series of

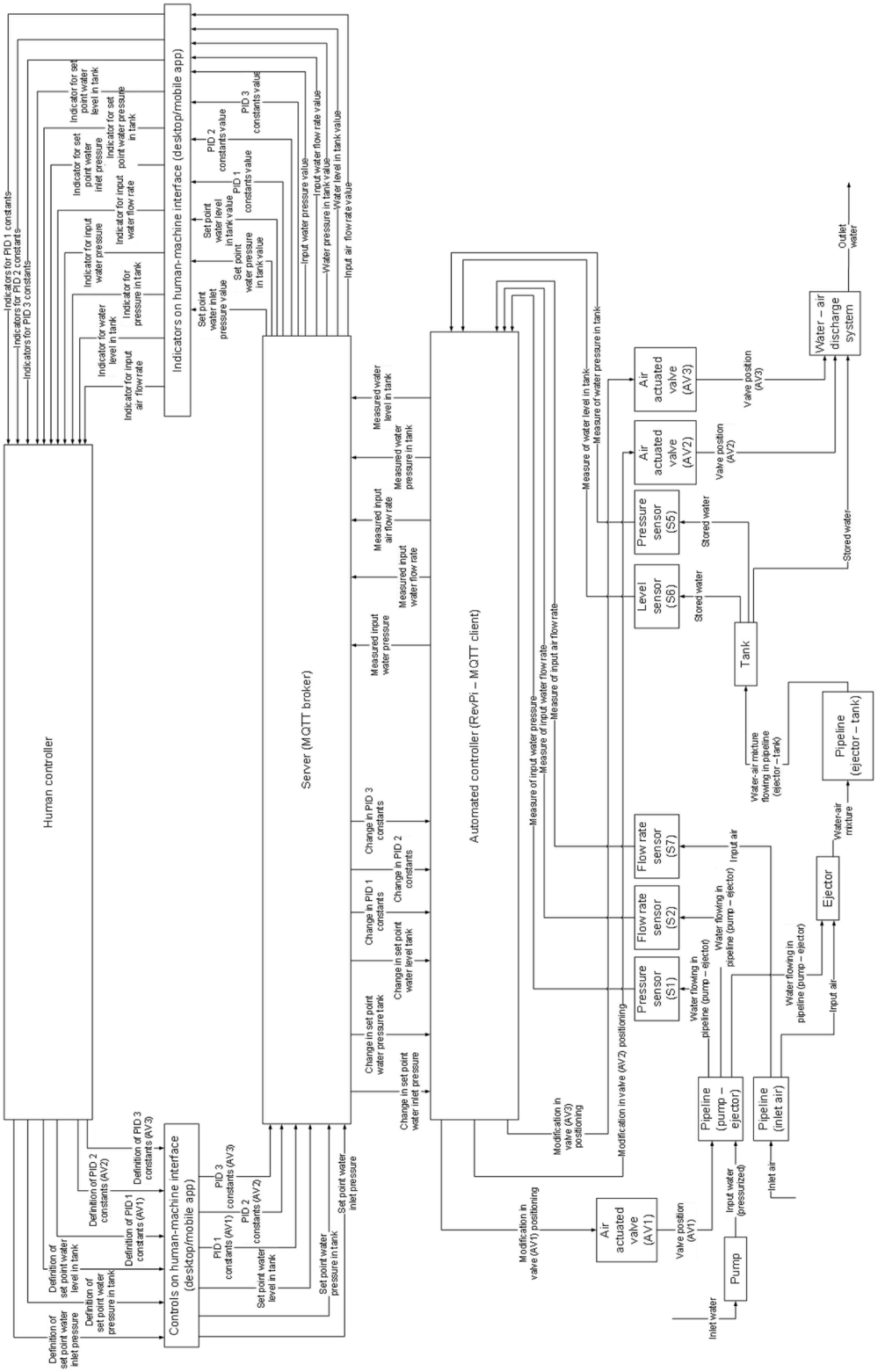

Fig. 3. SCS of the analysed CSTS.

three-dimensional coordinates for the pelvis sensor, referenced to a coordinate system anchored to a neutral (i.e., empty) fixed point. Starting from this spatial data, the operator's positions were contextualized into actions within the system, by introducing a set of control volumes, i.e., virtual three-dimensional zones positioned around key components of the plant, such as (e.g.) inspection points. Each control volume was defined as a non-overlapping parallelepiped, with specific origin coordinates and dimensions along the three axes. Accordingly, the operator's actions became an interaction with a given control volume, determined through a presence condition checking whether the pelvis sensor's coordinates fell within the control volume boundaries during a given time frame. This approach is aligned with the control volumes method, which is largely discussed and employed in literature [23]. The processed output of the motion capture system simulator has

the data structure reported in Table 2, with N being the total number of control volumes defined. The data table may have multiple rows, and their number is not fixed a priori but rather depends on the number of control volumes visited by the human operator during the HHIL simulation time.

On the other hand, the digitalization of the physical process, intended to allow a digital twin to replace the physical plant as the automated controller's target during destructive simulation scenarios, was implemented using the Python-based model developed by Mazzuto et al. [24]. Such digital model estimates the plant process variable by replicating the coupled behaviour of two main plant subsystems: the pump-ejector line, and the vertical separation tank. Accordingly, the pump-ejector model simulates how the water pressure drives the air suction, computing pressure values based on the water flow conditions, and enabling the simulated adjustment of the water inlet valve (i.e., AV1). In parallel, the vertical tank model computes the tank internal pressure and the water level over time, using this information to simulate the regulation of the two outlet valves positions (i.e., AV2, and AV3). The two subsystems are dynamically interconnected by a feedback loop since variations in the tank pressure and in the level of water must affect the ejector suction performance, too. The model is capable of returning a time series of values for all the simulated plant's components (i.e., sensors and actuators) which precisely integrate with the real ones coming from physical components. The resulting data structure is reported in Table 3 where M is the total number of time steps in which data are collected.

3.2.5. Define the resilience metrics

A set of resilience metrics was defined to track the differences between the nominal performance and the disrupted one by means of the eight process parameters of interest. To avoid oversimplification, at this stage a dedicated resilience metric has been proposed for each available system element, i.e., the one described by the S1, S2, S5, S6, and S7 measures, and by the AV1, AV2, AV3 behaviour. Each resilience metric followed a similar structure, specifically:

$$R_i = \frac{As_i - |As_i - Ad_i|}{As_i} \tag{1}$$

where As_i represents the area under the nominal performance curve, Ad_i represents the area under the disrupted performance curve, and $i \in$ [S1, S2, S5, S6, S7, AV1, AV2, AV3] identifies the eight different resilience indicators being defined.

The metrics were based on the well-established "area under the curve" approach [25], which is conceptually related to the measurement of resilience through the so-called "resilience triangle" [26].

Fig. 4 illustrates an exemplary i-th system performance using an hypothetical time series generated by the plant's digital model. Both the nominal (i.e., dotted line, cf. Fig. 4) and disrupted (i.e., solid line, cf. Fig. 4) performance curves are shown, with the resilience indicator R_i represented by the shaded area. At first glance, the definition of the resilience indicators may appear disconnected from the operator's behaviour. However, in relation to Fig. 4, it is important to identify five distinct phases in the disrupted curve that eventually indicate when and how the operator contributes to the overall system performance. Specifically:

Steady-state, cf. Fig. 4 [t_0, t_1). In a first phase of operations, the system behaves under a steady-state condition (i.e., P_s, cf. Fig. 4), with no cyber-attacks nor disruption occurring. As the intended analysis assumes that at some time a cyber-attack will succeed, this phase offers no opportunity for the operator's intervention. Therefore, operator-related data are not relevant in this phase assuming that neither positive nor detrimental effects can be put in place by them;
Under attack, cf. Fig. 4 [t_1, t_2). At a certain point in time (i.e., t_1, cf. Fig. 4), a cyber-attack starts affecting the plant performance. During this phase, the operator's role refers first to detecting the anomaly. The operator's detection time (i.e., t_2, cf. Fig. 4), directly influences the length of this phase: the sooner the operator identifies the attack, the shorter the "under attack" phase lasts;
Shutdown, cf. Fig. 4 [t_2, t_3). Once the attack is detected, the operator lets the plant enter in the shutdown phase, where the performance declines to a "zero" condition (i.e., P_0, cf. Fig. 4). Although the duration of this phase is not directly governed by operator's behaviour, it is indirectly affected by how quickly the operator reacted in the previous phase. Specifically, the time required to reach P_0 depends on the extent to which the system was impacted by the attack (i.e., P_d, cf. Fig. 4). For instance, if the performance curve in Fig. 4 represents the water level in the tank (i.e., S6), the earlier the attack is detected, the less the water level rises, making it possible to empty the tank quicker;
Plant off, cf. Fig. 4 [t_3, t_4). After the shutdown, the plant remains off, entering an unintended yet stable period. During this time, the operator assesses the attack's effect on the system, investigates potential malfunctions on components, and carries out the necessary recalibrations and diagnostics tasks. The length of this phase is directly linked to the duration of the operator's intervention: the quicker these tasks are completed, the sooner the plant can resume its operations;
Start-up, cf. Fig. 4 [t_4, ∞). Once all the recovery and verification activities are completed, the operator begins the start-up process (i.e., from t_4 onward, cf. Fig. 4), and the plant transitions back to its original steady-state (i.e., P_s, cf. Fig. 4). The transition can be non-linear and discontinuous, as mimicked by the line in Fig. 4. Since this phase is largely automated, the operator's action has no influence over its duration and, thus, the operator-related data are not included in this phase either.

Table 2
Structure of output data resulting from the human operator model based on the motion capture system.

