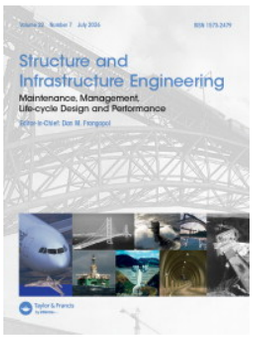

Structure and Infrastructure Engineering
Maintenance, Management, Life-Cycle Design and Performance

ISSN: 1573-2479 (Print) 1744-8980 (Online) Journal homepage: www.tandfonline.com/journals/nsie20

# A screening method to identify promising performance improvements of storm surge barriers: a case study in the Netherlands

**Leslie F. Mooyaart, Johan A. van den Bogaard , Alexander M.R. Bakker , Martine van den Boomen & Sebastiaan N. Jonkman**

**To cite this article:** Leslie F. Mooyaart, Johan A. van den Bogaard , Alexander M.R. Bakker , Martine van den Boomen & Sebastiaan N. Jonkman (07 Jul 2026): A screening method to identify promising performance improvements of storm surge barriers: a case study in the Netherlands, Structure and Infrastructure Engineering, DOI: 10.1080/15732479.2026.2696413

**To link to this article:** https://doi.org/10.1080/15732479.2026.2696413

STRUCTURE AND INFRASTRUCTURE ENGINEERING
https://doi.org/10.1080/15732479.2026.2696413



# A screening method to identify promising performance improvements of storm surge barriers: a case study in the Netherlands

Leslie F. Mooyaart[a], Johan A. van den Bogaard[b], Alexander M.R. Bakker[a,b], Martine van den Boomen[a,c] and Sebastiaan N. Jonkman[a,d]

[a]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands; [b]Rijkswaterstaat, Utrecht, the Netherlands; [c]Rotterdam University of Applied Sciences, Rotterdam, the Netherlands; [d]Texas A&M Galveston, Galveston, Texas, USA

**ABSTRACT**

Storm surge barriers are critical structures in coastal flood defences that operate only during expected storm surge events. Rising sea levels increase flood risk and intensify the need to improve barrier reliability. At present, reliability analysis supports risk reduction by identifying, quantifying, and prioritising failure scenarios. However, this scenario-driven approach can overlook effective performance improvements, given the wide range of potential measures that differ in scale and type. This study introduces a checklist of 85 candidate performance improvements, organised into eight categories, and proposes a novel screening method with three stages: (1) importance analysis to eliminate negligible improvements, (2) expert judgement to discard unpromising options, and (3) expert ranking of the remaining candidates. The method was applied to the Maeslant Barrier (Rotterdam, the Netherlands), yielding a ranked top-five set of improvements. The case study demonstrates that current practices are particularly poor at identifying low-cost improvements above the component level with little organisational burden. Although the proposed checklist and screening method are more comprehensive and structured than current approaches, additional case studies are needed to validate the method and assess its broader applicability to storm-surge barriers and other safety-critical infrastructure systems.

**ARTICLE HISTORY**
Received 15 December 2025
Accepted 19 April 2026

**KEYWORDS**
Safety-critical; reliability; availability; screening; storm surge barrier; flood defense; performance killer; performance improvement

## 1. Introduction

Storm surge barriers play a crucial role in protecting coastal cities such as Rotterdam (The Netherlands), London (UK), and New Orleans (USA) from coastal flooding. These large movable hydraulic structures, strategically placed in estuaries, close when extreme storm surges are predicted. In normal conditions, they remain open to facilitate ship passage and tidal flow (Mooyaart & Jonkman, 2017). These specific characteristics make maintaining these structures challenging (Jonkman et al., 2018; Kamps et al., 2024; Kharoubi et al., 2024).

Increasing coastal flood risk, driven by sea level rise and urbanisation, requires enhancing coastal flood defences (Hallegatte et al., 2013). For storm surge barriers, preserving the current storm surge barrier performance with sea level rise is already difficult as maintenance windows shorten (Trace-Kleeberg et al., 2023). Mooyaart et al. (2025) integrally evaluate operational, structural failures and those resulting from the barrier having a too low crest level. In Rotterdam, which is protected by the Maeslant barrier – and also for the Hollandsche IJssel barrier (Vader et al., 2024) – the probability of a failure to close emerged as the dominant failure mechanism. Mooyaart et al. (2023) demonstrate that an expensive performance improvement (an additional barrier in front of the Maeslant barrier, see Rijcken et al. (2023)) can already be economically beneficial with a rise in sea level of 0.2 to 0.3 metres. However, it is expected that many performance improvements are even more economically beneficial than an additional barrier.

In the Netherlands, the probability of a failure to close is estimated with a *reliability and availability analysis* (RA analysis), which uses fault and event trees that contain a thousand to ten thousand failure scenarios. In this paper, a failure scenario is defined as a group of failure events resulting in failure of the analysed object. In other industries, RA analyses are referred to as probabilistic safety assessments, quantitative risk analyses or other combinations of these words. RA analyses are applied to various safety-critical infrastructures and systems (nuclear plants, aeroplanes, space missions, chemical facilities) to analyse events with low probabilities and large consequences, where data to quantify risk are sparse (Bier & Cox, 2007). These types of analyses were also tested for urban drainage systems (ten Veldhuis et al., 2011) and timber bridges (Lokuge et al., 2019).

