Process Safety and Environmental Protection 172 (2023) 632–641

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.journals.elsevier.com/process-safety-and-environmental-protection

An approach towards the implementation of a reliable resilience model based on machine learning

Tomaso Vairo [a], Margherita Pettinato [a,*], Andrea P. Reverberi [b], Maria Francesca Milazzo [c], Bruno Fabiano [a]

[a] Department of Civil Chemical and Environmental Engineering, University of Genoa, via Opera Pia 15, 16145 Genoa, Italy
[b] Department of Chemistry and Industrial Chemistry Department, University of Genoa, via Dodecaneso 31, 16146 Genoa, Italy
[c] Department of Engineering, University of Messina, Contrada di Dio, 98166 Messina, Italy

ARTICLE INFO

Keywords:
AI explainability
Energy transition
Learning assurance
Hidden Markov Model
LNG fuel
Plant resilience

ABSTRACT

Machine Learning tools to enhance systems' resilience received an increased impetus driven by energy transition, climate change and digitalization, but critical challenges on system requirement definition and reliability of learning processes need to be addressed. This study proposes a systematic framework based on system engineering and focused on the reliability of the learning process of the Hidden Markov Model (HMM) coupled with the Baum-Welsh algorithm. The HMM hidden states may represent the precursors of accidental events, being the states between a regular performance and a failure of a sub-system. The Baum-Welch algorithm, estimating the parameters of the HMM, iteratively updates the estimates of the state transition and observation probabilities. The framework was applied to a real case of LNG bunkering, showing that the system can learn from incomplete data, improve the learning quality given a new set of observations, make predictions about the latent states and enhance system resilience. The novelty of this work lies in ensuring the learning process and contributing to the attainment of an explainable, robust, and interpretable data-driven approach.

1. Introduction

Even though "risk" and "resilience" are both terms with a long history, as pointed out by Aven (2022) their relation is strongly debated. The most acceptable approach is to consider risk as an aspect of resilience: the assessment and management of risk is rooted in preventing, or counteracting threats before they occur, whereas the core of resilience assessment and management is the system adaptation in the aftermath of threats. A crucial ability of a resilient industrial organisation is the anticipation of the system weak signals. Early detection, representing one of the main pillars of resilience (Pawar et al., 2021), refers to the recognition of precursors of undesired events and results in a more effective response to disturbances, which could otherwise potentially lead to dangerous and difficult situations (Jain et al., 2018). Detection and recognition of early warning signals should be emphasised besides design of error-tolerant equipment, plasticity of mind, and recoverability to ensure effective emergency responses. In this respect, artificial intelligence (AI) represents an important tool to continuously monitor the risk level of plants, enabling improved semi-automated hazard identification, more accurate risk assessment by pattern recognition, advanced statistics and revealing cause-effect structures (Pasman, 2021). Recently, data-driven approaches to risk analysis have become widespread, showing a good ability to dynamically represent the risk (Kamil et al., 2021) and to capture co-dependencies of complex systems (Mamudu et al., 2021; Sarbayev et al., 2019), possibly allowing early-fault detection on the basis of an effective solution to time-dependent sequence learning problem (Arunthavanathan et al., 2021). The most widely used approach involves the extension of Bayesian networks (BNs), originated from the exploration of uncertainty in the field of AI to the traditional risk assessment techniques (Meng et al., 2022; Nhat et al., 2020; Vairo et al., 2020; Yang et al., 2013). BNs proved to be one of the most effective theoretical models for representing uncertainty and reasoning (Guo et al., 2021) and showed satisfactory performance in the quantitative calculation of complex systems in applications, such as process safety assessment (Ghosh et al., 2020) and analysis of risk factors that may cause blowout accidents in deepwater drilling (Liu et al., 2021). Resilience can be considered as a forward and proactive defense (Dinh et al., 2012) with a focus on the dynamic assessment of the system's capacity at each moment of its life

* Corresponding author.
E-mail addresses: tomaso.vairo@edu.unige.it (T. Vairo), margherita.pettinato@edu.unige.it (M. Pettinato), andrea.reverberi@unige.it (A.P. Reverberi), mfmilazzo@unime.it (M.F. Milazzo), brown@unige.it (B. Fabiano).

https://doi.org/10.1016/j.psep.2023.02.058
Received 20 July 2022; Received in revised form 16 February 2023; Accepted 20 February 2023
Available online 21 February 2023

Nomenclature

a	state transition probability.
AI	artificial intelligence.
b	output probability.
BN	Bayesian network.
EASA	European Union Aviation Safety Agency.
EM	expectation-maximization algorithm.
ESD	emergency shut-down.
HMM	hidden Markov model.
i	variable counter.
j	variable counter.
LNG	liquefied natural gas.
ML	machine learning.
n	sequence counter.
PERC	powered emergency release coupling.
PLC	programmable logic controller.
s1, s2, sn	future time steps.
SE	system engineering.
SIMOPS	simultaneous operations.
S_n	states of the system.
t	time.
T	sequence length.
X_n	hidden state vector.
x_n	actual hidden state.
y	possible observation.
Y(n)	conditional probability distribution.
α	probability of seeing observations at time t, given the hidden state i.
β	probability of predicting all future observations.
θ	parameters.
ξ_{ij}	transition matrix.
Υ_{ij}	emission matrix.

(Linkov, Trump, 2019). On these grounds AI and ML techniques can provide the right support by analysing large amounts of data in real-time to provide reliable predictions, consider system evolution and make corrections based on new evidence (Paltrinieri et al., 2019). It should be noted that we are experiencing challenging times in a rapidly evolving risky world, where three drivers play a determining role, namely digitalization, climate change and energy transitions. Notably, the still-evolving Covid-19 pandemic demonstrated the need for robust tools to face with unexpected events by a critical and balance application of novel developments in data science and digital technologies with fundamental science and engineering principles (Fabiano et al., 2022). The introduction of new threats and hazards, lack of historical data, and the complex interaction between plants and the surrounding environment (Pasman and Fabiano, 2021), the support to expert knowledge of AI in monitoring, learning, predicting, and effectively responding is essential for building resilience, which is here declined as the ability to thrive in adversity, i.e., to absorb fluctuations without diminishing performance. More recently, the integration of prior knowledge of physical processes with machine learning attracted interest to improve ML robustness and capability in identifying early signals of hazardous deviations (Vairo et al., 2023). Despite the above-mentioned advantages of data-driven models, their actual applicability remains limited, due to a certain lack of explainability and, in some cases, a lack of interpretability of the results. Several factors can affect a machine learning model, including the quality and diversity of the training data, the complexity of the model, and the presence of overfitting or underfitting. To assess the resilience of complex systems through a machine learning model, it is important to carefully select and pre-process the training data, choose an appropriate model architecture, and use techniques such as regularization and cross-validation to prevent overfitting and improve generalization, as real-world environments are often complex and dynamic, and the model may fail to deliver accurate or reliable results. The issues of explainability and trustworthiness of Machine Learning (ML) models still represent an open challenge. In fact, by nature, AI can be considered a black box and trusting a ML agent involves opening the box to an extent related with its intended use (Samek and Müller, 2019). Four building blocks are identified for addressing the challenges of data-driven learning approaches (EASA, 2021), namely: AI trustworthiness analysis, learning assurance, AI explainability, and AI safety risk mitigation.