Control volume1	Control volume2	…	Control volumeN	Persistence [s]
1 if the human operator interacts with control $volume_1$; 0 instead	1 if the human operator interacts with control $volume_2$; 0 instead	…	1 if the human operator interacts with control $volume_N$; 0 instead	Time the human operator spend interacting with the control volume with value 1
…	…	…	…	…

3.2.6. Model the faults and their effects

Based on the selected UCAs listed in Section 3.2.3, two representative scenarios were formalized to serve as the foundation for the subsequent resilience analysis. In a way similar to the UCAs, a set of formally correct loss scenarios related to the selected UCAs (i.e., see Section 3.2.3) were first identified by the authors during the STPA-Sec/S analysis. Then, in consultation with the plant operators, only those deemed operationally relevant and feasible were retained. Among these, two were specifically highlighted by the plant operators as both relevant and feasible for simulation: the first related to UCA 3, and the second related to UCA 1 and UCA 2, simultaneously. As such, they were selected and here presented to evaluate the HHIL simulation approach. The scenarios are described below, detailing the faults and their technical implications, as well as the expected responses from the human operator when performing the HHIL simulations:

Scenario 1: false low reading of tank level. The feedback "Measured water level in tank" or "Measure of water level in tank" (cf. Fig. 3) indicates low level of water inside the tank. However, this reading is incorrect, and it result from a cyber-attack as (e.g.) a stealthy false data injection (FDI) attack that deliberately manipulates sensor readings, or a covert Man-in-the-Middle (MitM) attack that prevents

Table 3
Structure of output data resulting from the physical plant model by Mazzuto et al. [24].

Time [s]	S1 [bar]	S2 [m^3/h]	S5 [bar]	S6 [mm]	S7 [m^3/h]	AV1 [%]	AV2 [%]	AV3[%]
Time $step_1$	Inlet water pressure at Time $step_1$	Inlet water flow rate at Time $step_1$	Pressure in the tank at Time $step_1$	Water level in tank at Time $step_1$	Inlet air flow rate at Time $step_1$	Valve AV1 position at Time $step_1$	Valve AV2 position at Time $step_1$	Valve AV3 position at Time $step_1$
Time $step_2$	Inlet water pressure at Time $step_2$	Inlet water flow rate at Time $step_2$	Pressure in the tank at Time $step_2$	Water level in tank at Time $step_2$	Inlet air flow rate at Time $step_2$	Valve AV1 position at Time $step_2$	Valve AV2 position at Time $step_2$	Valve AV3 position at Time $step_2$
…	…	…	…	…	…	…	…	…
Time $step_M$	Inlet water pressure at Time $step_M$	Inlet water flow rate at Time $step_M$	Pressure in the tank at Time $step_M$	Water level in tank at Time $step_M$	Inlet air flow rate at Time $step_M$	Valve AV1 position at Time $step_M$	Valve AV2 position at Time $step_M$	Valve AV3 position at Time $step_M$

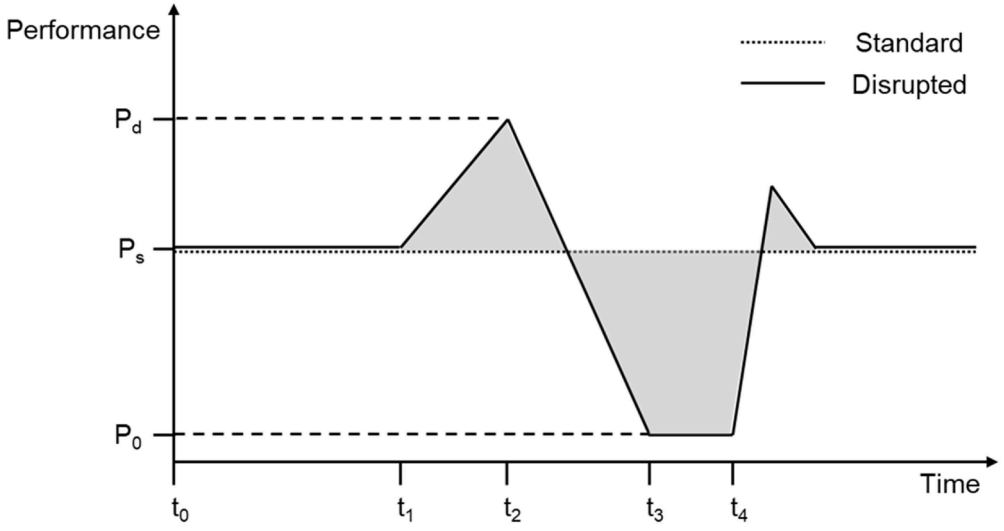

Fig. 4. Exemplary standard and disrupted performance curves used for calculating the resilience indicator, with key time moments and corresponding performance values.

sensor updates while simultaneously injecting malicious code or manipulating control setpoints to alter plant behaviour without detection. The anomalous situation prompts the automated controller to change the way AV2 is managed (i.e., UCA 3): assuming the tank to be underfilled, AV2 is closed to avoid additional water to draw out. However, the actual water level continues to rise, and as the water accumulates, the pressure in the tank increases. In this case, the correct provision of the feedback "Measured water pressure in tank" and "Measure of water pressure in tank" make the automated controller aware of the increase in pressure and it reacts by managing AV3 properly, eventually releasing the air excess to maintain a stable pressure level. However, the amount of water produced at the plant's output decreases due to the closing of AV2, and the rising water level could lead to the tank overflow. Upon detecting the sensor discrepancy and recognizing the associated overflood risk, the operator intervenes by disabling the automated controller to stop any further unintended control action. This manual intervention is aimed at bringing the plant to a stable state. Subsequently, the system is fully shutdown to allow the inspection and the resolution of the fault. This is done by recalibrating the compromised sensor and by checking the other plant's components functioning. After these tasks are completed, the system is restarted, the automated control is re-enabled, and the plant restores its operations under nominal conditions.

Scenario 2: false low reading of tank pressure. A wrong provision of the feedback "Measured water pressure in tank" or "Measure of water pressure in tank" (cf. Fig. 3 mislead the control of AV3. Again, this scenario can be verified due to various types of attacks as (e.g.) a stealth attack with FDI, a covert MitM attack, or even a denial-of-service (DoS) attack which prevents the controllers to have an updated vision on the pressure parameter. In the attempt to correct the reported low pressure, the automated controller closes the valve AV3 allowing less air to escape from the tank (i.e., UCA 1) and eventually making the internal pressure dangerously increase. In parallel, the management of inlet water through AV1 remains unchanged (i.e., UCA 2) but the increasing internal pressure impedes water to flow properly, thus compromising the ejector's ability to draw air. Upon detecting the abnormal system behaviour, the human operator is expected to intervene by disabling the automated controls, thereby manually stabilizing the plant in a safe configuration. If executed promptly, this manual control override prevents the situation from escalating and establishes the conditions for a controlled plant shutdown. The operator then proceeds with the shutdown and carries out the necessary restoration procedures (i.e., sensor recalibration) and checks to bring the plant back to its intended operational state. Once the issue has been resolved, the system is restarted, the automated control is reactivated, and the normal operating conditions are restored, with all process variables returning to their nominal values.

For clarity, the two scenarios will be referred respectively to as Scenario 1 and Scenario 2 throughout the remainder of the paper.

3.2.7. *Evaluate the system resilience performance*

The evaluation of the system's resilience performance was supported by an experimental campaign aimed at collecting operators' data within

the mock-up plant environment during the management of Scenario 1 and Scenario 2. Both expert and novice operators were involved in the simulations to capture a diverse range of human responses to the abnormal system conditions. A total of 8 operators participated in the experimental campaign, with 3 experts and 5 novices. All the participants had an engineering background, and the same initial training for operating the plant. Being the plant a mock-up used for research purposes, we considered 1 year of experience enough to be considered as an expert on the plant functioning. While no formal repeated or scheduled training was ever provided beyond the initial session, the operators continuously engage in research activities related to the plant. As such, participants with more than one year of experience working at the plant were classified as expert operators, whereas those with less than one year of experience were considered novices.

In a human-in-the-loop perspective, all the participants interacted with the plant using a combination of digital control panels and remote interfaces (i.e., desktop and mobile apps), and the plant's physical equipment (i.e., physical sensors and actuators). In each simulation session, the operators' actions were tracked using the motion capture system, generating data as outlined in Table 2. These temporal sequences were subsequently analysed and manually related to either the detection or the restoration phase (i.e., from t_1 to t_2, or from t_3 to t_4, cf. Fig. 4).

During the experimental campaign, the sensors equipped on the plant continuously recorded the physical process data, returning a data table as the one described in Table 3. In cases where some dangerous condition was being reached, the system automatically triggered a shutdown, stopping the real plant operations. In such instances, the digital simulation model was used to complete the remaining portions of the time series, ensuring a consistent dataset for the analysis.

Given these two types of data (i.e., the human-related data, and the plant performance data) a MATLAB script was developed to assess the whole CSTS resilience under varying operational conditions. Specifically, the script performed Monte Carlo simulations to compute the resilience indicators R_i distributions with respect to different operators' behaviours, drawn from the collected experimental data. The Monte Carlo simulation counted 500 iteration, being set conservatively guaranteeing $a < 5$ % error threshold [27].