RA analyses are criticised by those who analyse risks in safety-critical systems. The quality of them is often considered too low (Pasman et al., 2017; Pitblado, 1994; Rae et al., 2014; Thekdi & Aven, 2026). They do not include the

CONTACT Leslie F. Mooyaart L.F.Mooyaart@tudelft.nl Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15732479.2026.2696413.

quality of the safety culture and the true complexity of software reliability (Apostolakis, 2004; Leveson, 2004). Their validation is difficult given the low frequency that failure events occur (Aven & Heide, 2009; Goerlandt et al., 2017; Rae et al., 2014). In many articles (see for example: (Aven & Renn, 2009; Aven & Zio, 2013; Aven, 2017)), Aven warns that RA analyses consider risks as probability x consequence, ignoring the often weak state of knowledge supporting these claims. Despite criticising RA analyses, all authors mentioned in this paragraph – except for Leveson – emphasise its usefulness for risk management of safety systems such as storm surge barriers.

Rijkswaterstaat – the executive agency of the Ministry of Infrastructure and Water Management responsible for the maintenance and operation of Dutch storm surge barriers – has been challenged for several decades with exploring relevant performance improvements. After a thorough RA-analysis of the Maeslant storm surge barrier (NRG, 2003), several performance improvements were implemented. In a review (Horvat, 2006) five additional performance improvements were suggested of which some were executed. In the following decades, likewise, many performance improvements were recommended and occasionally implemented. However, the process was unsystematic, and it remained difficult to establish whether all relevant performance improvements were considered.

The current practice at Rijkswaterstaat and in reliability engineering in general (Aven, 2017; Henley & Kumamoto, 1981; Paté-Cornell, 2002) is to mitigate risk by searching for performance killers in the RA-analysis. *Performance killers* refer to either a single event or group of events that greatly contribute to overall unreliability, unavailability, and/or probability of failure on demand, the influence of which could be lowered by a single performance improvement. The identified performance killers are subsequently used to find performance improvements by *putting yourself into a creative mode*, as Haimes et al. (2002) nicely phrase.

To illustrate that finding performance killers can be difficult in fault and event trees, the fault tree of Figure 1 is used. A fault tree illustrates how both technical and human failure causes lead to an undesired top event, typically by applying AND- and OR-gates. AND-gates indicate that either all underlying failure events have to occur for the upper event to occur. OR-gates are used when only one underlying failure event has to occur for the upper event to occur. The fault tree represents an object that consists of two subsystems, of which one has two components. Customary formulas (Schüller et al. (1997), pages 9.30 to 9.40) are used to determine the unavailability of technical components.

Already, for this half-page fault tree (note that the RA model of the Maeslant barrier is over 1,200 pages), many potential performance killers can be identified, depending on how failure events are grouped and ranked. Five of these methods are presented in Figure 1. They are also used to discuss current approaches and their limitations in identifying performance killers. First, basic *failure events* and *failure scenarios* can be identified that kill performance. Then, *subsystems* containing multiple failure events already require grouping of failure events. Other useful grouping options include *variables*

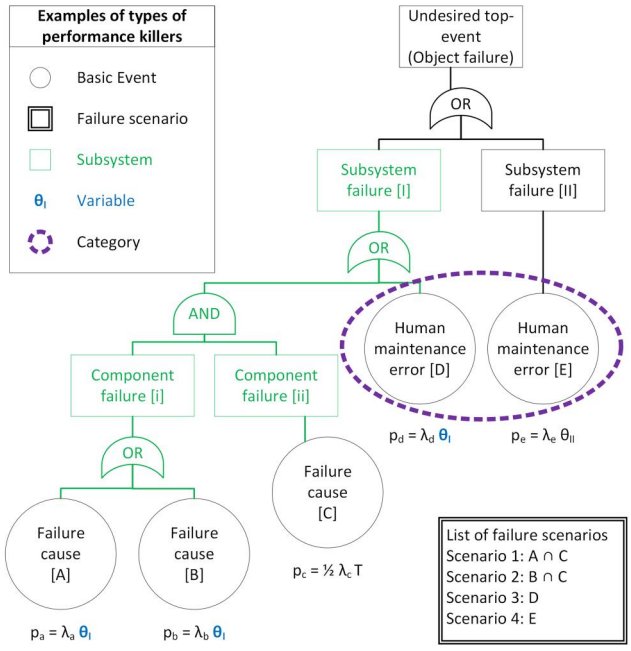


**Figure 1.** Example of a fault tree of a relatively simple system: an object with two subsystems of which one has two components. The fault tree shows how technical and human failure causes lead to object failure using AND-gates and OR-gates. With AND-gates both underlying failure events need to occur for the upper event to occur, while with OR-gates only one needs to occur. The fault tree aims to clarify that even for this relatively simple system, there are many performance killers of which five examples are highlighted: any basic event, any failure scenario, subsystem A, repair time $\theta_I$ and human maintenance errors.

affecting multiple failure events, such as repair times for multiple components, and failure events belonging to a specific *category*, such as human maintenance errors.

Rijkswaterstaat guidelines (Rijkswaterstaat, 2011, 2018a) propose to rank failure scenarios – in fault tree analysis referred to as minimal cut-sets – according to their contribution to the object's unreliability and/or unavailability. Failure scenarios have occasionally been grouped into subsystems (Horvat, 2006; Rijneveld, 2008). However, this approach can easily overlook basic failure events which affect multiple failure scenarios.