The *trustworthiness analysis* serves as a gate to the three other technical building blocks and encompasses the safety and security assessments, which are key elements of the trustworthiness analysis concept. All three assessments (safety, security and ethics-based) are not only preliminary steps but also integral processes towards approval of such innovative solutions. *Learning assurance* is intended to cover the paradigm shift from programming to learning, as existing development assurance methods are not suitable to cover learning processes specific to AI/ML. *AI explainability* deals with the capability to provide human with understandable and relevant information on how an AI/ML application is coming to its results. The *AI safety risk mitigation* block considers that we may not always be able to open the 'AI black box' to the extent required and that the safety risk may need to be addressed to deal with the inherent uncertainty of AI. The knowledge on AI explainability is still very scattered and deserve further investigation (Vilone and Longo, 2021), due to the prominent role that ML/AI are expected to gain in the next future. Starting from these premises, this paper focuses on the *learning assurance* issue, by "opening" the black box of the ML algorithm and providing answers to the following research challenges:

I. How does the framework fulfil the key requirements of learning process management, learning process verification and inference model verification?
II. Can the combination of expert knowledge and data-driven approach overcome critical issues related to data-driven models?
III. How to improve system resilience by supporting the decision-making process with the constant updating of the Hidden Markov Model (HMM) based on new observations?

2. Methodology

In a previous work by Vairo et al. (2021), a preliminary approach, based upon data-driven modelling, was outlined to assess system resilience by the identification of precursor events, i.e., referring to early detection of "system weak" signals during the operations. In the following, the novel framework and applied algorithms are detailed.

2.1. The Machine Learning framework

The logic diagram for the proposed resilience assessment framework, in terms of stepwise procedure, is depicted in Fig. 1, where the systems resilience capabilities (monitor, learn, anticipate, and respond) can be pointed out.

The model integrates into its sequential structure the four steps identified to achieve the specific resilience objectives, by employing the proposed framework. The focal point of ML models is represented by data investigation: all relevant dependencies, correlations and inference statistics can be extracted by building up a reliable data-driven model. Thus, it is possible to identify the significant perturbations and, by

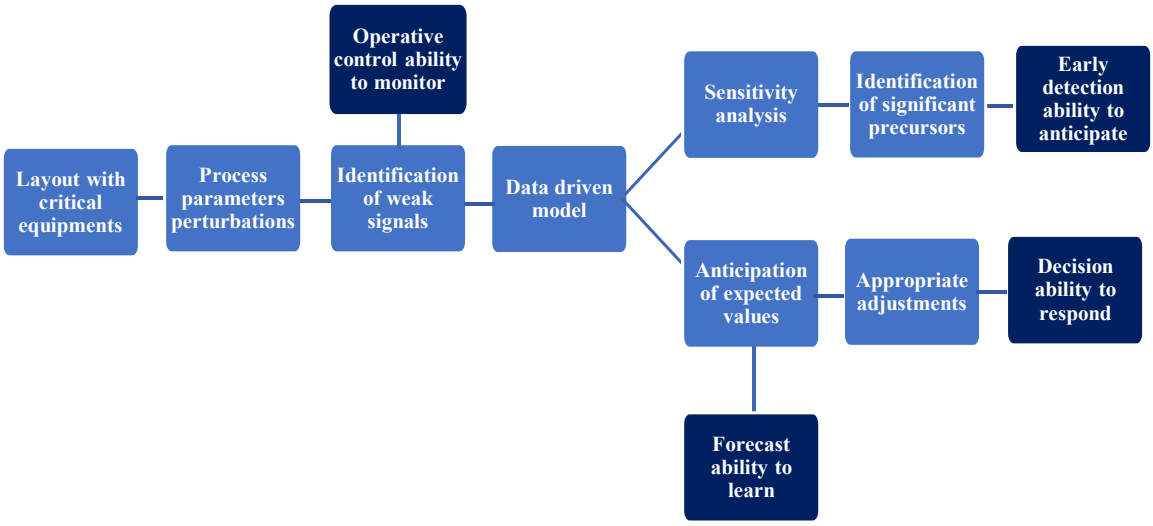

Fig. 1. The framework for resilience assessment.
(Adapted from Vairo et al., 2021).

training the model, anticipate the outcome of the system, to improve decision-making and promptly select the appropriate adjustments. Systems, by definition, deliver desired capability, while resilience addresses the delivery of such capability – in the face of adversity (Brtis and McEvilley, 2019). Systems interact with their environments and nominal environmental conditions often dominate the focus of Systems Engineering (SE) activities. The concept of resilience explicitly adds the assessment of adversity and requires a shift in the requirement analysis, architecture, and design methods to establish an approach addressing nominal and adverse conditions under which the system should operate. An influence diagram representing this meaning is depicted in Fig. 2.

The sources of adversity may be natural, technological, or human, and may include sources external to, or within the system. A high-level view of the steps to assess systemic resilience should include in-depth knowledge of the following items:

1. system architectures and/or designs;
2. system functional behaviour, data and control flows to deliver the required capability;
3. capabilities of interest, how are measured, and the required levels of delivery;
4. adversities that may affect the system;
5. system behaviour in response to adversities.

The term capability represents the system's ability to achieve the desired effects. This provides an umbrella term for considering many objectives and outcomes achieved by SE activities that are relevant to resilience, such as mission objectives, user needs, user requirements, system requirements, derived requirements, etc. The seven pillars of SE approach (Boardman and Sauser, 2008) are summarized as follows:

6. *Life Cycle*: it is by definition the time period between the concept phase and retirement.
7. *Gates*: are intended to ensure a safe progression along the project life cycle by providing intermediate checks during the development.
8. *Requirements*: connections between the problem space and the solution space.
9. *Perspectives*: integration of the different stakeholder's viewpoints in the early design stage.
10. *Trade-offs*: the project is a matter of decisions. The goal is to find the optimal solution satisfying all requirements.
11. *Modelling and simulations*: to forecast the efficiency and performance of the system.
12. *Operational effectiveness*: must be ensured in a long-term vision of the project.

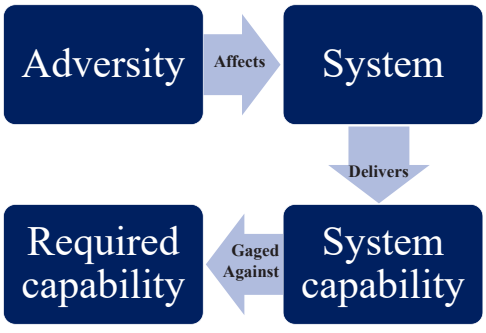

Fig. 2. Influences in Systems resilience.
Adapted from Brtis and McEvilley (2019)).

