

Adapting FMECA for digital reliability assessment of soft sensors in drinking-water systems

Akrivi Alexandraki , Grigorios Kyritsakas , Martine van den Boomen & Luuk Rietveld

To cite this article: Akrivi Alexandraki , Grigorios Kyritsakas , Martine van den Boomen & Luuk Rietveld (2026) Adapting FMECA for digital reliability assessment of soft sensors in drinking-water systems, Digital Water, 4:1, 2725912, DOI: [10.1080/28375807.2026.2725912](https://doi.org/10.1080/28375807.2026.2725912)

To link to this article: <https://doi.org/10.1080/28375807.2026.2725912>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 06 Sep 2026.



Submit your article to this journal [↗](#)



Article views: 93



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



Adapting FMECA for digital reliability assessment of soft sensors in drinking-water systems

Akrivi Alexandraki^a, Grigorios Kyritsakas^a, Martine van den Boomen^b and Luuk Rietveld^a

^aDepartment of Water Management, Delft University of Technology, Delft, The Netherlands; ^bDepartment of Hydraulic Engineering, Delft University of Technology, Delft, The Netherlands

ABSTRACT

The digitalisation of water utilities introduces new operational risks that conventional asset-management frameworks do not fully capture. As utilities increasingly study, test and implement machine-learning-based soft sensors for real-time monitoring and process optimisation, the reliability of these digital assets becomes essential for maintaining safe drinking-water production. This study presents an adapted Failure Modes, Effects and Criticality Analysis (FMECA) for soft sensors, focusing on information-related failure modes, such as data loss, model drift and alarm malfunction. A modified Risk Priority Number (RPN) integrates detectability through a severity-reduction factor, enabling quantification of the mitigating effect of early fault detection. The framework is demonstrated using the CT (Concentration × Time) ozone-exposure soft sensor for disinfection control in Leiduin drinking-water treatment plant of Waternet, using design specifications and expert-informed assumptions. Ten failure modes were identified across data acquisition, preprocessing, model inference and output interpretation. Critical modes included missing Supervisory Control and Data Acquisition (SCADA) input, model drift and alarm failure due to their potential impact on disinfection reliability. Incorporating detectability reduced the RPN by up to 80%; this shows that improvements in information visibility reduce risk more effectively than marginal changes in likelihood or severity. The framework supports ISO 55000-aligned asset-management strategies within FIWARE-based decision-support platforms.

ARTICLE HISTORY

Received 14 August 2026

Accepted 19 August 2026

1. Introduction

Drinking-water utilities are undergoing digital transformation to strengthen resilience and compliance under stricter regulatory requirements and increasing operational complexity. Within the European Horizon ToDrinQ project, digital decision-support tools are being developed to support risk-based management strategies for drinking-water systems (2024). In this context, FIWARE-based architectures are used to enable interoperable, real-time data exchange between operational systems and digital applications (FIWARE Foundation n.d). These developments create a strong technological basis for deploying predictive decision support in water treatment, but they also introduce new operational risks that conventional asset-management approaches do not fully capture.

Water utilities increasingly study, test and implement machine-learning (ML)-based soft sensors, for estimating key water-quality and treatment parameters that are difficult or expensive to measure directly. Soft sensors have been widely discussed as enabling technologies for monitoring and optimisation, but their development and long-term robustness remain challenging, particularly due to dependence on data quality and operating conditions (Brunner *et al.* 2021). In water treatment, specifically, performance degradation over time due to changing regimes and data distribution shifts (“drift”) has been identified as a critical challenge that can undermine the reliability of model outputs if not detected and managed (Shen *et al.* 2025). As a result, soft sensors should be treated as digital assets, whose reliability must be assessed and maintained through structured practices rather than assumed through model accuracy alone.