Alternatively, the importance measures of Fussell-Vesely (Vesely et al., 1983) and Birnbaum (1969) are well-known analytical tools for identifying basic event performance killers. The Fussell-Vesely importance measure indicates the extent to which a basic event affects object performance. The Birnbaum importance measure compares the object's performance in the case of success and failure of basic events. Rijkswaterstaat has occasionally applied these importance measures to prioritise the importance of basic events. Many authors (Borgonovo & Apostolakis, 2001; Dutuit & Rauzy, 2015; Liu et al., 2021; Mutar & Hassan, 2025; Rusnak et al., 2024; Scherb et al., 2017; Xu et al., 2020) extend upon these importance measures to find new ways to rank failure events. With this approach, however, grouping of failure events can be overlooked.

Therefore, Cheok et al. (1998) describe how to use importance measures for groups of basic events. They suggest defining groups for (sub)systems and components and evaluating them based on the size and constituency of these groups. The prevailing nuclear industry guideline (Nuclear

Energy Institute, 2018) still includes their recommendation. Moreover, several authors (Borgonovo & Apostolakis, 2001; Dutuit & Rauzy, 2015) acknowledge the possibility of importance measures to group events, but do not discuss how to group failure events.

In fact, the grouping method determines which performance killers are identified. For instance, prioritising components results in a list of the most critical components, prioritising repair times results in a list of the most critical repair times, and so on. For simple fault trees, such as those sketched in Figure 1, the multidimensionality of performance killers makes it difficult to systematically identify performance improvements. For large fault trees, it is unlikely that all relevant performance improvements are identified using the discussed approaches.

From analysing literature, it can be concluded that the success of finding relevant performance killers and performance improvements is currently determined by the creativity and quality of reliability experts. As there are many possible performance improvements that vary in scale and type, this approach can easily overlook important performance improvements. Thus, it is not known how to identify relevant performance killers and improvements in a systematic manner.

To overcome this issue, this paper investigates an alternative approach to find promising performance improvements that uses 1) a checklist of principal performance improvements (PPI's) and 2) a method to screen that checklist effectively. The remainder of this paper is structured as follows: Section 2 presents key definitions used throughout the study. Section 3 describes the research approach for developing the checklist. Section 4 introduces the three-stage screening method for applying the checklist. Section 5 demonstrates the application of the checklist and screening method to the Maeslant barrier case study. Section 6 discusses the method and section 7 concludes the paper.

## 2. Reliability and availability analysis of storm surge barriers

### 2.1. Definitions

This section provides some important definitions to clarify the checklist (Supplementary Appendix A) and the screening method presented in this paper. Four hierarchical levels of decomposition are defined: *object, object system, subsystem* and *component*. In this paper, the object is always the storm surge barrier. For a decomposition with more hierarchical levels than four, subsystems are ranked (i.e., subsystem level 1, subsystem level 2, etc.). Thus, components are the lowest level of detail in the object decomposition (Schüller et al., 1997). Each element of the object decomposition provides a particular function. A failure is defined as a loss of this element's functionality (Dummer et al., 1997). A *failure criterion* is used to define how much functionality can be lost before an element is considered failed. For instance, at a storm surge barrier with multiple gates, a closure of all gates except one can still be considered a successful closure if this event does not result in a coastal flood. In this example, the term partial failure can also be used to describe that part of the element's functionality is lost.

A reliability and availability analysis (RA analysis) establishes the reliability, availability and/or probability of failure on demand of an object. For storm surge barriers, the most important result is often the probability of a failure to close, which best aligns with the term probability of failure on demand. To avoid repetition, only this term is used to describe the main result of an RA analysis – and not reliability and availability. On the failure event level, four *unavailability types* are defined: 1) dormant failure[1], 2) unavailability due to repair, 3) failure during mission and 4) failure on demand. Here, unavailability is the probability that an object is in a failed state at a stated time under a given set of conditions (Li et al., 2024; Van Der Weide & Pandey, 2015). Based on guidelines for the nuclear industry (Iaea, 1992) and storm surge barriers (Rijkswaterstaat, 2018a), five *failure categories* are defined: 1) hardware failure, 2) external events, 3) common cause failure, 4) human error and 5) software error. Thus, hardware failures are all inherent failures of physical objects except for failures due to external events and common cause failure.

### 2.2. Set-up RA model

An RA model is part of an RA analysis that 1) illustrates how combinations of events result in undesired outcomes and 2) is used to quantify the probability and/or frequency of those outcomes. An RA model consists of three elements:

- A *logic model structure*, for instance, a fault or an event tree.
- A *failure event model structure*, for instance, the equations shown below the basic event in Figure 1.
- *Model variables*, for instance, the failure rate $\lambda$, the repair time $\theta$ and the test interval $T$ in Figure 1.

Failure event model structures can contain a model parameter, that is, a calibrated and/or validated value specifically belonging to that model. For instance, in the TOPAAS-model – a model specifically developed to determine the probability of software error at storm surge barriers (Van Manen et al., 2015) – the maximum probability of software error of $10^{-5}$ is such a model parameter. In contrast to model variables such as the development and test quality of software, the model validity is lost when the maximum probability is adjusted. Thus, adapting a model parameter is considered a change of the failure event model structure.