At each iteration, the script randomly picks up two human-related data tables, one for the detection activity, and one for the restoration activity. Then, the script draws randomly a plant operational performance curve, among the ones generated during the experimental campaign. The three inputs are merged considering the phases in which the human performance affects the plant's one (see Section 3.2.5): the detection time (i.e., the difference between t_2 and t_1, cf. Fig. 4) and restoration time (i.e., the difference between t_4 and t_3, cf. Fig. 4) are computed by summing up the last column of the detection and restoration temporal sequences, respectively. The two values are then used to truncate the "under attack", "shutdown", and "plant off" segment of the disrupted performance curve ensuring continuity and consistency between data leveraging again the plant's digital model. Specifically, the detection time identifies the time moment in which the attack is resolved and consequently the maximum performance degradation (i.e., t_2 and P_d, cf. Fig. 4). This latter is then used to initialize the "shutdown" phase. The restoration time, on the other hand, governs the duration of the "plant off" phase reflecting how long it takes for the operator to perform diagnostics, recalibrations, and prepare the system for its restart. Accordingly, each simulation iteration returns a three-dimensional data object containing: (i) the detection time value; (ii) the restoration time value; and (iii) a table structured as Table 3 representing the plant parameter curves generated during that specific simulation run.

The data generated through the simulation process is then used to compute the resilience indicators R_i. To this end, a reference dataset representing the nominal performance of the plant was created by recording approximately two hours of stable operations, during which no disruption was introduced. Any transient behaviour was excluded to ensure the dataset reflected only the steady-state system conditions. However, before comparing this reference time series with the disrupted performance curves, a set of adjustments were applied to ensure the most accurate and meaningful computation of system resilience.

The first post-processing step involved aligning the lengths of the two timeseries. Since the disrupted performance curves varied in duration depending on the simulated operator's behaviours (i.e., different detection and restoration times), the nominal performance dataset was cyclically trimmed to match the exact length of each disrupted curve. This ensured that the area under the standard performance curve As_i and the area under the disrupted performance curve Ad_i referred to the same temporal extent. Without this step, excessively long standard curves could artificially inflate As_i, pushing the resilience score R_i closer to 1, regardless of the actual extent of the system degradation.

The second adjustment addressed the temporal alignment of the two signals. Simply trimming both time series to the same length does not guarantee that periodic behaviours, such as (e.g.) oscillations around a given set-point, are phase-aligned. In other words, even if the two curves may share the same frequency, any time shift between them can result in a discrepancy in the point-by-point area comparison. This misalignment is particularly problematic when considering the absolute difference $|As_i - Ad_i|$ in Eq. (1), as it can lead to a significant underestimation of resilience even when the overall system behaviour remains basically the same. To mitigate this issue, the assessment incorporated the Dynamic Time Warping (DTW), a technique designed to align time series that may differ in speed or timing [28]. DTW works by identifying the optimal non-linear alignment between two sequences, minimizing the cumulative distance between their elements. Specifically, given two hypothetic time series:

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_a, \ \ldots, \ x_A) \tag{2}$$

and:

$$\boldsymbol{y} = (y_1, y_2, \ldots, y_b, \ \ldots, \ y_B) \tag{3}$$

the DTW constructs a cost matrix $\boldsymbol{D}$ in which each element $D(a, b)$ represents the distance between the a-th element of $\boldsymbol{x}$, and the b-th element of $\boldsymbol{y}$ calculated as the Euclidean distance between the two:

$$D(a, \ b) = (x_a - y_b)^2 \tag{4}$$

The goal of the procedure is to find a warping path $\boldsymbol{w}$ within $\boldsymbol{D}$:

$$\boldsymbol{w} = (w_1, w_2, \ldots, \ w_c, \ \ldots, \ w_C) \tag{5}$$

such as each $w_c = (a_c, \ b_c)$ aligns elements from $\boldsymbol{x}$ and $\boldsymbol{y}$ minimizing the cost function:

$$DTW(\boldsymbol{x}, \ \boldsymbol{y}) = min \sum_{c=1}^{C} D(a_c, b_c) \tag{6}$$

where $c = 1, \ldots, C$ iterates over all the couples a and b within the warping path $\boldsymbol{w}$.

As a constraint, the *DTW* path must satisfy: (i) the boundary conditions (i.e., start and end must be at the first and last elements), (ii) the continuity (i.e., steps of size 1), and (iii) the monotonicity (i.e., preserve the time order) of the timeseries $\boldsymbol{x}$ and $\boldsymbol{y}$.

A final refinement was meant to exclude non-anomalous segments from the resilience analysis, ensuring that the resilience metric would not be artificially underestimated simply because it tended to 1 due to large values of both As_i and Ad_i. Indeed, to avoid such phenomenon, the resilience assessment was focused specifically on the periods where a meaningful deviation in performance occurs via the implementation of a pragmatic anomaly detection procedure. This latter was empirically tuned and operates by identifying as anomalous any time segment where the disrupted curve deviates by at least 5 % from the nominal curve, and this deviation is kept for at least 50 consecutive timesteps. On the other

hand, an anomalous period is considered concluded only after 500 consecutive points fall below the 5 % deviation threshold. All non-anomalous time segments were excluded from the resilience calculation by setting their values to zero before computing the area under the curve.

For clarity, Fig. 5 presents the standard and disrupted time series of the component S6 (i.e., the water level in tank sensor) both before and after the application of the data post-processing procedures described above.

4. Simulation results and HHIL resilience assessment

The HHIL simulation runs provided valuable insights into the resilience performance of the CSTS, with a particular focus on the different behaviours of both the expert and the novice operators. This section presents the results obtained for the two simulation scenarios that are compared from two complementary perspectives. The first subsection examines the distribution of: (i) the detection time, (ii) the restoration time, and (iii) the resilience indicators defined in Section 3.2.5. The second subsection adopts a broader perspective, introducing two overall resilience metrics: one related to the amount of water produced by the process, and the other related to the volume of air drawn in. Indeed, it is important to recall that the system under analysis is an experimental plant simulating an extraction process, in which water and air represent oil and natural gas, respectively. As such, these two metrics are directly linked to the overall efficiency of the process being simulated.

4.1. Results distributions

To assess the statistical significance of the differences observed in results concerning experts and novices, a hypothesis testing approach was adopted. The null hypothesis (H_0) stated that "the resilience of expert and novice operators is equal", while the alternative hypothesis (H_1) stated that "the resilience of expert and novice operators is different". Each resilience indicator was first tested for normality using the Shapiro-Wilk test. Since no indicator yielded a *p*-value greater than 0.05, the assumption of normality was rejected. Consequently, the non-parametric Mann-Whitney U test was applied to compare each pair of

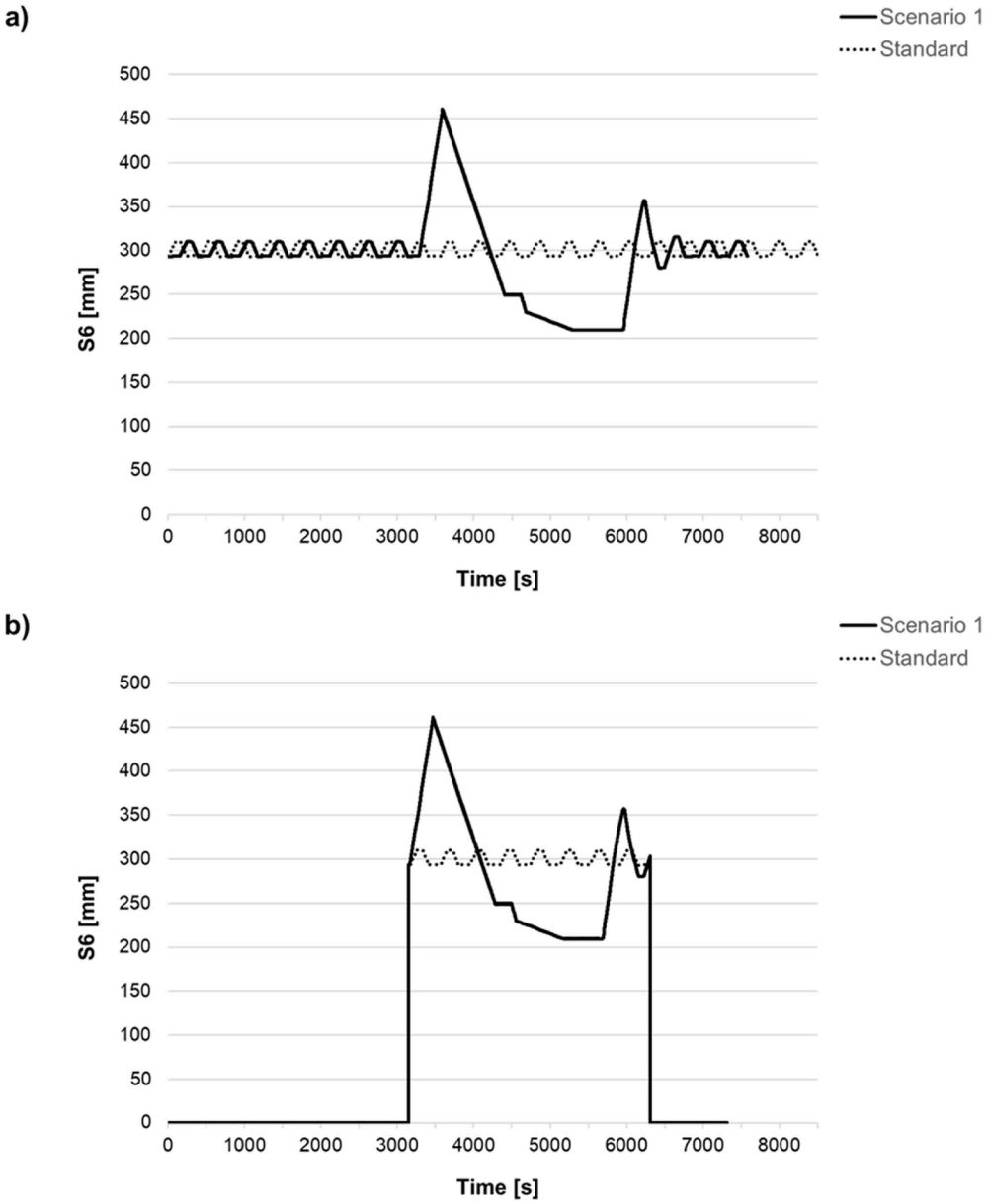

Fig. 5. Sensor S6 standard and disrupted (Scenario 1) timeseries before (a) and after (b) the post-processing procedure performed before the resilience calculation.

resilience indicators between novice and expert operators across both scenarios. The obtained *p*-values are reported in Table 4, data confirming the H_0 hypothesis (*p*-value > 0.05), thus showing no significance difference between the two distributions, are underlined in Table 4.