A broader view of the system can be attained according to the well-known V-model (Fig. 3), which is a useful guideline to manage the project, starting from the definition (the system is right), until the fulfilment of the defined requirements (it is the right system).

As depicted in Fig. 3, two sides can be identified. The left-hand one represents the analytical approach to the problem. Here a complex system is divided into sub-systems, which can be easily managed. In the right-hand side of Fig. 3, the sub-systems are combined back, so the whole system can be validated for ensuring that the requirements are met. The overall process corresponds to a stepwise sequence, namely: design, detail design, implementation, verification, and validation. Verification represents a critical item, as it ensures that all the steps to reach the goal have been well developed, testing as well whether both the assembled sub-systems meet the requirements and the whole system fulfils the designed performance.

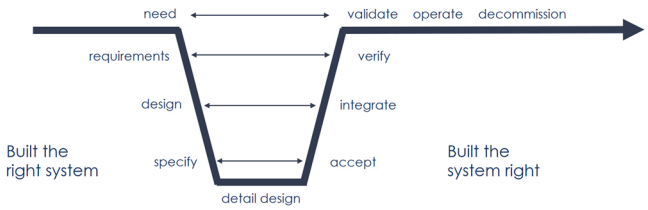

Fig. 3. The V-model based on the seven pillars of System Engineering.

2.2. The W-model

Recent enormous progress in ML has been made possible partly due to a simultaneous increase both in the amount of data available and in computational power even though at the burden of more complexity in ML models, possibly posing challenges in safety-critical domains. The most relevant challenges concerning ML trustworthiness can be summarized as follows (EASA, 2020):

1. traditional Development Assurance frameworks are not suitable for machine learning;
2. difficulties in keeping a comprehensive description of the intended function;
3. lack of predictability and explainability of the ML application behaviour;
4. lack of guarantee of robustness and of no 'unintended function';
5. lack of standardized methods for evaluating the operational performance of the ML applications;
6. the issue of bias and variance in ML applications;
7. complexity of architectures and algorithms;
8. adaptive learning processes.

As shown in Fig. 4, the modified W-model, here introduced starting from the above considerations, is focused on the selected key element, i.e., the learning process.

The steps of the W-shaped process, thoroughly investigated in the remainder of this paper are:

1. *Learning process management*, which includes all the steps required before the training: metrics, strategy to use for model selection, models/architectures to evaluate as well as the setup of software/hardware environment where the actual training takes place.
2. *Learning process validation*, where the outcome of the previous step, a single trained model, is evaluated on the test dataset. This evaluation includes understanding generalizability (performance guarantees) and failure cases, which can then be fed into a safety assessment.
3. *Inference model verification* and integration, where the desired properties of the deployed model are verified.

2.3. The resilience scenario

Scenarios are a useful way to represent the system needs, by describing the effect to be achieved and the reference environment and establishing the baseline for the measures, targets, and conditions (including adversities) by which acceptable capabilities will be judged. To be achieved and ensured, resilience must be effectively represented as a *system requirement*. The challenge is that resilience lumps the concepts of functional, performance and environmental requirements. This compound requirement must be captured, so standard SE practices can trade system resilience against other system properties expressed in the system requirements. Thus, specifying resilience requires that several parameters, whose aggregation is the resilience scenario, be identified, as detailed in the following (Jackson and Ferris, 2012):

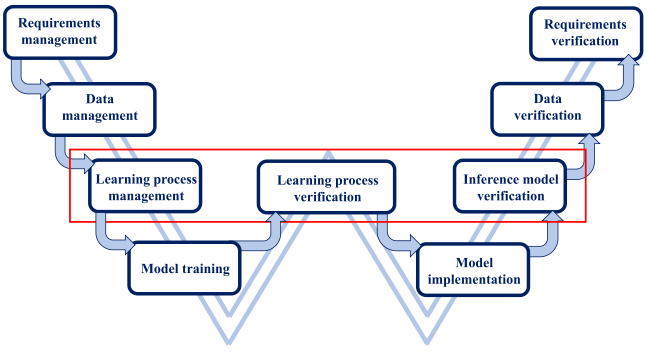

Fig. 4. The modified W-model.
Adapted from EASA (2021).

1. The capability of interest (note: a system may deliver several capabilities each of which may have different levels of resilience).
2. The measure(s) (and units) of the capability.
3. The target(s) (required amount) of the capability.
4. System modes of operation (e.g., operational, contingency, training, exercise, maintenance, update).
5. The adversity(s) being considered for this resilience scenario and the affecting level that the adversities can impose on the system.
6. Understanding of the effects that the adversity imposes on the system and how the system reacts to those effects in terms of its ability to deliver capability.
7. The timeframe of interest.
8. The required resilience (performance) of the capability in the face of each identified resilience scenario (e.g., expected availability, maximum allowed degradation, the maximum length of degradation, etc.).

Any of these factors and parameters may vary over the timeframe of the scenario, and this fact must be addressed by the systems engineer. The capability is likely to be a functional requirement of the system. Resilience then extends such requirements into a resilience scenario by adding environmental requirements (adversities) and performance requirements. For defining an applicative scenario and applying the abovementioned concepts of validation, verification, and assurance to the Resilience Assessment framework, we consider an LNG bunkering operating plant, where the HMM is the ML model to be verified. The selection is motivated also by the ambitious middle term goals defined by the RePower EU to manage the energy security supply, caused by the Ukraine war. Table 1 identifies the modelling information that needs to be captured during the various lifecycle stages to support the effective development and documentation of resilience scenarios and resilience requirements (Brtis and McEvilley, 2019). This modelling information is the main point of the proposed resilience assessment framework depicted in the already mentioned Fig. 1.

2.3.1. System description: the physical system

The LNG transfer unit is equipped with:

1. quick-release hooks;

Table 1
System safety requirements at each life-cycle stage.

Life cycle stage	Information	Resilience Assessment
Mission and Stakeholder Needs Analysis	Insert adversities in the context diagram as actors. Insert resilience scenarios as use cases.	Operative control: ability to monitor
Stakeholder Requirements	Develop use case interaction diagrams to document the interaction of actors and architectural modules during the scenarios. Develop sequence diagrams to represent the activity flow during scenarios.	Decision: ability to respond
System Requirements	Develop activity diagrams to show the states of the system (and adversities) during scenarios.	Early detection: ability to anticipate
Architecture and System Design	Develop state models of the scenarios. Model events and signals among the architectural nodes.	Forecast: ability to learn

2. fenders;
3. dock monitoring system to check the ship's position and speed of approach, weather and sea conditions;
4. pier control room.

The quick-release hooks will be installed on the dock. All hooks are capable of moving both vertically and horizontally and each of them is designed to be released independently of the other. The pier control room is equipped with control apparatuses for the emergency stop of the LNG transfer, for the release of the LNG transfer connection for the remote control of the fire extinguishing system. The Ship-to-Shore connection is used to reciprocally exchange Emergency Shut-Down (ESD) between the ship and the ground system.