The logic model establishes failure scenarios $S_i$, which contain one or multiple failure events $(A, B, \ldots)$. When failure scenario probabilities $P_i$ are low and independent from each other, the probability of failure on demand $P_f$ can be approximated with:

$$P_f = 1 - \prod_{i}^{n}(1 - P_i) \approx \sum_{in}^{n} P_i \tag{1}$$

### 2.3. Fussell-Vesely importance measure

To effectively screen principal performance improvements (PPI's), the Fussell Vesely Importance Measure $I_{FV}$ is used to establish the maximum effect a PPI has on the

probability of failure on demand. Here, the RA model in Figure 1 is used to demonstrate how the Fussell-Vesely Importance Measure $I_{FV}$ is determined for a basic event, groups of events and variables. This model has four failure scenarios, which have the following probabilities:

$$P_1 = P(A \cap B) = \lambda_a \cdot \theta_I \cdot \frac{1}{2} \cdot \lambda_c \cdot T$$
$$P_2 = P(A \cap C) = \lambda_b \cdot \theta_I \cdot \frac{1}{2} \cdot \lambda_c \cdot T$$
$$P_3 = P(D) = \lambda_d \cdot \theta_I$$
$$P_4 = P(E) = \lambda_d \cdot \theta_{II}$$

In which $\lambda$ are failure rates of components or human error [per hour], $\theta$ the repair time [hour], and $T$ the test interval [hour], relevant for dormant failures. Indices correspond to components and subsystems where failure rates, repair times, and test intervals that failure events belong to (see Figure 1). With equation (1), the probability of failure on demand for this object is found:

$$P_f = \sum_1^4 P_i = \frac{1}{2} \cdot (\lambda_a + \lambda_b) \cdot \lambda_c \cdot \theta_I \cdot T + \lambda_d \cdot \theta_I + \lambda_e \cdot \theta_{II}.$$

For basic events, the Fussell-Vesely Importance Measure $I_{FV}$ is the sum of all failure scenarios that contain that event relative to the probability of failure on demand. For failure event $A$ in Figure 1, the Fussell-Vesely Importance Measure $I_{FV}$ is:

$$I_{FV}(A) = \frac{\sum P_i(A)}{P_f} = \frac{P_1}{P_f} = \frac{\frac{1}{2} \cdot \lambda_a \cdot \lambda_c \cdot \theta_I \cdot T}{P_f} \quad (2)$$

For group of failure events, the Fussell-Vesely Importance Measure $I_{FV}$ is the sum of all failure scenarios that contain at least one of the failure events relative to the probability of failure on demand. The Fussell-Vesely Importance Measure $I_{FV}$ of human maintenance errors $(D \cup E)$ is, thus:

$$I_{FV}(D \cup E) = \frac{\sum P_i(D \cup E)}{P_f} = \frac{P_3 + P_4}{P_f} = \frac{\lambda_d \cdot \theta_I + \lambda_e \cdot \theta_{II}}{P_f} \quad (3)$$

For variables affecting multiple events, the Fussell-Vesely Importance Measure $I_{FV}$ is calculated in an opposite manner. For the variable (in the example of Figure 1, repair time $\theta_I$), the probability of failure on demand without the influence of this variable $(\theta_I = 0)$ is determined. This value $P_f(\theta_I = 0)$ is then used to attain its Fussell Vesely Importance Measure $I_{FV}(\theta_I)$ in the following manner:

$$I_{FV}(\theta_I) = 1 - \frac{P_f(\theta_I = 0)}{P_f} = 1 - \frac{\lambda_e \cdot \theta_{II}}{P_f} \quad (4)$$

In the hypothetical case that each failure scenario has the same probability $(P_1 = P_2 = P_3 = P_4 = 25\%\ P_f)$, multiple performance killers can be identified. Failure cause $C$ and the failure category human maintenance errors are included in half of the failure scenarios, resulting in a Fussell-Vesely value $I_{FV}$ of 50%. Subsystem $I$ and its repair time $\theta_I$ affect three failure scenarios and, thus, their maximum effect $I_{FV}$ is 75%. This hypothetical case again illustrates that even a simple object can have multiple performance killers, depending on how it is analysed.

## 3. Development of a checklist

The checklist of principle performance improvements (PPI's) was developed in three steps:

1. Literature study
2. Dialogues
3. Systematic structuring

The checklist was revised after the case study (see section 5.4 for motivations). The final version is presented in supplementary appendix A.

### 3.1. Literature study

The literature only reveals a few structured lists to enhance the performance of safety-critical infrastructure. Lists were discovered for defending against common cause failures (Paula et al., 1990; Parry, 1991) and best practices for the reliability of spillway gates of dams (Lewin et al., 2003), which resulted in an initial list of 32 PPI's. Due to the scarcity of performance improvement lists, the literature study was extended.

Model variables were studied of human reliability models (Barnes et al., 2000; Heslinga, 2004; NUREG, 2004; Swain & Guttmann, 1983; US Nuclear Regulatory Commission, 2006), software reliability models (Van Manen et al., 2015; Rijkswaterstaat, 2018b), external event screening methods (Knochenhauer & Louko, 2003), a repair model (Kiel et al., 2023) and a guideline on CCF modelling (International Electrotechnical Commission, 2021). For these variables, corresponding PPI's were proposed. For example, software reliability models (United States Nuclear Regulatory Commission, 2011; Van Manen et al., 2015) consider the development process as a factor influencing the probability of software errors. Subsequently, 'redeveloping software' was added as a PPI to our checklist. Excluding human reliability performance improvements, this step resulted in 22 additional PPI's.