Conventional reliability-centred and asset-management frameworks have been developed for physical assets, where failure manifests through mechanical degradation, wear or breakage (Hastings 2015). Classical reliability

CONTACT Akrivi Alexandraki ✉ A.A.Alexandraki@tudelft.nl Department of Water Management, Delft University of Technology, Stevinweg 1, Delft 2628 CN, The Netherlands

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

and quality methods, including Failure Modes, Effects and Criticality Analysis (FMECA), similarly focus on tangible component failures and historically observed failure rates (Dhillon 2007). In contrast, soft sensors fail primarily through information-related mechanisms, such as missing or corrupted input data, miscalibrated upstream sensors, preprocessing or integration errors, model drift, outdated training data, or malfunctioning alarms. These failure mechanisms can remain hidden while producing plausible-looking outputs, thereby introducing safety risks through incorrect operational decisions. In the water sector, this is particularly relevant because ML tools are increasingly proposed for monitoring and water quality management, yet their reliability under real operational constraints (e.g., Supervisory Control and Data Acquisition (SCADA) variability) is still a key barrier to broader adoption (Reneeth *et al.* 2025).

In this context, reliability assessment must explicitly account for the quality and visibility of information flows and the ability to detect failures early. This paper addresses that gap by adapting classical FMECA to the digital domain and embedding it within an asset-management perspective consistent with ISO 55000 principles (ISO 55000 2014). Building on established FMECA foundations (Dhillon 2007) and reliability-centred asset-management logic (Hastings 2015), the method reframes failure analysis around functional information-flow blocks (input acquisition, preprocessing, model inference and output interpretation) and integrates detectability as a severity-reduction factor to quantify the risk-mitigating value of early fault detection. Therefore:

- FMECA was reformulated for digital soft sensors, focusing on reliability of information flows rather than mechanical components.
- Detectability was explicitly integrated via a severity-reduction factor within the Risk Priority Number (RPN) formulation, enabling quantification of the protective value of observability and early warning.
- The framework was demonstrated through a safety-critical ozonation case study, highlighting dominant risks (e.g., missing inputs, drift, alarm failure) and showing how detectability-driven measures can reduce realised risk.

2. Methodology

An adapted FMECA framework was developed to assess the reliability of ML-based soft sensors used in drinking-water treatment. The analysis was structured around four functional blocks common to digital prediction pipelines:

- Input acquisition: collection of real-time data from SCADA systems and physical sensors;
- Preprocessing: filtering, validation, scaling and handling of missing or anomalous data;
- Model inference: execution of the trained ML model and generation of predictions;
- Output interpretation and alerts: comparison against operational thresholds and triggering of alarms or fallback logic.

Failure modes were identified within each block and characterised in terms of their potential effects on treatment performance, compliance and public health.

2.1. Severity, likelihood and detectability

Each failure mode was evaluated using three parameters: Severity (S), Likelihood (L) and Detectability (D), following an asset-management logic consistent with ISO 55000 (2014) and classical FMECA guidance (Dhillon 2007, Hastings 2015).

2.1.1. Severity (S)

Severity represents the potential consequence of a failure on system performance, drinking-water quality and compliance. A discrete powers-of-two scale from 1 to 64 was used to ensure consistent risk stratification across digital assets. Table 1 presents the severity scale and descriptors applied in this study, with severity levels ranging from negligible operational nuisance ($S = 1$) to critical public-health risk or regulatory non-compliance ($S = 64$).

Table 1. Severity (impact) scale and qualitative descriptors used for assessing the consequences of soft-sensor failure modes.

Score of Severity	Label	Descriptor/Examples
1	Negligible	No noticeable consequence; no service effect; no cost.
2	Minor	Small nuisance; negligible time or cost loss.
4	Moderate	Measurable service degradation or minor process deviation; no regulatory impact.
8	Significant	Non-permanent injury or minor environmental issue; near-miss quality deviation.
16	Major	Service loss requiring intervention; single target excursion; reportable incident.
32	Severe	Quality non-compliance with potential health consequence; repeated excursions; customer impact.
64	Critical	Public-health risk; widespread non-compliance; structural or systemic failure.

Table 2. Likelihood scale based on qualitative MTBF bands for digital soft sensors.