The connection between the ship and the plant takes place via a loading arm, with two independent lines: one for the liquid phase (LNG) discharged from the ship to the plant and a flexible line for the gas phase (steam return from the plant to the ship); vice versa, the steam to the plant and the LNG to the ship, during the bunkering of a barge.

The loading arm is equipped with all control and safety devices with critical elements identified as follows:

5. a quick release system (PERC);
6. a PLC, dedicated to the loading arm and connected to the plant control system, integrated into the Hydraulic Processing Unit;
7. the arm will be connected to the ship by means of 2 flanged connections, one for the liquid and one for the vapour.

The operations of connection/disconnection of the loading arm are monitored through the control system (pressure gauges and thermometers). The liquid and the vapour line, made of low-temperature carbon steel, have respectively nominal diameters of 10″ and 8″.

As argued by Zarei et al. (2017), the most critical factors affecting performance variability represent a mandatory issue for safety and resilience application in complex systems, which cannot rely only on prior or posterior probabilities. The main causes of loss of containment during bunkering reside in the coupling operation of the bunkering manifold to the receiving vessel and are due to damage to the connection pipe during normal operations and SIMOPS (simultaneous operations). During bunkering operations, a loss of containment can occur in different sections of the process. In particular, the situations that can lead to a loss of containment concern failures of critical equipment and failures of the receiving vessel.

2.3.2. System description: the ML system

HMM is a statistical Markov model in which the system being modelled is assumed to be a Markov process –X – with unobservable ("hidden") states (Sipos, 2016). HMM assumes that there is another process Y whose behaviour "depends" on X. The goal is to learn about X by observing Y.

HMM stipulates that, for each time instance n_0, the conditional probability distribution of $Y_{(n0)}$ given the history $\{X_n = x_n\}_{n<=n0}$ must not depend on $\{x_n\}_{n<=n0}$ (i.e., it is a Markov process).

The probabilistic parameters of an HMM are:

1. X - Hidden states;
2. y - possible observations;
3. a - state transition probabilities;
4. b - output probabilities.

Here, the hidden states, are the states between a regular performance and a failure of a sub-system. The only known states are the first (the component is performing well) and the last (the component fails), and the hidden states in between may represent the precursors of accidental events. The possible observations of the system are the process variable values.

3. Results and discussions

3.1. Explaining the learning process

The core element of the proposed approach is the explainability, validation, and verification of the learning process (the red square in Fig. 4), which is performed by the Baum-Welch algorithm.

In such a model, two parts must be trained: the Markov Chain and the observations.

When no knowledge is available, the Baum-Welch algorithm can be used to fit an HMM. However, the Baum-Welch algorithm does not always give the right answer. So, for ensuring the training process, some knowledge must be added to the process.

An HMM is structured into two parts:

1. an underlying Markov Chain that describes the logical sequence of the system states. This underlying state is the element of interest. If there are k states in the HMM, then the Markov Chain consists of:
   - a k × k matrix describing the probabilities for the system to shift from a state S1 to a state S2;
   - a k-length vector describing the probabilities to start off in each of the states;
2. a probability model allowing to compute P[Observation|State], the probability of seeing an observation O if it is assumed that the underlying state is S. Unlike the Markov Chain, which has a fixed format, the model for P[Observation|State] can be arbitrarily complex.

The external expert knowledge can add a-priori information on the transition probability. To a large degree, these two moving parts can be considered independently.

After the phase of labelled data collection, (the sequence of observations and a knowledge of what the underlying state is), training the HMM breaks down into two independent problems:

3. firstly, train the Markov Chain with the labels;
4. secondly, divvy up the observations based on what state they were in and train P[Observation|State] for each state.

If the state labels for our data are reliable, then training the HMM is straightforward. However, usually we just have the sequence of observations with only a little knowledge regarding the actual system state. So, we can guess what the state labels are and train an HMM using those guesses. Then, we use the trained HMM to make better guesses at the states and re-train the HMM on those better guesses. This process continues until the trained HMM goes to convergence. This back-and-forth, between using an HMM to guess state labels and using those labels to fit a new HMM, is the core of the *Baum-Welch algorithm*. In order to test the capability of the framework, we refer to the real-case scenario of LNG bunkering, where the relevant observations are: pressure, temperature and flow rate. A simplified layout of the plant section is represented in Fig. 5.

The investigation will be limited to the leakage hazard originating in the part of the system between the two flanges of the connecting hose, technically indicated as "LNG transfer system", from v_1 to v_2. Table 2 summarizes the operating parameters and standard values during operations.

The bunkering operations can maintain a constant temperature by managing the boil-off vapours. Thus, pressure is the relevant parameter to be monitored for inferring the system state.

The probability of seeing all the observations so far, given a time *t* and a hidden state *i,* can be defined as follows:

$$\alpha_i(t) = P\ (Y_1 = y_1, \ldots, Y_t = y_t, X_t = i|\theta) \quad (1)$$

Additionally, given a time *t* and a hidden state *i*, the probability of predicting all the future observations is:

$$\beta_i(t) = P\ (Y_{t+1} = y_{t+1}, \ldots, Y_T = y_T, X_t = i,\ \theta) \quad (2)$$

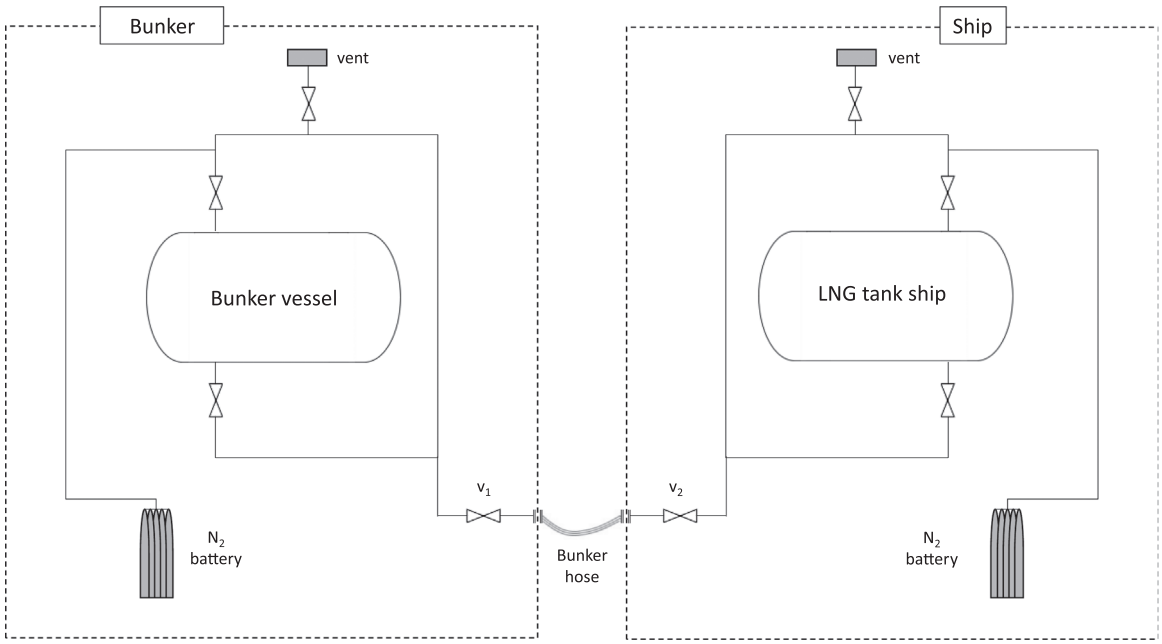

Fig. 5. Schematization of the system layout.