The literature study into human reliability models highlighted the need to cluster PPI's for practical use. The OPSCHEP human reliability model, used for storm surge barriers, has 36 variables (Heslinga, 2004). While these variables stem from nuclear industry studies (e.g. Barnes et al. (2000)), different models often employ slightly different variables. Moreover, these studies present lists of best practices and important considerations, some of which may not be included in the model itself. Attempting to comprehensively cover all these variables and aspects would result in an unwieldy list of over a hundred PPI's for human reliability alone, many of which would be closely related. Based on this insight, PPI's of human reliability and other categories were clustered, which resulted in a list of approximately fifty PPI's.

### 3.2. Dialogues

The initial checklist, based on the academic literature review only, was found incomplete, lacking crucial PPI's implemented at Dutch storm surge barriers. This highlighted a gap between academic literature and real-world practice in storm surge barrier management.

To address this gap, dialogues were conducted with co-authors and reliability experts of Rijkswaterstaat. An important example was that high software probability errors could be resolved by human recovery. This revealed that performance killers in a certain category could be addressed by solutions from another category. Based on this insight, six PPI's were added (e.g. hardware improvements for software, software improvements for human error, etc.). In this stage, operational performance improvements such as preventive closure of the barrier in case of malfunctioning were identified as well, resulting in five more PPI's.

Finally, the option to adapt the RA model was mentioned. Based on this suggestion, options were studied to reduce conservatism in RA analyses, which Aven (2016) convincingly argues is desirable. Three PPI's were found of this type: using less conservative model variables (Hegseth et al., 2021), less conservative model structures (Gomes & Beck, 2021) and less strict failure criteria (dialogues with co-authors). These principles can be applied to the models of each failure category (human, software, hardware, external event, CCF) and, thus, resulted in fifteen PPI's.

### 3.3. Systematic structuring

The literature study and dialogues resulted in about eighty PPI's. To validate whether this long-list was near complete, three key *improvement dimensions* were defined (see Figure 2), which are the result of a combination of literature study and scientific reasoning:

- Improvement types: model, system properties, maintenance, and operation
- Failure categories: hardware, common-cause-failure, external events, human reliability, and software reliability
- Hierarchical levels: object, object system, subsystem, and component

The hierarchical levels were utilised to incorporate PPI's across the checklist. For example, PPI's were included such as testing at different hierarchical levels and implementing more reliable equipment at each level.

Although the key dimensions nicely illustrate the large number of possible performance improvements, it proved difficult to use to identify relevant performance improvements. Therefore, the concept of the checklist was returned to. This list was structured into eight categories based on these key improvement dimensions:

1. Reducing model conservatism
2. Hardware
3. Common-cause-failure

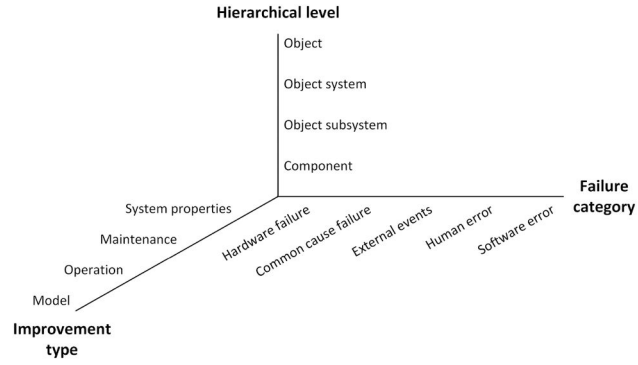


Figure 2. Three dimensions of improvement that were used to structure the checklist.

4. External events
5. Human reliability
6. Software reliability
7. Maintenance
8. Operation

## 4. Three-stage screening method

This section describes the developed method to screen the checklist for most promising performance improvements. The method is inspired by three well-known methods in reliability and availability analyses: HAZOP (International Electrotechnical Commission, 2016) for the general set-up, external events screening (Knochenhauer & Louko, 2003) for efficient selection from a long-list and FMECA (International Electrotechnical Commission, 2018) for prioritising improvements.

The method consists of three-stages (see Figure 3), which are:

- Stage 1: Analyse importance aiming to remove PPI's with no or only a marginal effect and supporting the subsequent screening stages.
- Stage 2: Select performance improvements with expert evaluation.
- Stage 3: Rank performance improvements with expert scores.

Expert sessions are included in the method as assessing the feasibility of PPI's and case-specific performance improvements requires object knowledge and expertise from different disciplines. For instance, adding redundancy can be applied to both mechanical and electrical systems, thus requiring multidisciplinary knowledge. Object knowledge was assumed to be essential to assess the costs and feasibility of a performance improvement.