On this premises, Figs. 6 and 7 include the box plots for the resilience indicators related to the eight plant's component (cf. Table 1) in Scenario 1 and Scenario 2, respectively. Additionally each figure details the difference between the expert (cf. Figs. 6**a** and 7**a**) and novice (cf. Figs. 6**b** and 7**b**) operators' distributions. For completion, numerical values (mean, standard deviation, minimum value, maximum value, and 25 %, 50 %, and 70 % percentiles) are included in the Tables 5–8.

In the first scenario, the cyber-attack compromises the sensor S6, responsible for monitoring the tank's water level. The expert operators identify the failure significantly faster than novices, with an average detection time of 261.798 s against 433.982 s: a difference of nearly three minutes. This gap points at the expected experts' superior performance, and to their deeper understanding of the plant's normal behaviour. Accordingly, for them it was much easier to capture the difference in the plant's functioning by means of (e.g.) subtle noise when the disruption occurred. Moreover, the experts exhibit overall a much lower variability in detection times, with a standard deviation of 56.155 s versus 105.556 s for novices, indicating, overall, a more consistent and effective response. In contrast, looking at the restoration times, they are nearly identical between the two groups, on average: 30.716 s for experts against the 30.612 s for novices, both with minimal variation (1.654 versus 1.679), as expected from the *p*-value in Table 4. This result may suggest that once the fault is recognized, the recovery procedures are well-defined and equally accessible to both groups, which basically act the same.

The resilience indicators related to the plant's technical components offer further distinctions. If looking at the compromised sensor S6, expert operators achieve a significantly higher resilience score (0.818 versus 0.706), along with a lower variability (0.039 versus 0.074), pointing at a more reliable capacity to stabilize the system despite losing direct visibility on the water level. This may indicate that the experts are likely compensating the missing information by interpreting indirect signals or inferring other system states. For sensors S1, S2, and S5, resilience values are broadly similar. Novices slightly outperform experts in S1, but the difference remains negligible (0.883 versus 0.877), and the variability remains low for both (0.015 versus 0.014). Regarding S2, all operators fail in maintaining an adequate input water flow rate, with average resilience values as similar as low (0.375 versus 0.376). In S5 instead, both groups show an equally high average resilience (0.881 for both), suggesting an effective pressure management even under the attack on S6. The three indicators on sensors S1, S2, S3 show a shared intention among operator to prioritize the pressure management, which is more critical, instead than the flow one. Experts also show a slightly better performance in S7 (0.755 versus 0.732), with a bit lower variability (0.013 versus 0.018), indicating better control over the air intake.

Table 4
Mann-Whitney *p*-value results for resilience variables in the two scenarios. Underlined values represent no significant difference between the expert and novice distributions.

Variable	*p*-value (Mann-Whitney) for Scenario 1	*p*-value (Mann-Whitney) for Scenario 2
Detection time	1.254×10^{-120}	1.515×10^{-93}
Restoration time	$\underline{2.494 \times 10^{-1}}$	$\underline{4.206 \times 10^{-1}}$
R_{S1}	5.838×10^{-15}	1.767×10^{-5}
R_{S2}	2.071×10^{-2}	$\underline{3.833 \times 10^{-1}}$
R_{S5}	9.264×10^{-19}	$\underline{6.093 \times 10^{-2}}$
R_{S6}	8.387×10^{-111}	$\underline{9.833 \times 10^{-1}}$
R_{S7}	8.621×10^{-82}	1.353×10^{-7}
R_{AV1}	6.088×10^{-48}	8.876×10^{-6}
R_{AV2}	1.563×10^{-31}	2.560×10^{-2}
R_{AV3}	8.521×10^{-5}	2.189×10^{-15}

The resilience indicators related to the valve controls reveal some additional strategic differences between the two types of users. The experts exhibit a higher resilience in the behaviour of AV1 with an average of 0.334 compared to 0.281 for novices, with basically equal variability (0.051 versus 0.052). This may reinforce the idea that expert operators place a stronger effort on stabilizing the input when the direct feedback on the tank level is compromised. They also outperform novices in the AV2 and AV3 controls, which handle the air and water discharge. Although AV2 shows a really low resilience, experts attain a higher mean value (0.084 versus 0.049) and a slightly greater variability (0.048 versus 0.037), implying a potentially riskier approach. In AV3, experts again show slightly higher average resilience (0.487 versus 0.480) with a bit reduced variability (0.021 versus 0.026), reflecting an albeit minimal steadier control.

In the second scenario, the cyber-attack targets S5, which is instrumental to monitor the pressure inside the tank, a central variable for ensuring system safety. Again, on average expert operators detect the failure more quickly (238.170 s versus 288.622 s), almost one minute faster than novices, reinforcing their higher situation awareness. Restoration times remain not significantly different between the groups (cf. Table 4), indicating a possible consistent application of the recovery protocols even in Scenario 2. Notably, when looking at the technical-related resilience indicators, the difference between the two is in this case much less evident, with expert operators displaying even a slightly lower resilience in some components (i.e., S1 and S7), all of which are closely linked to the pressure management in both the pipes and the tank. This may suggest a more aggressive control strategy, where rapid interventions introduce temporary instabilities that ripple through the system. Despite this, both groups demonstrate high average resilience in S5 (0.803 for both), reflecting an effective system recovery strategy following the disruption on tank pressure. On the contrary, both the groups demonstrate an extremely low average resilience value for S1 (0.246 versus 0.249) which was among the best performing in Scenario 1. Moreover, also in this case, the input flow rate remains the most disrupted performance metric overall (on average, 0.059 for both), underlining the widespread impact of the pressure sensor compromise up to the earlier stages of the system. Valve control related results once again indicate that experts prioritize input stability, with slightly higher resilience in AV1 (0.241 versus 0.228). Conversely, they exhibit a marginally lower resilience in AV2 (0.212 versus 215) and AV3 (0.342 versus 3.64), suggesting an albeit minimal trade-off consisting in sacrificing the tank stability – which was unsensed because of the cyber-attack – to regain control over the input pressure parameter. However, the three resilience indicators related to the valves control remains among the lower, in a way similar to Scenario 1.

4.2. Overall resilience performance

As mentioned above, this paragraph takes a broader perspective on the CSTS resilience by means of: (i) the output water flow rate, which relates to the system's ability to sustain a consistent water (that in the experimental layout mimics the actual plant's oil output); and (ii) the input air flow rate, which relates to the system's capacity to maintain the suction efficiency at the ejector inlet, eventually ensuring a stable pressure support for the reservoir.

Again, the two dimensions were checked for consistency performing the Shapiro-Wilk test at first. None of the dimensions passed the test (*p*-value < 0.05) thus the Mann-Whitney U test was conducted. Similarly to Table 4, the obtained results are presented in Table 9.

Figs. 8 and 9 include two scatterplots related to the aforementioned metrics for Scenario 1, where the resilience measures are plotted against the operators' detection time. Only the detection time was considered as the results in Section 4.1 shows how the variations in the restoration time are not significant. Both plots clearly demonstrate that rapid anomaly detection is the most critical factor driving system resilience.

In the water flow plot (cf. Fig. 8), the overall performance loss

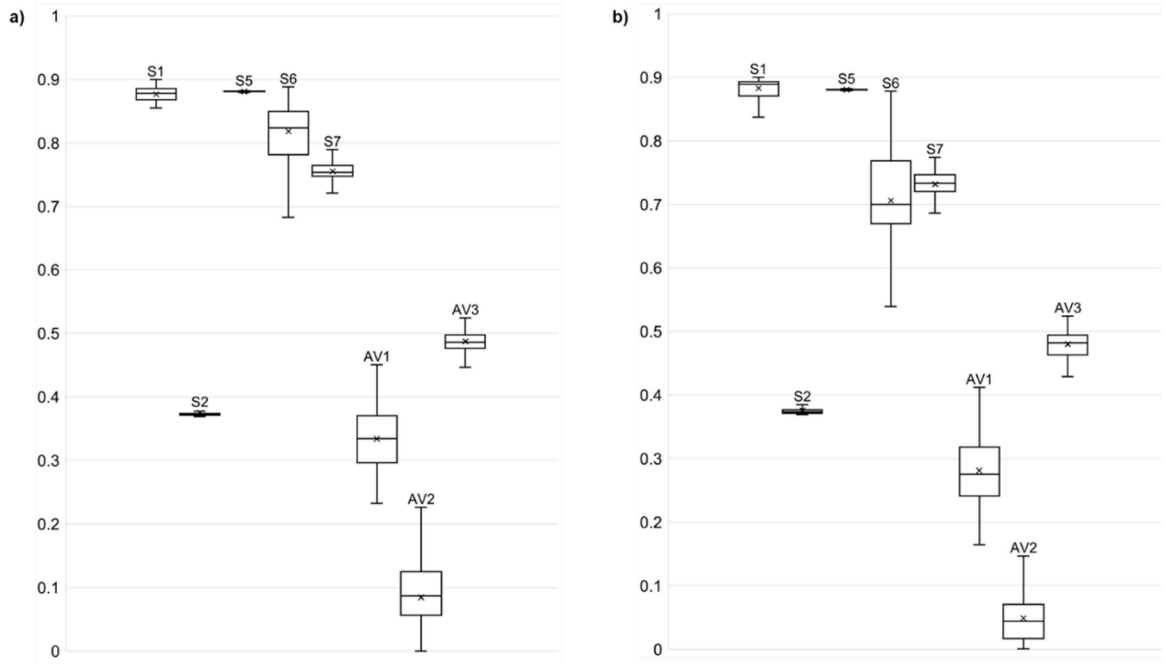

Fig. 6. Distribution of resilience indicators per plant's components in Scenario 1 grouped by expert (a) and novice (b) operators.