Score of likelihood	Indicative MTBF (years)	Typical frequency (per year)
1	≥ 100	≤ 0.01 (once per century)
2	30–100	~ 0.01 –0.03
4	10–30	~ 0.03 –0.1
8	3–10	~ 0.1 –0.33
16	1–3	~ 0.33 –1
32	0.08–1	1–12 (annual to monthly)
64	≤ 0.08	≥ 12 (monthly or more frequent)

2.1.2. Likelihood (*L*)

Because empirical failure statistics are rarely available for emerging digital assets, likelihood was estimated using qualitative Mean Time Between Failures (MTBF) bands. This preserves the interpretive intent of classical FMECA occurrence ratings while accommodating data limitations typical of early deployment stages. Table 2 summarises the likelihood scale, mapping MTBF ranges to discrete likelihood scores.

2.1.3. Detectability (*D*) as a severity-reduction factor

In classical FMECA, detectability is treated as an independent multiplicative parameter. In this study, however, detectability was not modelled as a separate factor. Instead, it was implemented as a severity-reduction fraction $r_D \in [0, 1]$, representing the proportion of impact mitigated through early detection and response. This formulation reflects the reality of digital systems: rapid detection does not reduce the probability of failure, but it can substantially limit its realised consequences. Table 3 maps qualitative detectability levels to numerical reduction fractions r_D .

Table 3. Mapping of qualitative detectability levels to severity-reduction fraction r_D , used to quantify the mitigating effect of early fault detection.

Level	Descriptor	Typical means	Range of (r_D)	Midpoint (r_D)
D1	Immediate/Automatic	Real-time alarms with auto-fallback	0.75–1.00	0.875
D2	Very high	Redundant checks, shadow models	0.50–0.75	0.625
D4	High	Routine monitoring with fallbacks	0.25–0.50	0.375
D8	Moderate	Dashboard-based anomaly review	0.10–0.25	0.175
D16	Low	Manual inspection required	0.00–0.10	0.050
D32	Very low	Off-line analysis needed	0.00–0.00	0.000
D64	Undetectable	No indicators until failure	0.00–0.00	0.000

The resulting effective severity is defined as:

$$S_{\text{eff}} = S \times (1 - r_D) \quad (1)$$

2.1.4. Risk Priority Number (RPN) formulation

The final Risk Priority Number (RPN) is calculated as

$$\text{RPN}_{\text{final}} = L \times S \times (1 - r_D) \quad (2)$$

This formulation preserves the interpretability of traditional FMECA screening thresholds, while explicitly rewarding improvements in information visibility, monitoring and alarm responsiveness. A classical

two-factor screening matrix (Severity $\times$ Likelihood) provides a rapid overview of which combinations of consequence and frequency fall into unacceptable risk zones before detectability is considered.

2.2. Case study

The framework was applied to the CT (Concentration $\times$ Time) ozone-exposure soft sensor used for disinfection control at Leiduin drinking-water treatment plant of Waternet water company, the only public water utility serving Amsterdam and its surroundings (Kyrtsakas *et al.* 2025). The soft sensor estimates ozone exposure based on SCADA inputs, such as ozone dose, water flow through the reactors, residence time and water temperature. CT is an operational control parameter: underestimation risks insufficient microbial inactivation, while overestimation can lead to excessive chemical dosing and by-product formation.

Operational knowledge, typical ozone-control dependencies and platform diagnostics were used to identify failure modes and assign qualitative scores. The analysis was conducted pre-deployment, providing a proactive assessment of digital reliability risks.