### 4.1. Screening stage 1: Importance analysis

The first aim of the importance analysis is to remove PPI's which do not or only marginally affect the probability of failure on demand for the case considered. Generically applicable PPI's do not need to have an effect on a specific case. For example, the RA analysis of the considered case

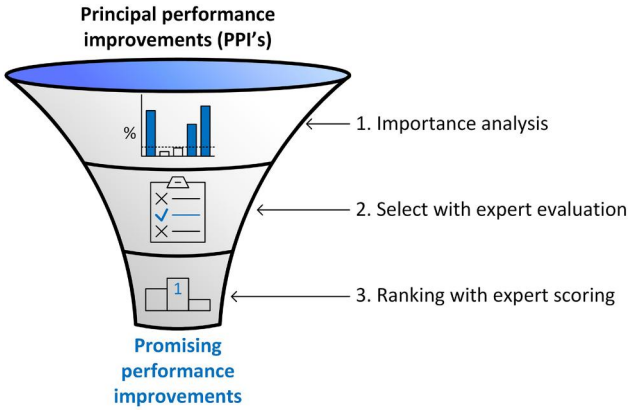


Figure 3. Schematic of the three-stage screening method.

might assume a time-constant failure rate in all hardware failure models. Hence, PPI's that reduce ageing do not have any effect.

*Screening criteria* are set to exclude PPI's that only have a small effect on the probability of failure on demand. Nuclear Energy Institute (2018) prescribes that screening criteria depend on size and constituency of the group of events. They mention criteria of 1% and 5% as typical examples for components and subsystems, respectively. Both these screening criteria are used in the case study.

A secondary aim of the importance analysis is to support decision-making in subsequent screening stages. For this purpose, performance killers are identified by determining the Fussell-Vesely Importance Measure of the following (groups of) failure events and/or variables:

1. Basic events
2. Failure categories
3. Unavailability types
4. The object decomposition
5. Combinations the object decomposition and failure categories
6. Combinations the object decomposition and unavailability types
7. Variables affecting multiple basic events

The results of the importance analysis are described in a report, which is shared with and reviewed by the team of experts.

### 4.2. Screening stage 2: Select performance improvements

The main objective of this step is to select performance improvements. In an expert session, the session leader, together with the experts, uses four *screening questions* to assess PPI's:

1. Is the PPI applicable to the case considered?
2. Is the PPI expected to have a substantial effect on reliability, availability and/or probability of failure on demand?
3. Does the PPI impede other functions of the object in an unacceptable manner?
4. Are there any other reasons why the PPI is unpromising?

The PPI is only selected for the next screening stage when the answers to the first two questions are positive, and the last two are negative. The last question can be used to rule out performance improvements that are already known to be worse than others. For example, replacing the entire storm surge barrier is more costly and less effective than adding a redundant one. Important motivations to rule out PPI's – especially in the case of the final question – are noted.

The second task of this session is to make PPI's case-specific, which is achieved in two ways: 1) per PPI, experts are asked to provide case-specific examples and 2) per category, experts check whether important performance improvements were missed.

The prepared information on performance killers, that is, the Fussell-Vesely Importance Measures for all (groups of) events and variables, is used in two manners. First, for those PPI's that just passed the screening criterion, it can be judged whether the PPI is likely to have a substantial effect on performance. Second, the most critical object systems, object subsystems and components are defined to find case-specific examples.

The results of the session are reported and shared with the experts for verification.

### 4.3. Screening stage 3: Ranking

The third screening stage is designed to narrow down a relatively long list of performance improvements to a top selection of five. This top-5 could later be worked out in finer detail.

This second expert session ranks remaining performance improvements, using scoring criteria. Each scoring criterion is scored from 1 (lowest) to 5 (highest). The performance improvements that greatly improve performance at low cost, score highest.

Four screening criteria are applied:

1. Maximum theoretical effect
2. Effectiveness
3. Costs
4. Feasibility

The first screening criterion establishes the theoretical maximum effect of a performance improvement. Generally, this directly results from the importance analysis using the Fussell-Vesely importance measure. In some cases, experts can comment on the chosen maximum effect. For instance, experts can argue that a PPI only reduces long – and not short – repair times. In that case, only the Fussell-Vesely importance measures of long repair times need to be considered rather than all unavailability due to repair.

Table 1. Classes used for scoring criteria in case study.

| Criterion | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|---|---|---|---|---|---|
| Maximum effect | < 5% | 5%–10% | 10%–20% | 20%–40% | >40% |
| Effectiveness | <10% | 10%–30% | 30%–60% | 60%–90% | >90% |
| Cost [M€] | >25 | 5–25 | 1–5 | 0.1–1 | <0.1 |
| Feasibility | Very diff. | Difficult | Normal | Easy | Very easy |

The second screening criterion – effectiveness – is used to judge the improvement percentage given the maximum effect on the probability of failure on demand. For instance, a software replacement affects a set of the minimal cut-sets which together determine 30% of the probability of failure on demand (criterion 1). Experts expect that the software replacement reduces this contribution by about 50% (criterion 2). They use the value of 50% to score the second criterion.

The third screening criterion is cost. These costs are life cycle costs relative to the current situation. The cost classes should be distinctive and are, therefore, case-dependent. The leader proposes initial values for scoring cost based on performance improvements defined in the first session (screening step 2), which is then verified by the experts.

The final screening criterion is feasibility. This criterion aims to capture aspects excluded in the other criteria. For example, a performance improvement can take a long time to implement, be politically difficult or have a negative effect on another function of the object. More important than with other criteria is that these considerations are noted.

The total score is the product of the scores per criterion. The results of the session and the entire process are reported and checked by experts. Table 1 shows the values for the scoring criteria that were applied in the case study. With respect to the first two criteria, it is expected that most performance improvements only affect a portion of minimal cut-sets while sometimes having a large effect on them. Therefore, the highest score for maximum effect is already achieved with percentages higher than 40%, while improvement percentages need to be higher than 90% to achieve the highest score.