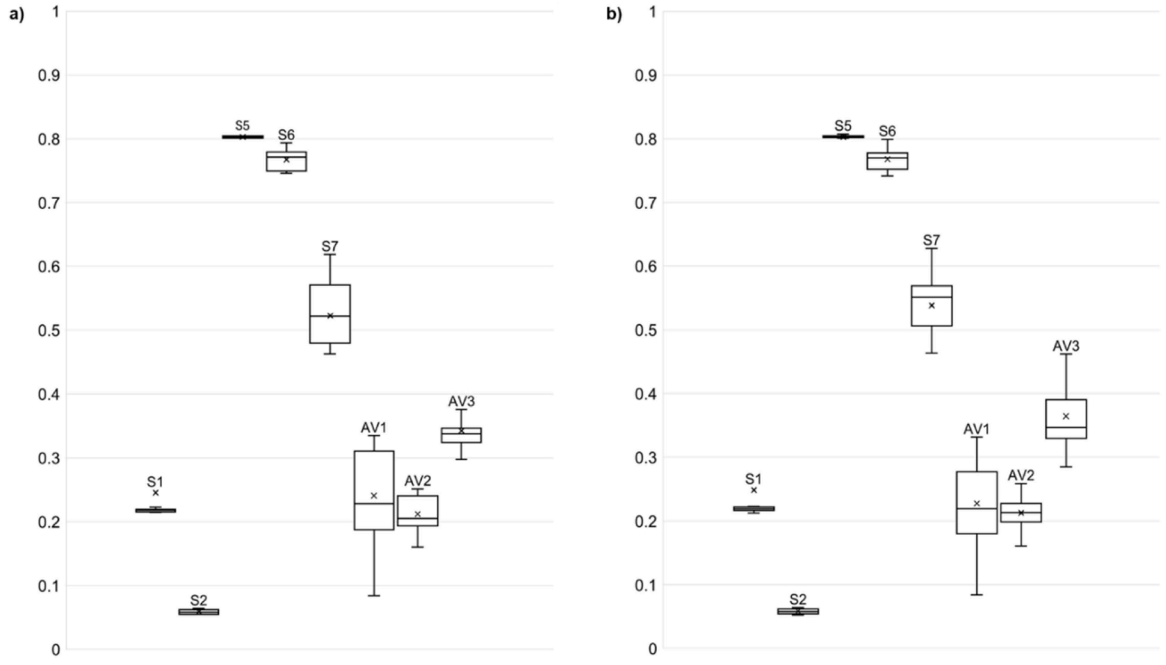

Fig. 7. Distribution of resilience indicators per plant's components in Scenario 2 grouped by expert (a) and novice (b) operators.

remains substantially high for all operators. Experts are clustered within a detection window of 150 to 400 s, with an inter-quartile range (IQR) equal to 81 s, and resilience values largely falling between 0.40 and 0.55. This indicates their ability to identify the issue promptly, and their consistency in controlling the water output, regardless of any variation in detection time. This consistency is further confirmed by a low IQR in the resilience score (i.e., equal to 0.027). In contrast, the novice operators exhibit a much wider distribution, with most detection times ranging from 350 to 700 s (IQR = 150.5 s), and even lower resilience scores, generally between 0.15 and 0.45. Their resilience variability is also slightly higher, as indicated by a larger IQR (i.e., equal to 0.139), highlighting a less consistent control over the water output. This supports the idea that the detection delay is a major driver for recovery effectiveness. However, several data points from novice operators show a sort of rebound in resilience for detection times beyond 600 s, reaching levels comparable to those associated with much earlier detections. While this apparent recovery may not directly reflect improved operators' skills, it remains a notable result. One possible explanation is that some novices, when faced with ambiguous signals or minor anomalies, adopt a passive "let-it-run" approach, delaying the intervention out of caution. During this period of inaction, given that the cyber-attack targets the level sensor, only, the plant's PID controllers continue to regulate pressures and flows, keeping the system relatively stable until corrective actions are finally taken. Nevertheless, a more intriguing

Table 5
Numerical results for the distributions of detection time, restoration time, and resilience indicators in Scenario 1 for the expert operator.

R_{AV3}	0.487	0.021	0.429	0.476	0.486	0.498	0.577
R_{AV2}	0.084	0.048	0.001	0.057	0.087	0.125	0.226
R_{AV1}	0.334	0.051	0.232	0.296	0.334	0.370	0.451
R_{S7}	0.755	0.013	0.720	0.747	0.754	0.765	0.798
R_{S6}	0.818	0.039	0.675	0.782	0.824	0.849	0.888
R_{S5}	0.881	0.001	0.873	0.881	0.881	0.881	0.882
R_{S2}	0.375	0.007	0.369	0.371	0.373	0.374	0.393
R_{S1}	0.877	0.015	0.776	0.868	0.878	0.886	0.900
Restoration time [s]	30.716	1.654	28.000	29.000	30.000	32.000	33.000
Detection time [s]	261.798	56.155	143.000	220.000	254.000	301.000	417.000
	Mean	**Standard deviation**	**Minimum**	**25 %**	**50 %**	**75 %**	**Maximum**

Table 6
Numerical results for the distributions of detection time, restoration time, and resilience indicators in Scenario 1 for the novice operator.

R_{AV3}	0.480	0.026	0.429	0.463	0.482	0.494	0.524
R_{AV2}	0.049	0.037	0.001	0.017	0.044	0.071	0.190
R_{AV1}	0.281	0.052	0.165	0.242	0.275	0.318	0.446
R_{S7}	0.732	0.018	0.687	0.720	0.733	0.747	0.774
R_{S6}	0.706	0.074	0.481	0.670	0.700	0.768	0.878
R_{S5}	0.881	0.001	0.879	0.880	0.881	0.881	0.882
R_{S2}	0.376	0.007	0.369	0.371	0.373	0.377	0.393
R_{S1}	0.883	0.014	0.820	0.870	0.889	0.893	0.900
Restoration time [s]	30.612	1.679	28.000	29.000	30.000	32.000	33.000
Detection time [s]	433.982	105.556	184.000	358.500	426.500	509.000	709.000
	Mean	**Standard deviation**	**Minimum**	**25 %**	**50 %**	**75 %**	**Maximum**

Table 7
Numerical results for the distributions of detection time, restoration time, and resilience indicators in Scenario 2 for the expert operator.

R_{AV3}	0.342	0.036	0.279	0.324	0.338	0.346	0.460
R_{AV2}	0.212	0.024	0.161	0.194	0.205	0.241	0.252
R_{AV1}	0.241	0.056	0.084	0.188	0.228	0.310	0.335
R_{S7}	0.523	0.043	0.463	0.480	0.522	0.571	0.618
R_{S6}	0.768	0.015	0.746	0.750	0.771	0.779	0.793
R_{S5}	0.803	0.001	0.801	0.801	0.803	0.804	0.805
R_{S2}	0.059	0.003	0.055	0.055	0.058	0.063	0.064
R_{S1}	0.246	0.065	0.215	0.215	0.218	0.220	0.405
Restoration time [s]	1590.710	118.542	1429.000	1442.000	1606.000	1737.000	1739.000
Detection time [s]	238.170	18.253	193.000	225.750	240.000	252.000	280.000
	Mean	**Standard deviation**	**Minimum**	**25 %**	**50 %**	**75 %**	**Maximum**

Table 8
Numerical results for the distributions of detection time, restoration time, and resilience indicators in Scenario 2 for the novice operator.

R_{AV3}	0.364	0.046	0.285	0.330	0.347	0.390	0.462
R_{AV2}	0.215	0.019	0.161	0.199	0.213	0.228	0.259
R_{AV1}	0.228	0.054	0.084	0.181	0.219	0.277	0.332
R_{S7}	0.538	0.037	0.463	0.506	0.551	0.569	0.628
R_{S6}	0.768	0.016	0.742	0.752	0.770	0.778	0.799
R_{S5}	0.803	0.001	0.801	0.802	0.803	0.804	0.808
R_{S2}	0.059	0.004	0.053	0.055	0.058	0.063	0.064
R_{S1}	0.249	0.062	0.213	0.216	0.219	0.223	0.457
Restoration time [s]	1597.218	120.752	1429.000	1442.000	1606.000	1737.000	1739.000
Detection time [s]	288.622	40.528	207.000	257.000	283.000	315.000	392.000
	Mean	**Standard deviation**	**Minimum**	**25 %**	**50 %**	**75 %**	**Maximum**

Table 9
Mann-Whitney *p*-value results for resilience related to the output water flow rate, and resilience related to the input air flow rate in the two scenarios.