Table 2
Operative parameters.

Parameter	Operative value
Pressure	5–6 bar(g)
Temperature	-162 °C
Transfer rate	600 m^3/h

Furthermore, given the observations, the emission probability at a given state *i* at time *T* is:

$$\Upsilon_i(t) = P(X_T = i|Y, \theta) \quad (3)$$

And, similarly, the transition probability from state *i* to state *j* at time *t* is:

$$\xi_{ij}(t) = P\ (X_t = i,\ X_{t+1} = j|Y,\ \theta) \quad (4)$$

To learn the HMM model, we need to know which states best describe the observations. The emission probability Υ, namely the probability of state *i* at time *t* given all the observations, accounts for the previous reasoning. Once the distribution of Υ and ξ (θ) are refined, it is possible to perform a point estimate (θ: α, β) on what will be the best transition and emission probability.

HMM learning process is schematically depicted in Fig. 6: one set of parameters is fixed to improve others and the iteration continues until the solution converges. Fig. 7 shows a graphical example of the attained best explanation of observations.

Within the context of the LNG bunkering scenario, the system safety is undoubtedly the required capability. Safety is the required emerging property the system must have, to ensure the desired performances (accident prevention, environmental protection, occupational health &

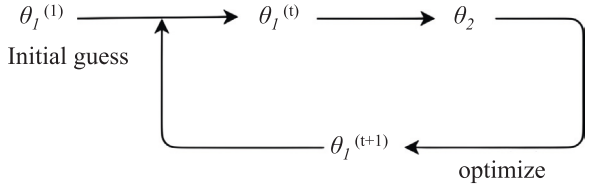

Fig. 6. Graphical schematization of the HMM learning process.

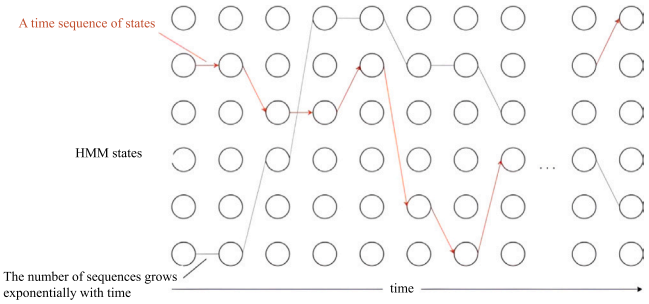

Fig. 7. The best explanation for observations.

safety, production quality, regulatory compliance, etc.) in a sustainable way. Thus, an appropriate resilience metric, accordingly to Brtis and McEvilley (Brtis and McEvilley, 2019), is the expected availability of required capability. On this basis, three possible hidden states have been defined:

1. STATE 1: the system delivers the required capability;
2. STATE 2: the system delivers a portion of the required capability;
3. STATE 3: the system is no longer able to deliver the required capability.

3.1.1. *First guess*

As a first guess, to determine Υ and ξ (emission and transition probabilities) according to the described learning process, appropriate set points for the process variables (from expert knowledge) should be defined as follows: pressure alarm at 6.8 bar, pressure block at 8 bar.

So, the first guess is that the system states are:

1. STATE 1: P < 6.8 bar
2. STATE 2: 6.8 bar < P < 8 bar
3. STATE 3: P > = 8 bar.

3.1.2. *Learning*

As previously outlined, the Baum-Welch algorithm, classified as an expectation-maximization (EM) algorithm, is the selected learning tool,

calculating each time-step probability of being in each state. The HMM is trained to be on average the best fit over all the probabilistic guesses. In the *expectation* step, the forward and the backward formulae give the expected hidden states according to the observed data; the *maximization* step updates formulae and then tunes the parameter matrices to best fit the observed data and the expected hidden states. Subsequently, these two steps are iteratively repeated, by a Markov Chain Monte Carlo (MCMC) sampling, until attaining parameters convergence, or until the model has reached the required accuracy requirement.

The issue of the Baum-Welch algorithm is the same drawback affecting the EM algorithm in general, i.e., solutions that are only locally optimal.

There are several ways to mitigate this issue:

1. simplification of the HMM. The number of local optima can grow exponentially with the number of parameters in the HMM. If the parameter size is reduced, the number of local optima reduces as well. If the observations are multinomial, it could mean reducing the number of possible observations;
2. use field-knowledge intuition to have an initial HMM that is similar to what the outcome is expected to be. Different local optima often have large qualitative differences from each other. On a gross level, the HMM that ultimately converges is likely to resemble the first one and to be consistent with knowledge-based intuitions.

   Several different convergence criteria can be used to determine when the Baum-Welch algorithm has converged to a satisfactory solution. In the model development, a combination of the below criteria is adopted by setting both a maximum number of iterations and a threshold on the change in the log-likelihood.
3. A fixed number of iterations: the algorithm can be run for a fixed number of iterations and then stopped, regardless of whether convergence is achieved.
4. A termination criterion on the log-likelihood: the algorithm can be stopped when the change in the log-likelihood between iterations falls below a given threshold.
5. A termination criterion on parameters: the algorithm can be stopped when the change in the estimated parameters between iterations falls below a given threshold.

The goal is to design a model that is simple but not simpler, i.e., suitable to mitigate local optima but complicated enough to describe the investigated process. Additionally, this process already contains in its framework the sensitivity analysis, which conversely represents a pivotal step in approaches using simple Bayesian networks (Chang et al., 2018).

3.1.3. Prediction

As shown in Table 3, once the parameters of the HMM are learned from data, given a partially observed data sequence, the posterior distribution over the hidden states is inferred. This is a filtering task that can be carried out using the forward algorithm. This posterior distribution enables uncovering the hidden state of the system, as data streaming in is observed. Consequently, it is possible to compute the most probable state sequence path, which is updated with newly observed data (Hofmann and Tashman, 2020).

As new data are streaming in, an estimate of the state over a certain future horizon is calculated as well as the most probable time at which the system will enter the *failure* state (terminal state).

Table 3
Dataset head.