## 5. Case study: Maeslant barrier

The Maeslant barrier is a storm surge barrier in the Netherlands that protects Rotterdam and Dordrecht against coastal floods. The barrier consists of two floating sector gates that, in their closed position, span over the 360-meter-wide navigation canal: the New Waterway.

The floating sector gates are operated in the following manner (see Figure 4): In normal conditions, the gates are located in a dry dock. When a storm surge is expected the dry dock and gate are prepared to close by jacking up the ball joint, filling the dock, and opening the dock door. The closure of the New Waterway commences when the sector gates sail out and are horizontally positioned in the New Waterway. The closure is realised by sinking the gates onto a sill. To achieve high reliability, the barrier closes automatically but can be operated by staff in the event of a malfunction.

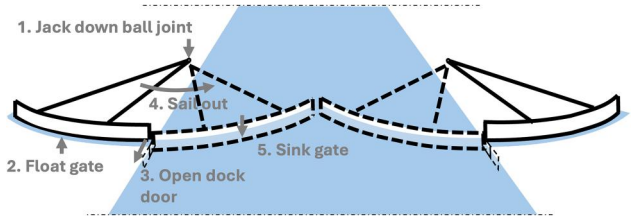


Figure 4. Schematic bird-eye view diagram of Maeslant barrier floating sector gates illustrating the five-step closure procedure. Solid lines sketch the gates in closed position. Dotted lines in closed position. Light blue is used for water surfaces. Grey arrows indicate the main movements to get from the open to closed position.

In this case study, the aim was to reduce the probability of a failure to close. Mooyaart et al. (2025) showed that this failure mechanism is dominant. Moreover, the Maeslant barrier has been struggling to meet the design and later legal requirements for this probability of failure to close. The original design requirement was 1/1000 on demand. Quickly after delivery of the barrier to Rijkswaterstaat in 1997, reliability experts started to doubt whether the required design probability was achieved. Between 2003 and 2008, the probability of a failed closure estimate was reassessed and ultimately determined to be 1 in 100 per request, which was deemed acceptable at the time. Meeting this lower requirement has remained a struggle, resulting in several performance improvements and many model adjustments (Rijkswaterstaat, 2022).

Initially, a case was preferred to best reflect the current situation. In that manner, sensible performance improvements could be suggested to improve the Maeslant barrier. However, two difficulties were encountered. First, due to confidentiality issues, it was uncertain whether study results could be published. Second, more importantly, it proved impossible to understand the current RA analysis, including all the model adjustments made over the last few decades, as the underlying documentation had not been updated.

Therefore, a historic RA analysis of 2005 was selected for the case study. Although this RA analysis also had quality issues, for instance, the name of a component in the system analysis report could be inconsistent with the fault tree report and the fault tree model, this analysis had the highest available quality. An independent consultant reviewed this RA analysis (Horvat, 2006) and deemed it the 'best available'.

Using a historical case also had an important advantage. Making use of a historical case enables comparison with current approaches. The implemented performance improvements can be compared with those proposed by the method. Moreover, if the method identifies other performance improvements, this could also lead to valuable insights for the barrier.

For the case study, only the fault tree model of the RA analysis was used. The 2005 RA analysis consisted of a fault tree model and a separate database with failure scenarios to model human recovery actions. The database is difficult to use, and there is no report on it. As the human recovery actions were excluded from the fault tree model, it was practically impossible to determine Fussell-Vesely importance measures of them. Therefore, the fault tree model was used which finds a probability of a failure to close of 4/100 on demand.

Several aspects were taken into consideration when selecting the experts. First, the expert group needs to have a multidisciplinary technological background, including civil, mechanical, and electrical engineering. Furthermore, expertise in both the design and the operation & maintenance stages was needed. This requirement limited the options for selecting experts. To minimise potential bias due to the historical nature of the case, one expert was included who was not involved in the RA analysis during the period of 2003–2008. Finally, given the study's objective and budget, a group size of more than six was deemed undesirable. These considerations resulted in a selection of six experts, of whom five were able to attend.

### 5.1. Screening stage 1: Importance analysis

The initial phase of the screening process involved sharing a comprehensive report with the expert panel. This report contained:

1. The study's primary objective: identifying performance improvements to achieve the original design requirement of 1/1000 on demand
2. Eight detailed tables presenting importance analyses of:
   - Failure rates
   - Failures on demand
   - Repair times
   - Test intervals and mission times
   - Single events
   - Failure scenarios
   - Object decomposition
   - Top-5 per hierarchical level

Figures 5 and 6 along with Table 2 summarise this information. Here, Table 2 describes the top 3 critical items, for which the Fussell-Vesely Importance Measure is shown in Figure 6. For example, the failure rate belonging to 'object system 1: measurements' has a Fussell-Vesely Importance Measure of just over 6%.

Based on this analysis, two screening criteria were applied. For larger groups of failure events (e.g., object systems, failure categories, unavailability types), a screening criterion of 5% was applied. For smaller groups of failure events (e.g., components, repair times), a criterion of 1% was used. The first criterion excluded all PPI's related to human reliability and external events from the subsequent screening stage (see Figure 5). The second criterion rejected four other PPI's for further evaluation. This approach ensured that only the PPI's that potentially substantially affect the probability of failure on demand progress to the next stage.