Variable	*p*-value (Mann-Whitney) for Scenario 1	*p*-value (Mann-Whitney) for Scenario 2
Output water flow rate	6.874×10^{-117}	6.687×10^{-20}
Input air flow rate	8.621×10^{-82}	1.353×10^{-7}

interpretation is that less experienced operators may, paradoxically, benefit from a less constrained and more improvisational management approach. They might be forced to adapt without relying on established routines or prior expectations, employing a more flexible and exploratory strategy that, under stress, results in unexpectedly better performance. This result suggests how extensive experience – though generally advantageous – could sometimes lead to more rigid, less adaptive behaviours that ultimately limit system's resilience itself. This could be argued also with the experts' resilience plots, which appear more normally distributed.

The air input flow rate scatterplot in Fig. 9 reinforces the broader conclusions commented above, but with a less evident rebound distortion, and much higher resilience values. Experts resilience scores range between 0.75 and 0.80 (IQR = 0.017), reflecting an overall good

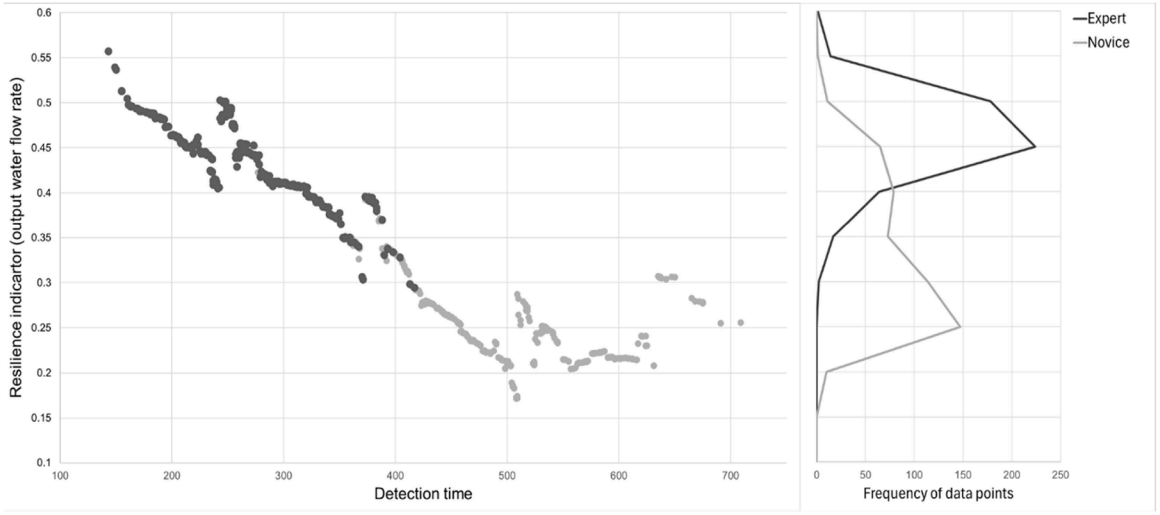

Fig. 8. Overall process performance resilience metric (i.e., output water flow rate) for Scenario 1 distributed by detection time for expert (dark grey) and novice (light grey) operators with frequency of occurrence on the right. In the frequency plots, the count of points within each range is represented at the upper bound of the range.

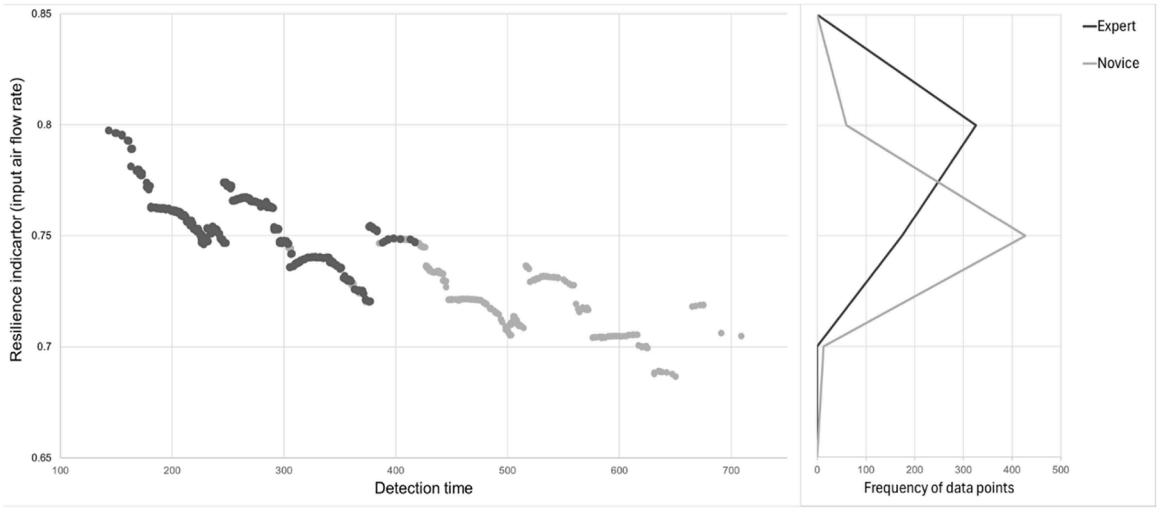

Fig. 9. Overall process performance resilience metric (i.e., input air flow rate) for Scenario 1 distributed by detection time for expert (dark grey) and novice (light grey) operators with frequency of occurrence on the right. In the frequency plots, the count of points within each range is represented at the upper bound of the range.

management of the suction performance. In contrast, the novices show a greater resilience decline as detection delays increase, dropping down to a resilience value of 0.70 or lower, for detection times beyond 500 s. However, their variability remains minimal, as demonstrated by the low IQR in the resilience distribution (i.e., equal to 0.026).

Similarly, Figs. 10 and 11 shows the scatter plots of the detection time against the resilience computed by means of the water output flow rate and the air input flow rate, for Scenario 2, in which the tank-pressure sensor S5 is compromised. As for Scenario 1, differences in the restoration times were not significant, and only the detection time was considered in the analysis.

In both plots, expert operators are tightly clustered within a detection window of 200 to 350 s (IQR = 26 s), achieving a moderate yet stable resilience performance. Their resilience values range approximately between 0.50 and 0.65 for both the water output and the air input flow rates. The distributions of the two resilience dimensions show minimal variabilities, with IQR = 0.053 for the output water flow rate, and IQR = 0.091 for the air inlet. Interestingly, novice operators exhibit a similar overall pattern, with resilience indicators stabilizing within a comparable range of 0.45 to 0.60, and overall low IQRs (i.e., 0.017 for the water output, and 0.063 for the air inlet), yet demonstrating a wider variability in detection time (IQR = 59.25 s). However, a notable divergence appears in the water output flow rate where around 350 s, their performance shows a clear decline, stabilizing at a lower average of approximately 0.50.

This threshold-driven drop in performance may reflect an important characteristic of the pressure-sensitive dynamics investigated in Scenario 2. As pointed out in the previous findings, the early detection proves to be an even critical aspect in this case. Once a sensor breach goes undetected beyond a certain point, the pressure oscillations intensify, preventing both a manual adjustment, and the PID-based stabilization. This leads to the identification of a kind of “non-returning point” (i.e., when detection goes over 350 s), to be avoided at all costs, where the system control becomes significantly more difficult. In this context, the passive “let-it-run” approach that may have benefited novices in Scenario 1 appears counterproductive in Scenario 2. The pressure-based disruption brings the system to move more, and more permanently from its target values, making it increasingly difficult to

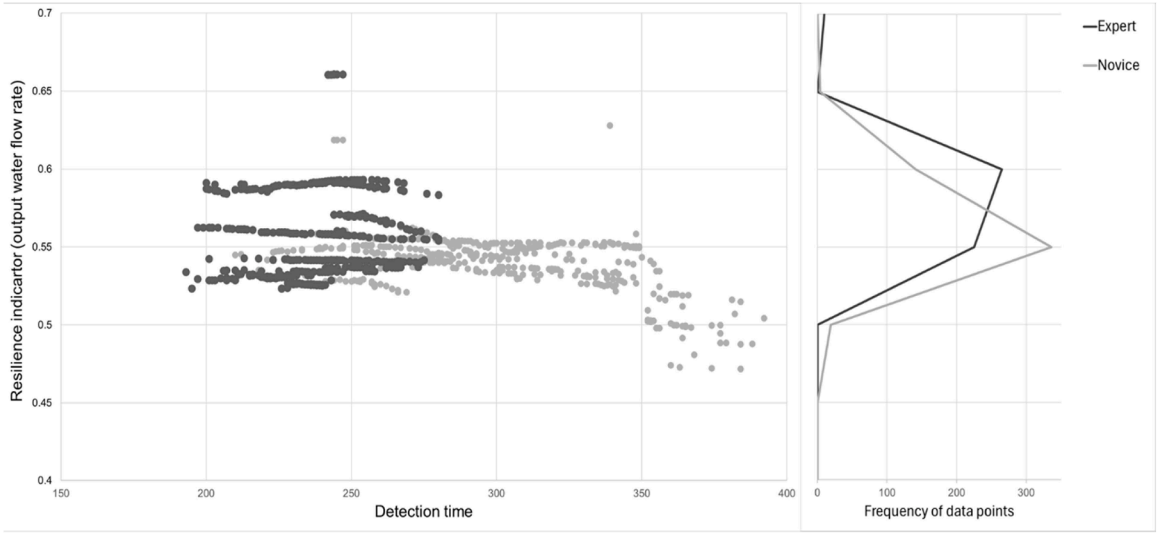

Fig. 10. Overall process performance resilience metric (i.e., output water flow rate) for Scenario 2 distributed by detection time for expert (dark grey) and novice (light grey) operators with frequency of occurrence on the right. In the frequency plots, the count of points within each range is represented at the upper bound of the range.