3. Results and discussion

3.1. Identified failure modes for CT

Applying the adapted FMECA framework to the CT (Concentration $\times$ Time) ozone-exposure soft sensor resulted in the identification of failure modes associated with the four functional blocks as defined in Section 2. For the CT application, the most relevant failure modes included:

- Missing or delayed SCADA input data, affecting ozone dose, water flow, residence time, or temperature signals;
- Upstream sensor miscalibration, leading to biased or systematically incorrect inputs, such as an overestimated residual ozone concentration due to drift in the ozone analyser or an underreported flow rate from a miscalibrated flow meter, resulting in incorrect CT estimation;
- Unfiltered spikes or noise in input signals propagating through the prediction chain;
- Model drift, caused by changing operating conditions or water-quality regimes over time;
- Failure of the low-CT compliance alarm, caused by incorrect configuration of the regulatory CT setpoint or loss of signal transmission to the SCADA dashboard, resulting in the absence of warning when the predicted CT drops below microbial inactivation requirements;
- System-level failures, such as Application Programming Interface (API), server, or container crashes interrupting model execution.

These failure modes form the basis for the subsequent quantitative risk prioritisation using severity, likelihood and detectability scores.

3.2. Baseline risk prioritisation (Severity $\times$ Likelihood)

Initial risk screening was performed using the classical two-factor FMECA approach (Severity $\times$ Likelihood). This screening highlights failure modes that combine high potential impact with non-negligible frequency and therefore require further attention. The results show that missing SCADA input data and model drift emerged as the most critical risks at this stage, combining high severity with moderate likelihood. Alarm failure and system-level outages also ranked highly due to their potential impact on disinfection reliability. These results confirm that the most considerable risks originated from information degradation rather than from algorithmic malfunction alone.

3.3. Effect of detectability

When detectability was incorporated using the severity-reduction formulation (Equation 2), a substantial redistribution of risk is observed. Failure modes equipped with strong monitoring, redundancy, or

Table 4. Comparison of baseline (RPN_{screen}) and detectability-adjusted (RPN_{final}) risk scores for key CT soft-sensor failure modes.

Failure Mode	S	L	Detectability (D , (r_D))	RPN_{screen}	RPN_{final}
Missing SCADA data	64	8	D1 ($r_D = 0.875$)	512	64
Sensor miscalibration	32	4	D2 ($r_D = 0.625$)	128	48
Unfiltered spikes	8	16	D4 ($r_D = 0.375$)	128	80
Model drift	32	8	D1 ($r_D = 0.875$)	256	32
Alarm not triggered	64	2	D2 ($r_D = 0.625$)	128	48
API/Server crash	64	2	D2 ($r_D = 0.625$)	128	48

automated fallback mechanisms exhibited considerably reduced final numbers of RPN_{final} compared to their baseline screening values.

In particular, real-time data-availability checks and automated drift-detection mechanisms reduced RPN values by up to 80%. In contrast, poorly detectable failure modes remained in high-risk zones even when their likelihood was relatively low. This demonstrates that detectability acts as the dominant driver of realised digital risk, outweighing marginal changes in failure frequency or severity. The baseline and detectability-adjusted risk scores for the key CT soft-sensor failure modes are summarised in Table 4.

3.4. Visualisation of detectability-driven risk reduction

The influence of detectability on risk prioritisation is further illustrated using multi-dimensional FMECA visualisation (Figure 1). The three-dimensional matrix shows how risk escalated when detectability decreased, even for moderate severity-likelihood combinations. This confirms that digital reliability cannot be adequately assessed using two-factor screening alone.

Complementary two-dimensional heatmaps (Figure 2) demonstrate that transitions between detectability levels resulted in abrupt shifts between acceptable and unacceptable risk zones. These visualisations