Time	Pressure (Barg)	Temperature (°C)
2020–10–21 10.10.00.000	5.88	− 162.02
2020–10–21 10.15.00.000	5.91	− 162.03
2020–10–21 10.20.00.000	5.99	− 162.05

Those inferential tasks are important as they provide a picture of the current state of the system, as well as a forecast of when each sub-system will most likely fail. This information will then be used to optimize the decision-making process, modify some process parameters or interrupt the process itself. The cross-validation method is adopted to validate the performance of the Baum-Welch algorithm: the algorithm is trained on a portion of the verified data and tested on a held-out portion. This method can provide a more robust estimate of the model's performance, as it allows for a more realistic assessment of the model's ability to generalize to new data. In terms of loss, the Baum-Welch algorithm does not explicitly minimize a loss function, likewise some other ML algorithms. Instead, it iteratively updates the parameters of the HMM aiming at maximizing the likelihood of the observations. The log-likelihood of the observations can be thought of as a measure of the "loss" or discrepancy between the observations and the model, with a higher log-likelihood indicating a smaller loss. As the algorithm runs, the log-likelihood should generally increase, indicating that the model is becoming a better fit for the data.

In Fig. 8, the prediction of the future system states sequence is shown in form of immediate readability. The future time steps are s1, s2 and s3, after 5, 10, and 15 min respectively from the current time. Results indicate that the system will be in STATE 2 (as previously defined) after 5 min, then, after 10 min, it will enter the failure state (STATE 3), when it is no longer able to deliver the required capability. Table 4 shows the predicted state sequences given the observation on system pressure.

Table 5 provides the first state transition prediction for each system subcomponent and the corresponding update of failure probability. The bold value is the state that has the highest expected probability, according to the Baum-Welch HMM formulation.

3.1.4. From prediction to action

Choosing the best action requires considering not only immediate effects but also long-term effects, which are not known in advance. Sometimes, actions with poor immediate effects can have better long-term ramifications, so an optimal decision requires the right trade-off between immediate effects and future rewards, dynamically connected to variable uncertainty. The goal of the optimal decision is to determine the best action to take for ensuring the desired system capability at any given point in time, given the uncertainty of current and future states.

The constant updating of the HMM with the new observations allows predicting the short- and long-term effects of the actions. Fig. 9 graphically shows the predicted state sequence at the current time.

Analogously, Fig. 10 shows the updated sequence prediction after decreasing the LNG bunkering rate.

3.1.5. Defining system requirements

As a core consideration, system resilience must be effectively represented as a system requirement. The challenge is that the concept of *system safety* aggregates the considerations of functional, performance and environmental requirements. This compound requirement must be captured, so standard system engineering practices can trade resilience against other system properties expressed in the system requirements. Any of these factors and parameters may vary over the timeframe of the scenario and the systems engineer must address this evidence. The capability is likely to be a functional requirement of the system. Resilience then extends such requirements into a scenario by adding environmental requirements (adversities) and performance requirements (Brtis and McEvilley, 2019).

Table 6 summarizes the resilience assessment outputs for each life cycle stage to demonstrate how the resilience assessment framework responds to system safety requirements introduced in Table 1.

4. Conclusions

To ensure Machine Learning models work as intended, what is needed can be pointed out in correctly designed data, correctly designed

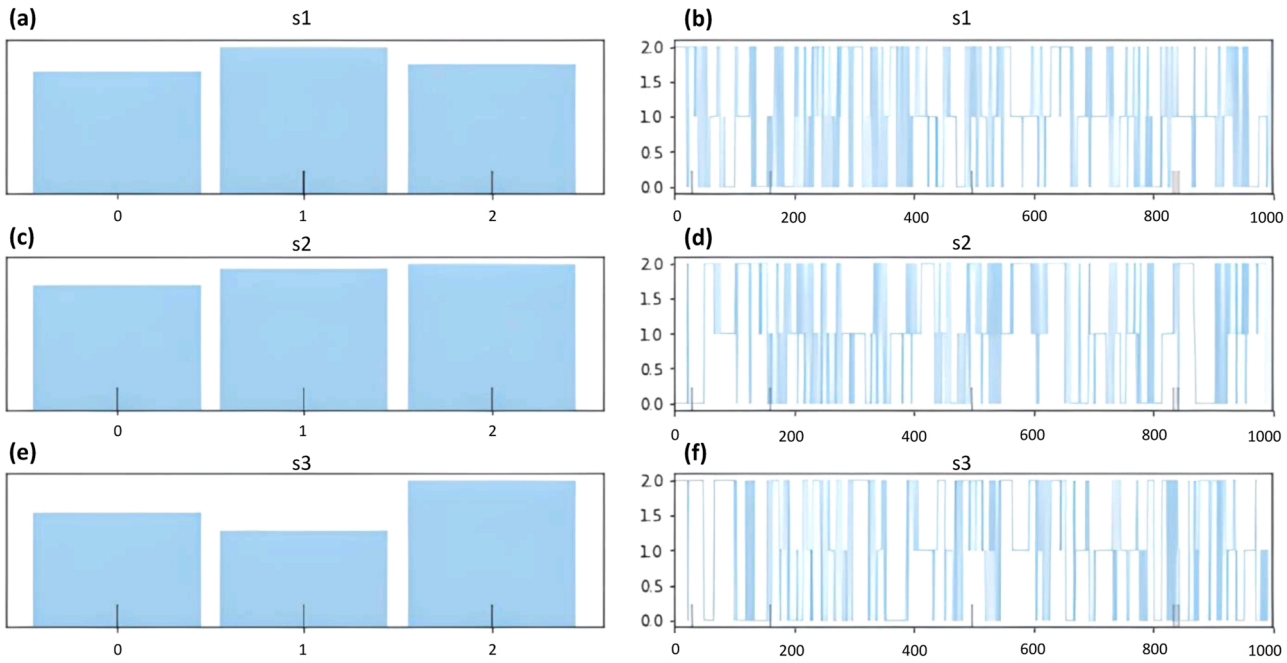

Fig. 8. Prediction of the future system state. (a), (c), (e) represent the most probable expectations obtained from the inferential samplings given in (b), (d), and (f) respectively, where each different state is reported as a function of the number of samplings.

Table 4
Predicted sequences given the observed pressure values.

Observed sequence	Highest probability of observing sequence 1–2 if state is 1	Highest probability of observing sequence 2–3 if state is 2	Highest probability of observing sequence 2–2 if state is 2	Highest probability of observing sequence 1–1 if state is 1	Predicted sequence
P1-P1	0.4127	0	0	0.5873	1–1
P1-P1	0.3591	0	0	0.6409	1–1
P1-P2	0.4647	0	0	0.5353	1–1
P1-P1	0.4899	0	0	0.5101	1–1
P1-P1	0.3561	0	0	0.6439	1–1
P1-P1	0.3125	0	0	0.6875	1–1
P1-P2	0.5782	0	0	0.4218	1–2
P2-P2	0	0.4099	0.5901	0	2–2
P2-P3	0	0.4147	0.5853	0	2–2
P2-P3	0	0.5026	0.4974	0	2–3