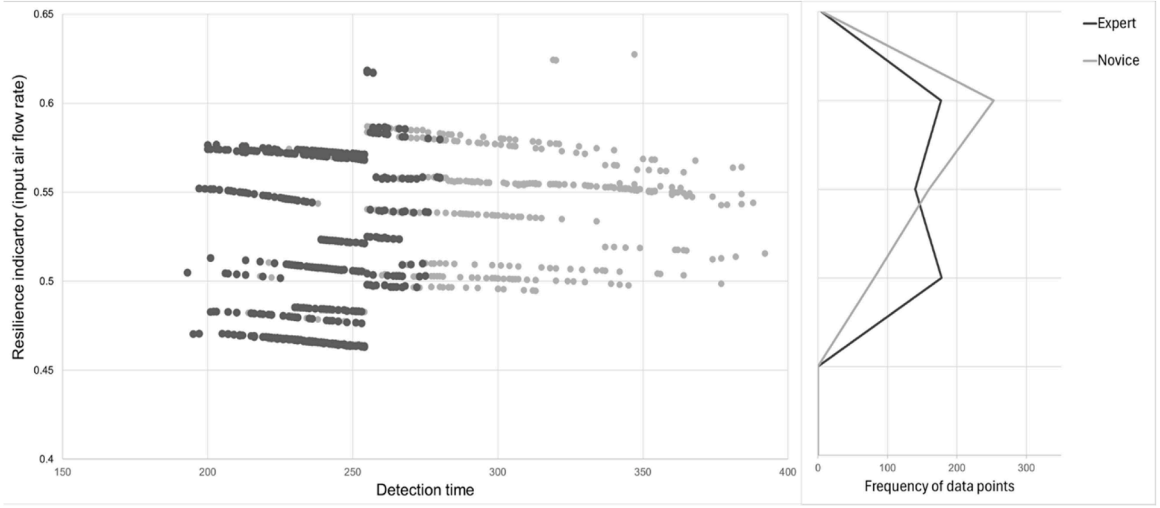

Fig. 11. Overall process performance resilience metric (i.e., input air flow rate) for Scenario 2 distributed by detection time for expert (dark grey) and novice (light grey) operators with frequency of occurrence on the right. In the frequency plots, the count of points within each range is represented at the upper bound of the range.

recover and sustain high resilience performance if the intervention is too delayed.

5. Discussion

This study set out to examine how resilience analysis, supported by HHIL experimentation and informed by STAMP-inspired modelling, can contribute to understanding interactions within a CSTS during the occurrence of cyber-attacks. Although the experimental setting was simulated and non-critical, the obtained findings nonetheless offer several insights of practical relevance for decision-makers and system managers.

Indeed, a central result concerns the pivotal role of humans in determining system resilience. Across the two scenarios explored in Section 4, human decision-making (by means of their ability to detect anomalies early) proved to be an influential determinants of overall system performance. Experts generally reacted faster and more reliably. However, their strategies tended to be more rigid, which is advantageous under expected conditions but can limit adaptability in situations that diverge from their experience. In contrast, novice operators, despite being slower and more variable in their actions, occasionally demonstrated unexpectedly resilient performance, especially in Scenario 1, by adopting more flexible and improvised strategies. While it should be recalled that this study was conducted in an experimental setting, this contrast echoes the tension described by Grote et al. [29] between controllability and adaptability, and aligns with Woods' Theory of Graceful Extensibility [30], which views resilience as the ability to extend performance as system boundaries are approached or exceeded. From a system management and design perspective, these findings underscore the need to cultivate both reliable control and adaptive flexibility within CSTSs' elements. Training remains essential for building operator competence, but strict notions alone cannot ensure resilient behaviour under surprising conditions. System architectures and interfaces must therefore embed affordances that support flexibility and accommodate adaptation, to let resilience emerge not simply as a property of skilled individuals, but as an engineered feature within the

broader CSTS.

The study also highlights several concrete opportunities for system and interface improvements. During various simulations, experts were often able to quickly detect faults by triangulating information from multiple sources not displayed on the current operator panel. These experts' heuristics could potentially be formalized into clear visual cues embedded within the interface design, thereby enhancing novices' situation awareness by making such tacit knowledge more accessible and explicit. Similarly, the significant impact observed when individual sensors (e.g., S5) failed, suggests that both redundancy and estimation-based substitutions would strengthen operators' situational awareness and reduce vulnerability to cascading disruptions. The wide-ranging effects of cyber-attacks point to the need for more holistic automation strategies: instead of the current point-to-point valve logic, a more coordinated control approach could help maintain global system stability, while modular segmentation might limit the propagation of disturbances, especially in larger industrial contexts.

5.1. Limitations

Several limitations should be considered when interpreting the findings of this study. First, all experiments were conducted in a simplified, simulated CSTS environment. While this controlled setting enabled a systematic examination of cyber-attack dynamics, it inevitably abstracts away much of the complexity found in real-world infrastructures, which feature richer variability. It is important to stress how the generalizability of the results obtained is therefore limited.

A limitation concerns the modelling of human behaviours. Although the paper conceptually integrates a STAMP perspective with HHIL experimentation and resilience analysis, the representation of human factors in the case study is – intentionally – simplified. Behavioural inference only relied on the time operators spent within predefined control volumes, and motion capture served to verify their physical presence in these zones. This creates a gap between the potential richness of the conceptual approach and the simplicity of the behavioural data ultimately captured. Moreover, assuming that physical proximity to the control area directly reflects actions is a notable simplification, as operators may respond (and thus decide with respect) to cues unrelated to spatial location, such as (e.g.) auditory signals and alarms, peripheral perception, or communications with colleagues. Consequently, in the presented case study, the human factor remains partially a "black box" limiting the depth with operators' strategies can be interpreted.

The sampling of human modelling was similarly limited. Operators were grouped into experts and novices based on the time they spent working in the mock-up plant, without capturing in depth the broader diversity of real-world operational profiles. Important dimensions such as (e.g.) teamwork dynamics, or fatigue and stress, were not included. In a real plant, a more comprehensive human factor modelling should be considered to ultimately derive both physically and cognitively engagement during the cyber disruptions.

The analytical framework adopted for resilience quantification also introduces its own constraints. The study relies heavily on the resilience triangle (cf. Section 3.2.5), a proxy that compresses performance complexity into a single geometric measure, which has been recently acknowledged to have limited representational power [31]. While it can mask important dimensions of resilience, particularly those related to adaptation dynamics and recovery pathways, it remains a simple communicable proxy for capturing system performance under disruption. While the approach has been applied carefully and critically, being aware of its limitations, the resilience triangle is here used just as an accessible entry point. It is recognized that more nuanced approaches, such as multi-dimensional performance envelopes or adaptive capacity metrics, would likely provide a more comprehensive understanding of system behaviour under attack.

Finally, the scope of cyber-attacks and system configurations examined was objectively – yet intentionally – limited to just show an application of the method. Only three UCAs, and two attack scenarios were analysed. Clearly, other attacks may have yielded different outcomes, and would deserve further dedicated investigations.

Taken together, these limitations suggest that the present work should be regarded primarily as a demonstration of the proposed methodological approach for human-hardware digital twin representation, rather than a comprehensive model of human-system performance under cyber-attack conditions.

6. Conclusion

This paper introduced a HHIL simulation approach to enhance the systemic yet quantitative resilience assessment of CSTS, building upon the STPA-Sec/S methodology. By integrating human behaviour and physical process data into a simulation environment, the proposed approach aimed at offering a more realistic representation of complexity in system facing cyber-related disruptions. The approach was tested through an experimental oil and gas plant managed by two distinct operator groups (i.e., experts and novices). The application demonstrated that the HHIL simulations were capable of effectively capturing the systemic interplay between human decision-making and system performance. Overall, the obtained results indicated that expert operators generally responded more quickly and consistently, while novices occasionally displayed adaptive behaviours that led to unexpectedly resilient outcomes. These findings highlight the importance of balancing technical expertise while fostering the adaptive capacity of operators, as both dimensions appear closely linked to ensuring the overall CSTS resilience and thus, graceful extensibility.

Being here emerged only as hints due to the several limitations acknowledged in the presented case study (see Section 5.1), future research shall explore in more detail how these dimensions actually relate to real system resilience. At present, the implementation is limited to custom-designed experimental environments, and integrating it into real safety-critical industrial systems is left as an open challenge for future research.

In addition, future developments could aim to complement the HHIL analysis with more established human factors methodologies such as (e.g.): the Hierarchical Task Analysis (HTA), the Cognitive Work Analysis (CWA), the Human Error Analysis (HEA), or even the more recent Structured Exploration of Complex Adaptations (SECA). The integration of such approaches may provide a more structured decomposition of operator tasks, a deeper understanding of their cognitive constraints, and a systematic identification of human-related vulnerabilities and responses. Although some of these approaches may partially appear in contrast with the principles of Resilience Engineering and Safety-II, they could nonetheless offer a structured scaffold for navigating complexity in a more guided manner. In highly intricate or safety-critical scenarios, this additional layer of methodological structure may represent a valuable complement to resilience-oriented analyses, enhancing both their interpretability and their analytical depth.

Moreover, as noted by recent STAMP-related studies, further developments could focus on reducing the inherent subjectivity being present in STAMP/STPA-based analyses, providing measurable, reproducible insights while mitigating the method's reliance on expert judgment alone. Although the use of simulations in the current study is meant to help reducing the analysts' subjectivity, future research may further reinforce this point by exploring how different methods such as (e.g.) fuzzy modelling [32], reinforcement learning [33], or formal causal modelling [34] can be combined with the HHIL framework to enable a more rigorous and objective path to resilience assessment in complex CSTSs.