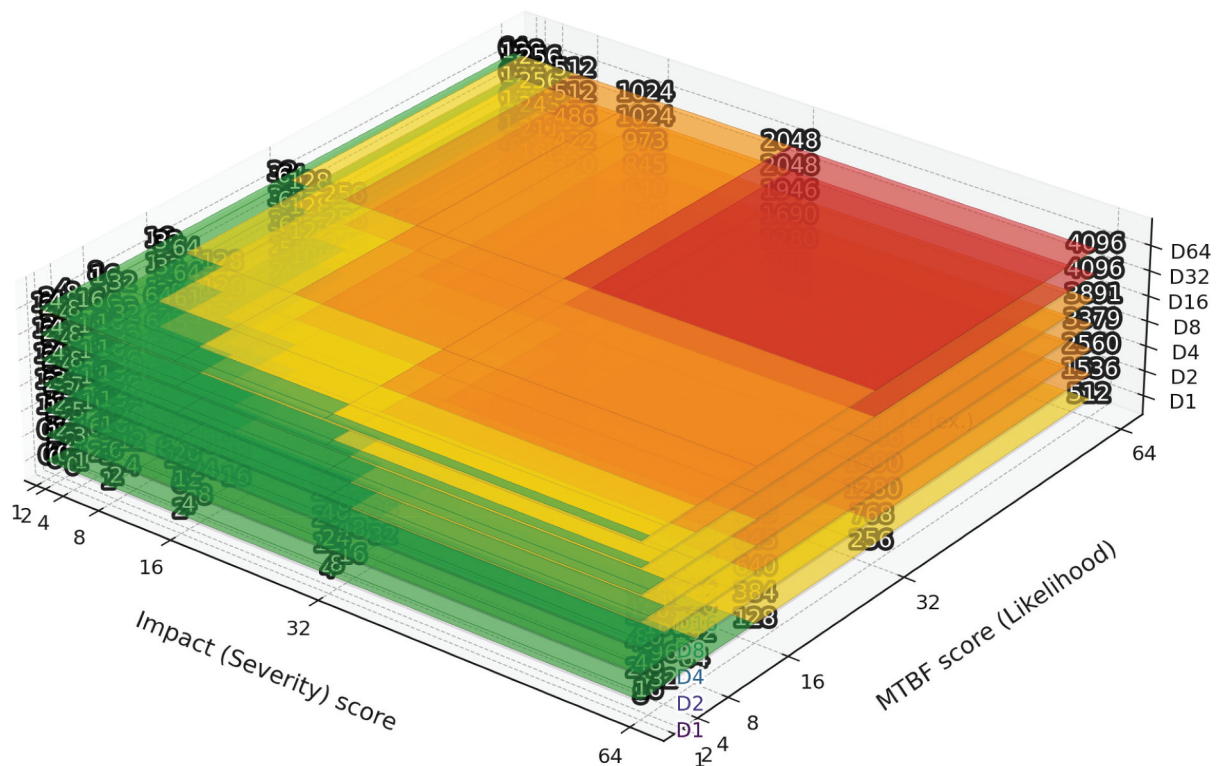


Figure 1. Three-dimensional FMECA risk matrix showing the combined effect of severity (impact), likelihood (MTBF score) and detectability level (D1–D64) on the final risk Priority Number (RPN_{final}). Colour intensity represents increasing risk, illustrating how decreasing detectability shifts severity-likelihood combinations into higher risk zones.