learning process and correctly designed model. Each of these three elements may become a deal breaker if not correctly performed and properly verified. A combination of expert knowledge and data-driven methods is required for designing the right model, identifying local optima stuck, and interpreting the output for translating it into actions. The main contribution of this research is the development of a SE lifecycle framework based on the definition of system resilience through a series of system requirements that allow the system to deliver the required capability. The explored case study, of actual interest under the drivers of energy transition, climate change and war crises, exhibits high potentiality to model the performance of a complex system in terms of system resilience and safety. The abductive inferential framework, as declined in the proposed Hidden Markov Model, proved to be a very useful tool to deal with incomplete observations and update the prediction at each newly observed data. The approach is by definition explainable, being strongly model-based, rather than strictly data-driven. It stands to reason that the data-driven part of the model is fundamental to obtaining reliable predictions through the learning process, accomplished by properly applying the customized Baum-Welch algorithm. As recently pointed out (Yu et al., 2022) when dealing with real complex operations, a clear pre-definition of acceptable thresholds depending on the consequence selected hazards and cost implications is required and this item needs further refinement of the here presented approach. Predictive models and resilience engineering are two fields that often intersect, as predictive models can be used to identify potential failures or disruptions in complex systems and to develop strategies for mitigating their impact. To be effective in resilience engineering, predictive models need to be accurate and reliable, as well as resilient themselves. This means that the models should be able to continue performing well even in the face of changing circumstances or unexpected inputs. The current limitation of the model is that the prediction of the system's states is not directly associated with a specific critical variable (apart from the fact that in the case examined, the critical parameters were few), and therefore, future developments go towards integrating reinforcement learning techniques into the model, so that it is not only able to predict future states, but also to propose, and possibly implement, the most appropriate actions to intercept deviations. Overall, predictive models can be a valuable tool in resilience engineering, helping to early identify potential failures and disruptions and to develop strategies for mitigating their impact.

As a concluding remark, even if further validation on complex systems and real case studies is required and upon further refinement, the proposed framework may represent a fundamental tool to support the decision process for achieving system safety and designing resilient systems, under the requirement of explainability, verification and validation.

Table 5
Actual and predicted state, and updated failure probability.

Root Component	Actual state t0 (prob.)	Updated failure prob.	Predicted State t1 (prob.)
SHORESIDE VALVES			
Safe	**0.721**	1.6×10^{-8}–1.6×10^{-5}	0.455
Intermediate	0.222	(94%HPD)	**0.501**
Fail	0.047		0.044
PUMP			
Safe	**0.633**	1.8×10^{-8}–1.8×10^{-5}	**0.527**
Intermediate	0.166	(94%HPD)	0.301
Fail	0.201		0.172
SHORESIDE PIPELINE			
Safe	0.487	1.7×10^{-8}–1.7×10^{-5}	0.392
Intermediate	**0.491**	(94%HPD)	**0.487**
Fail	0.022		0.121
HOSE			
Safe	**0.355**	5.7×10^{-9}–1.8×10^{-5}	0.329
Intermediate	0.331	(94%HPD)	**0.481**
Fail	0.314		0.190
SHIPSIDE VALVES			
Safe	0.212	8.7×10^{-9}–1.7×10^{-5}	0.188
Intermediate	**0.692**	(94%HPD)	0.396
Fail	0.096		**0.416**
SHIPSIDE PIPELINE			
Safe	**0.411**	1.2×10^{-8}–1.7×10^{-5}	0.350
Intermediate	0.307	(94%HPD)	**0.399**
Fail	0.282		0.251

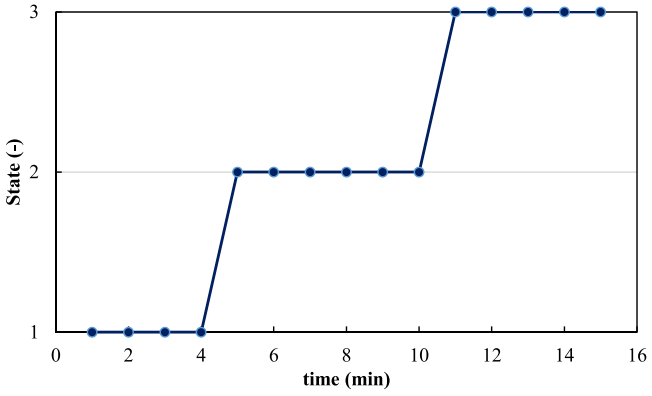

Fig. 9. Current time sequence prediction in the hypothetical scenario.

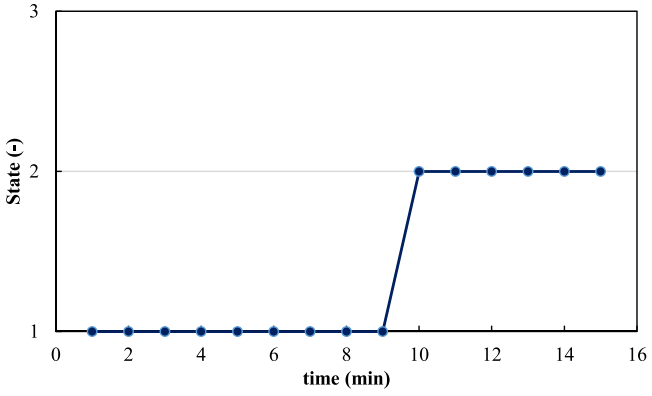

Fig. 10. Sequence prediction after corrective action in the hypothetical scenario.

Table 6
Resilience assessment outputs for each life cycle stage.

Life cycle stage	Resilience Assessment Output
Mission and Stakeholder Needs Analysis	The adversities are inserted by identifying the perturbations.
	The resilience scenarios are inserted by evaluating the weak signals in the safety assessment context.
Stakeholder Requirements	The use case interaction diagrams are rooted in the anticipation of expected values.
	The activity flow describes how the anticipation capability influences the safety assessment.
System Requirements	The system states are predicted by the ML system.
Architecture and System Design	States, event, and signals represent the training datasets for design and develop the dynamic model.

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 was funded by INAIL, within the frameworks of the call BRIC/2019/ID 02 (Project DYN-RISK) and of the call BRIC/2021/ID03 (Project DRIVERS).

References

Arunthavanathan, R., Khan, F., Ahmed, S., Imtiaz, S., 2021. A deep learning model for process fault prognosis. Process Saf. Environ. Prot. 154, 467–479. https://doi.org/10.1016/j.psep.2021.08.022.

Aven, T., 2022. On some foundational issues concerning the relationship between risk and resilience. Risk Anal. 42, 2062–2074. https://doi.org/10.1111/risa.13848.

Boardman, J., Sauser, B., 2008. Systems thinking. Coping with 21st century problems. Taylor & Francis, Boca Raton.

Brtis, J., McEvilley, M.A., 2019. Systems engineering for resilience, MP1909495. Syst. Eng. Aerosp. https://doi.org/10.1016/c2018-0-00485-5.

Chang, Y., Chen, G., Wu, X., Ye, J., Chen, B., Xu, L., 2018. Failure probability analysis for emergency disconnect of deepwater drilling riser using Bayesian network. J. Loss Prev. Process Ind. 51, 42–53. https://doi.org/10.1016/j.jlp.2017.11.005.

Dinh, L.T.T., Pasman, H., Gao, X., Mannan, M.S., 2012. Resilience engineering of industrial processes: principles and contributing factors. J. Loss Prev. Process Ind. 25, 233–241. https://doi.org/10.1016/j.jlp.2011.09.003.