In conclusion, we believe the proposed HHIL approach represents a little yet meaningful step forward in bridging qualitative system-theoretic analysis with more quantitative assessments. By providing a more realistic understanding of human-machine interdependencies in complex systems, this approach lays an important foundation for the

development of safer and more resilient industrial systems in an increasingly complex and digitalized world.

CRediT authorship contribution statement

Francesco Simone: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marco Bortolini:** Writing – review & editing, Software, Resources, Data curation. **Giovanni Mazzuto:** Writing – review & editing, Software, Resources, Data curation. **Giulio Di Gravio:** Writing – review & editing, Supervision. **Riccardo Patriarca:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been funded by the European Union – NextGenerationEU under the National Recovery and Resilience Plan (PNRR) – Mission 4 Education and Research – Component 2 From research to business - Investment 1.1, Prin 2022 Notice announced with DD No 104 of 2/2/2022, entitled RESIST – RESilience management to Industrial Systems Threats, proposal code 2022YSAE2X – CUP B53D23006650006.

Data availability

Data will be made available on request.

References

[1] Dobra Z, Dhir KS. Technology jump in the industry: human–robot cooperation in production. Indus Robot 2020;47(5):757–75. https://doi.org/10.1108/IR-02-2020-0039.
[2] Hollnagel E, Wears RL, Braithwaite J. From Safety-I to Safety-II: a white paper. The resilient health care net. USAAustralia: University of Southern Denmark, University of FloridaMacquarie University; 2013. p. 1–32.
[3] Leveson N. Engineering a safer world: systems thinking applied to safety. Choice reviews online. The MIT Press; 2012. https://doi.org/10.5860/choice.49-6305.
[4] Baxter G, Sommerville I. Socio-technical systems: from design methods to systems engineering. Interact Comput 2011;23(1):4–17. https://doi.org/10.1016/j.intcom.2010.07.003.
[5] Rajkumar R, Lee I, Sha L, Stankovic J. Cyber-physical systems: the next computing revolution. In: Proceedings - design automation conference; 2010. p. 731–6. https://doi.org/10.1145/1837274.1837461.
[6] Yaacoub J-PA, Salman O, Noura HN, Kaaniche N, Chehab A, Malli M. Cyber-physical systems security: limitations, issues and future trends. Microprocess Microsyst 2020;77. https://doi.org/10.1016/j.micpro.2020.103201.
[7] Patriarca R, Falegnami A, Costantino F, Gravio GD, Nicola AD, Villani ML. WAx: an integrated conceptual framework for the analysis of cyber-socio-technical systems. Saf Sci 2021;136(105142). https://doi.org/10.1016/j.ssci.2020.105142.
[8] Patriarca R, Chatzimichailidou M, Karanikas N, Gravio GD. The past and present of System-Theoretic Accident Model and Processes (STAMP) and its associated techniques: a scoping review. Saf Sci 2022;146:105566. https://doi.org/10.1016/j.ssci.2021.105566.
[9] Ebrahimi H, Zarei E, Ansari M, Nojoumi A, Yarahmadi R. A system theory based accident analysis model: sTAMP-fuzzy DEMATEL. Saf Sci 2024;173:106445. https://doi.org/10.1016/j.ssci.2024.106445.
[10] Sun H, Wang H, Yang M, Reniers G. Dynamic risk assessment of chemical process systems using the System-theoretic accident model and process approach (STAMP) in combination with cascading failure propagation model (CFPM). Saf Sci 2024;171:106375. https://doi.org/10.1016/j.ssci.2023.106375.
[11] Zhu T, Meng C, Han X, Wang Y, Dang J, Chen H, Qi M, Zhao D. A risk assessment framework for water electrolysis systems: mapping system theoretic process analysis (STPA) and event tree analysis (ETA) into fuzzy Bayesian networks (FBN). Process Saf Environ Protect 2025;194:306–23. https://doi.org/10.1016/j.psep.2024.11.117.
[12] Nakhal Akel AJ, Campari A, Paltrinieri N, Patriarca R. STheBaN - system-theoretic bayesian approach for the evaluation of inspections workability in hydrogen operations. J Loss Prev Process Ind 2025;97:105687. https://doi.org/10.1016/j.jlp.2025.105687.
[13] Liao W, Qiao Y, Dong T, Gou Z, Chen D. A human reliability analysis method based on STPA-IDAC and BN-SLIM for driver take-over in level 3 automated driving. Reliab Eng Syst Saf 2025;254:110577. https://doi.org/10.1016/j.ress.2024.110577.
[14] Bensaci C, Zennir Y, Pomorski D, Innal F, Lundteigen MA. Collision hazard modeling and analysis in a multi-mobile robots system transportation task with STPA and SPN. Reliab Eng Syst Saf 2023;234:109138. https://doi.org/10.1016/j.ress.2023.109138.
[15] Simone F, Akel AJN, Gravio GD, Patriarca R. Thinking in systems, sifting through simulations: a way ahead for cyber resilience assessment. IEEE Access 2023;11:11430–50. https://doi.org/10.1109/ACCESS.2023.3241552.
[16] Nakhal Akel AJ, Simone F, Lombardi M, Costantino F, Di Gravio G, Tronci M, Bortolini M, Mazzuto G, Patriarca R. Dealing with I5.0 complexity: cyber-socio-technical systems modelling and analysis. In: *Proceedings of the summer school Francesco Turco*. XXIX summer school Francesco Turco. Scopus; 2024. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105005405421&partnerID=40&md5=0c6594dfc1ba0427787b9809adb46dff.
[17] Davis E, Marcus G. The scope and limits of simulation in automated reasoning. Artif Intell 2016;233:60–72. https://doi.org/10.1016/j.artint.2015.12.003.
[18] Björck F, Henkel M, Stirna J, Zdravkovic J. Cyber resilience – fundamentals for a definition. Adv Intell Syst Comput 2015;353:311–6. https://doi.org/10.1007/978-3-319-16486-1_31.
[19] Simone F, Nakhal Akel AJ, Lombardi M, Di Gravio G, Patriarca R. Human-hardware In the loop (HHIL) STAMP-based simulations to model cyber-physical complexity in experimental high-risk plants. In: 11th European STAMP workshop and conference: advancing safety in a complex world; 2024.
[20] Di Carlo F, Mazzuto G, Bevilacqua M, Ciarapica FE, Ortenzi M, Donato LD, Ferraro A, Pirozzi M. A process plant retrofitting framework in industry 4.0 perspective *. IFAC-PapersOnLine 2021;54(1):67–72. https://doi.org/10.1016/j.ifacol.2021.08.007.
[21] European Commission. Technical guidance on the application of 'do no significant harm' under the Recovery and Resilience Facility Regulation. Online 2021;C(c).
[22] Bortolini M, Ferrari E, Gamberi M, Galizia FG, Giannone E. A human-digital twin model to track human motion in an experimental Cyber-socio-technical system. Procedia Comput Sci 2025;253:1373–81. https://doi.org/10.1016/j.procs.2025.01.199.
[23] Bortolini M, Faccio M, Gamberi M, Pilati F. Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes. Comput Ind Eng 2020;139:105485. https://doi.org/10.1016/j.cie.2018.10.046.
[24] Mazzuto G, Pietrangeli I, Ortenzi M, Ciarapica FE, Bevilacqua M. Leveraging Digital Twin for operational resilience in the oil and gas industry. Array 2025;27:100443. https://doi.org/10.1016/j.array.2025.100443.
[25] Hosseini S, Barker K, Ramirez-Marquez JE. A review of definitions and measures of system resilience. Reliab Eng Syst Saf 2016;145:47–61. https://doi.org/10.1016/j.ress.2015.08.006.
[26] Tierney K, Bruneau M. Conceptualizing and measuring resilience: a key to disaster loss reduction. TR News 2007;250:14–7.
[27] Driels MR, Shin YS. Determining the number of iterations for Monte Carlo simulations of weapon effectiveness. Monterey, California: Naval Postgraduate School; 2004. http://hdl.handle.net/10945/798.
[28] Dynamic time warping. In: Müller M, editor. Information retrieval for music and motion. Springer; 2007. p. 69–84. https://doi.org/10.1007/978-3-540-74048-3_4.
[29] Grote G, Kolbe M, Waller MJ. The dual nature of adaptive coordination in teams: balancing demands for flexibility and stability. Organ Psychol Rev 2018;8(2–3):125–48. https://doi.org/10.1177/2041386618790112.
[30] Woods DD. The theory of graceful extensibility: basic rules that govern adaptive systems. Environ Syst Decis 2018;38(4). https://doi.org/10.1007/s10669-018-9708-3.
[31] Eisenberg DA, Seager TP, Alderson DL. The rebound curve is a poor model of resilience. PNAS Nexus 2025;4(3):pgaf052. https://doi.org/10.1093/pnasnexus/pgaf052.
[32] Nakhal AAJ, Patriarca R, De Carlo F, Leoni L. A System-Theoretic Fuzzy Analysis (STheFA) for systemic safety assessment. Process Saf Environ Protect 2023;177:1181–96. https://doi.org/10.1016/j.psep.2023.07.014.
[33] Chang J, Kwon R, Kwon G. STPA-RL: Integrating Reinforcement Learning into STPA for Loss Scenario Exploration. Appl. Sci. 2024;14(7). https://doi.org/10.3390/app14072916.
[34] Riccardi L, Compare M, Mascherona R, Zio E. Structural causal modeling and STPA for the risk analysis of a rail system powered by H2 fuel. Reliab Eng Syst Saf 2025:256. https://doi.org/10.1016/j.ress.2024.110758.