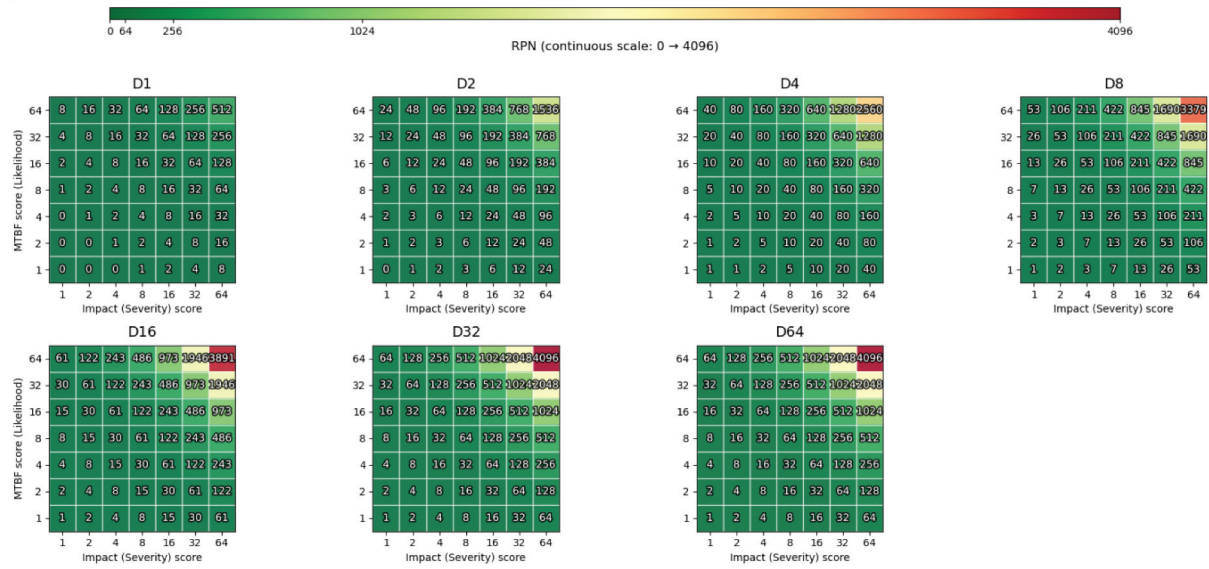


Figure 2. Severity-likelihood heatmaps of RPN_{final} across detectability levels (D1–D64), showing the redistribution of risk as detectability decreases for identical severity-likelihood combinations.

reinforce the conclusion that information visibility and early fault detection are the most effective levers for digital-risk reduction.

3.5. Sensitivity analysis

A tornado-style (Figure 3) sensitivity analysis shows that reductions in detectability consistently led to larger increases in RPN_{final} than equivalent one-step changes in severity or likelihood across most failure modes. Sensitivity was evaluated by applying one-step increases in severity and likelihood and a 0.25 reduction in detectability from a baseline $r_D = 0.50$. In particular, failure modes related to

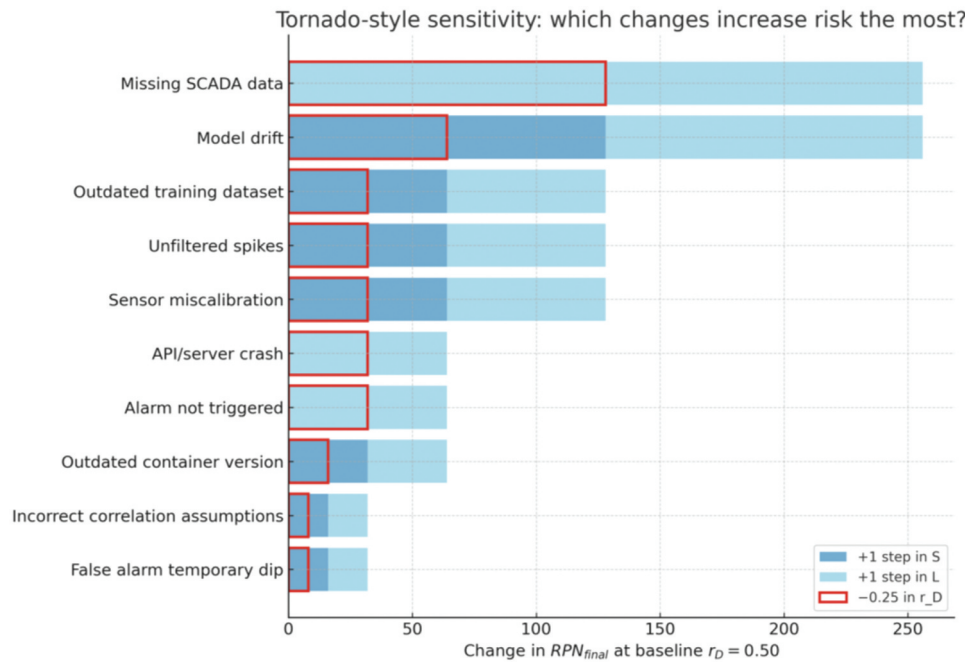


Figure 3. Tornado-style sensitivity analysis of detectability-adjusted risk (RPN_{final}), showing the relative influence of changes in severity (S), likelihood (L), and a reduction in the detectability factor (r_D), on the final risk score.

missing SCADA data and model drift exhibited strong sensitivity to detectability, indicating that loss of observability rapidly escalated risk even when baseline likelihood remained moderate. This confirms that, for digital assets, reliability is primarily governed by information visibility and early fault detection.