EASA, 2020. Concepts of design assurance for neural networks (CoDANN). Public Rep. Extr. Version 1, 0.

EASA, 2021. EASA Concept Paper: First usable guidance for Level 1 machine learning applications- A deliverable of the EASA AI Roadmap 0–144.

Fabiano, B., Hailwood, M., Thomas, P., 2022. Safety, environmental and risk management related to Covid-19. Process Saf. Environ. Prot. 160, 397–399. https://doi.org/10.1016/j.psep.2022.02.035.

Ghosh, A., Ahmed, S., Khan, F., Rusli, R., 2020. Process safety assessment considering multivariate non-linear dependence among process variables. Process Saf. Environ. Prot. 135, 70–80. https://doi.org/10.1016/j.psep.2019.12.006.

Guo, X., Ji, J., Khan, F., Ding, L., Tong, Q., 2021. A novel fuzzy dynamic Bayesian network for dynamic risk assessment and uncertainty propagation quantification in uncertainty environment. Saf. Sci. 141, 105285 https://doi.org/10.1016/j.ssci.2021.105285.

Hofmann, P., Tashman, Z., 2020. Hidden markov models and their application for predicting failure events. In: Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J. (Eds.), Computational Science – ICCS 2020. Springer, Cham, Switzerland, pp. 464–477. https://doi.org/10.1007/978-3-030-50420-5_35.

Jackson, S., Ferris, T.L.J., 2012. Resilience principles for engineered systems. Syst. Eng. 14, 305–326. https://doi.org/10.1002/sys.

Jain, P., Rogers, W.J., Pasman, H.J., Keim, K.K., Mannan, M.S., 2018. A Resilience-based Integrated Process Systems Hazard Analysis (RIPSHA) approach: part I plant system layer. Process Saf. Environ. Prot. 116, 92–105. https://doi.org/10.1016/j.psep.2018.01.016.

Kamil, M.Z., Taleb-Berrouane, M., Khan, F., Amyotte, P., 2021. Data-driven operational failure likelihood model for microbiologically influenced corrosion. Process Saf. Environ. Prot. 153, 472–485. https://doi.org/10.1016/j.psep.2021.07.040.

Linkov, I., Trump, B., 2019. The science and practice of resilience. Springer Cham, Amsterdam, NL.

Liu, Z., Ma, Q., Cai, B., Liu, Y., Zheng, C., 2021. Risk assessment on deepwater drilling well control based on dynamic Bayesian network. Process Saf. Environ. Prot. 149, 643–654. https://doi.org/10.1016/j.psep.2021.03.024.

Mamudu, A., Khan, F., Zendehboudi, S., Adedigba, S., 2021. Dynamic risk modeling of complex hydrocarbon production systems. Process Saf. Environ. Prot. 151, 71–84. https://doi.org/10.1016/j.psep.2021.04.046.

Meng, H., An, X., Xing, J., 2022. A data-driven Bayesian network model integrating physical knowledge for prioritization of risk influencing factors. Process Saf. Environ. Prot. 160, 434–449. https://doi.org/10.1016/j.psep.2022.02.010.

Nhat, D.M., Venkatesan, R., Khan, F., 2020. Data-driven Bayesian network model for early kick detection in industrial drilling process. Process Saf. Environ. Prot. 138, 130–138. https://doi.org/10.1016/j.psep.2020.03.017.

Paltrinieri, N., Comfort, L., Reniers, G., 2019. Learning about risk: machine learning for risk assessment. Saf. Sci. 118, 475–486. https://doi.org/10.1016/j.ssci.2019.06.001.

Pasman, H.J., 2021. Early warning signals noticed, but management doesn't act adequately or not at all: a brief analysis and direction of possible improvement. J. Loss Prev. Process Ind. 70, 104272 https://doi.org/10.1016/j.jlp.2020.104272.

Pasman, H.J., Fabiano, B., 2021. Highlights and an impression of process safety evolutionary changes from the 1st to the 16th LPS present and future of the European loss prevention and safety promotion in the process industries. Process Saf. Environ. Prot. 147, 80–91.

Pawar, B., Park, S., Hu, P., Wang, Q., 2021. Applications of resilience engineering principles in different fields with a focus on industrial systems: A literature review. J. Loss Prev. Process Ind. 69, 104366 https://doi.org/10.1016/j.jlp.2020.104366.

Samek, W., Müller, K.-R., 2019. Towards explainable artificial intelligence. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, (eds.).,. Springer Cham, Switzerland, pp. 5–22. https://doi.org/10.1007/978-3-030-28954-6.

Sarbayev, M., Yang, M., Wang, H., 2019. Risk assessment of process systems by mapping fault tree into artificial neural network. In: J. Loss Prev. Process Ind, 60, pp. 203–212. https://doi.org/10.1016/j.jlp.2019.05.006.

Sipos, I.R., 2016. Parallel Stratified MCMC Sampling of AR-HMMs for Stochastic Time Series Prediction, in: 4th Stochastic Modeling Techniques and Data Analysis International Conference with Demographics Workshop (SMTDA2016). Valletta, pp. 295–306. https://doi.org/10.1038/ajh.2009.223.

Vairo, T., Reverberi, A.P., Fabiano, B., 2020. From risk assessment to resilience assessment. An application to a hazmat storage plant. Chem. Eng. Trans. 82, 151–156. https://doi.org/10.3303/CET2082026.

Vairo, T., Gualeni, P., Reverberi, A.P., Fabiano, B., 2021. Resilience dynamic assessment based on precursor events: Application to ship lng bunkering operations. Sustainability 13, 6836. https://doi.org/10.3390/su13126836.

Vairo, T., Cademartori, D., Clematis, D., Carpanese, M.P., Fabiano, B., 2023. Solid Oxide Fuel Cells for shipping: a Machine Learning model for early detection of hazardous system deviations. Process Saf. Environ. Prot. 172, 184–194. https://doi.org/10.1016/j.psep.2023.02.022.

Vilone, G., Longo, L., 2021. Notions of explainability and evaluation approaches for explainable artificial intelligence. Inf. Fusion 76, 89–106. https://doi.org/10.1016/j.inffus.2021.05.009.

Yang, M., Khan, F.I., Lye, L., 2013. Precursor-based hierarchical Bayesian approach for rare event frequency estimation: a case of oil spill accidents. Process Saf. Environ. Prot. 91, 333–342. https://doi.org/10.1016/j.psep.2012.07.006.

Yu, M., Pasman, H.J., Erraguntla, M., Quddus, N., Kravaris, C., 2022. A framework to identify and respond to weak signals of disastrous process incidents based on FRAM and machine learning techniques. Process Saf. Environ. Prot. 158, 98–114.

Zarei, E., Azadeh, A., Khakzad, N., Aliabadi, M.M., Mohammadfam, I., 2017. Dynamic safety assessment of natural gas stations using Bayesian network. J. Hazard. Mater. 321, 830–840. https://doi.org/10.1016/j.jhazmat.2016.09.074.