4. Conclusions

This study demonstrated that classical FMECA can be meaningfully adapted to assess the reliability of machine-learning-based soft sensors by explicitly accounting for information-related failure mechanisms. While baseline severity-likelihood screening identified critical failure modes, such as missing SCADA data and model drift, incorporating detectability revealed that realised risk depends more strongly on monitoring and observability than on failure frequency alone.

From an asset-management perspective, these findings indicate that ML-based soft sensors should be managed as critical digital assets, with priority given to data validation, monitoring coverage and alarm performance rather than solely to improvements in model accuracy. The proposed framework supports ISO 55000-aligned decision-making by providing a structured basis for prioritising reliability measures that directly reduce operational risk.

Integrating detectability into FMECA provides a practical and scalable approach for digital reliability assessment in drinking-water systems. Application to a CT ozone-exposure soft sensor demonstrated that improvements in observability can reduce RPN values by up to 80%. The framework offers a reproducible foundation for managing soft sensors within decision-support platforms and can be extended to other digital assets and treatment processes in the future work.

The analysis is based on expert-informed qualitative scoring and pre-deployment assumptions; future work should incorporate empirical failure data from operational use to validate and refine the proposed framework. Overall, the key finding of this study is that incorporating detectability into digital reliability assessment fundamentally reshapes the understanding and management of risk. In this context, early fault detection emerges not merely as a supportive function, but as a primary mechanism for reducing operational risk in soft sensor applications.

Acknowledgments

This work was funded by the European Union under the Horizon Europe programme under grant agreement no. 101082035. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Author contributions

CRedit: **Akrivi Alexandraki**: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing; **Grigorios Kyritsakas**: Supervision; **Martine van den Boomen**: Validation; **Luuk Rietveld**: Supervision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

During the preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-5.6 Thinking; accessed July 2026) solely for language refinement, including checks of grammar, syntax, spelling, clarity and consistency. The tool was not used to generate research data, conduct analyses, derive results, create figures or select references. All AI-assisted revisions were critically reviewed and verified by the authors, who take full responsibility for the accuracy, originality, integrity and final content of the manuscript.

Funding

This work was funded by the European Union under the Horizon Europe programme under Grant Agreement No. 101082035.

References

- Brunner, V., *et al.*, 2021. Challenges in the development of soft sensors for bioprocesses: a critical review. *Frontiers in bioengineering and biotechnology*, 9, 722202. doi:[10.3389/fbioe.2021.722202](https://doi.org/10.3389/fbioe.2021.722202).
- Dhillon, B.S., 2007. *Applied reliability and quality*. London: Springer.
- FIWARE Foundation, *FIWARE Catalogue for Smart Applications*. FIWARE Foundation, n.d.
- Hastings, N.A.J., 2015. *Physical asset management*. Cham: Springer.
- ISO 55000:2014, *Asset management — overview, principles and terminology*.
- Kyritsakas, G., *et al.*, 2025. *Machine learning prediction of ozone exposure for improving the ozonation process in a drinking water treatment plant*. Sheffield, UK: University of Sheffield Data Repository. doi:[10.15131/shef.data.29921090.v1](https://doi.org/10.15131/shef.data.29921090.v1).
- Reneeth, F., *et al.*, 2025. Intelligent effluent management: AI-based soft sensors for organic and nutrient quality monitoring. *Processes*, 13, 1664. doi:[10.3390/pr13061664](https://doi.org/10.3390/pr13061664).
- Shen, Y., *et al.*, 2025. Drift correction method of wastewater treatment model based on transfer learning across time scales. *Huan Jing Ke Xue*, 46 (1), 318–326.
- ToDrinQ Project. 2024, *Toolkit for Adaptable, Resilient Installations Securing High Quality Drinking Water*. EU Horizon Europe.