Many categories, unavailability types, variables and elements (e.g. object systems, subsystems and components) remained relevant to consider for the next stage. Figure 5 shows that hardware, software and CCF all contributed highly to the probability of failure on demand. Although failures on demand and unavailability due to repair dominate, test intervals and mission times also considerably affect the result. Figure 6 illustrates that all top-5 object systems remained above the 5% screening criterion and all other top-5 items above the 1% screening criterion.

### 5.2. Screening stage 2: Select performance improvements

In this stage, an expert session was conducted to systematically evaluate the remaining PPI's of the checklist. The session involved a multidisciplinary team of five experts. The session leader guided the experts through each PPI, applying the four screening questions as outlined in the method. From the original checklist, 38 performance improvements were selected as potentially applicable and promising based on the response to these screening questions. The experts utilised the Fussell-Vesely Importance Measures to assess the potential impact of each improvement on the barrier's probability of failure on demand. For each selected PPI, the experts provided case-specific examples relevant to the

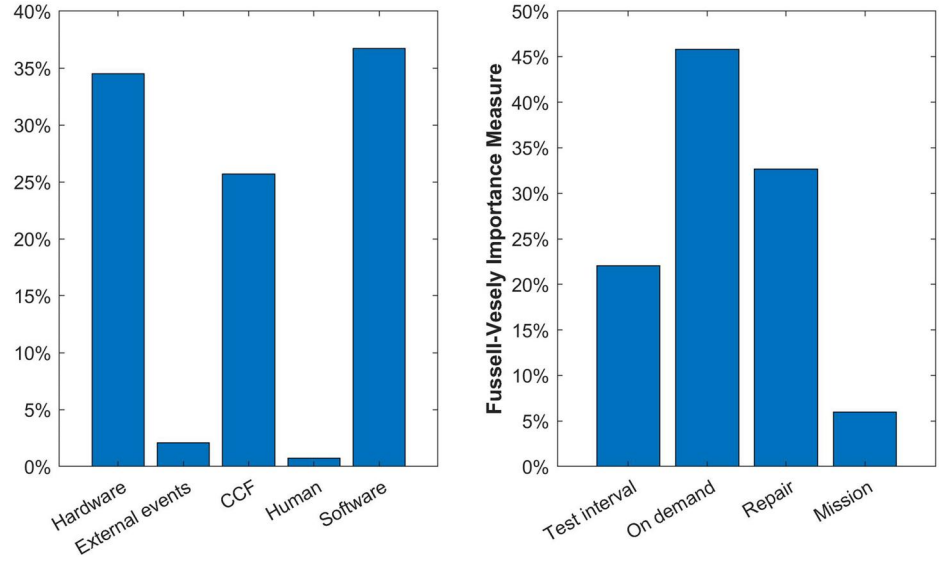


Figure 5. Fussell-Vesely importance of failure categories (left) and unavailability types (right) of the Maeslant barrier case study.

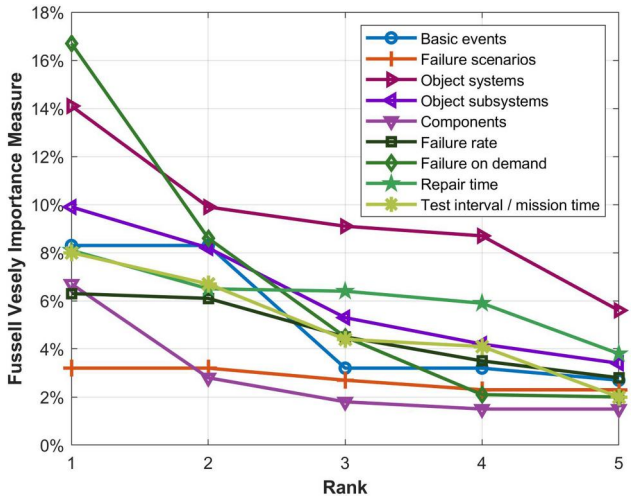


**Figure 6.** Fussell-Vesely importance of top-5 basic events, failure scenarios, elements of object decomposition and variables (failure rate, failure on demand, test interval, mission time and repair time) of the Maeslant barrier case study.

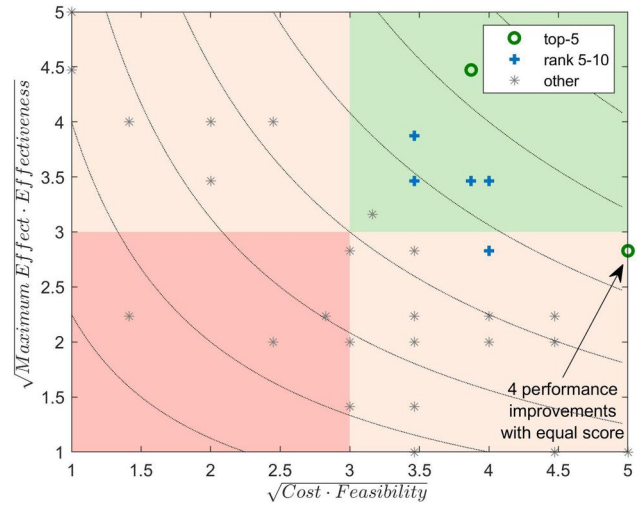


**Figure 7.** Quadrant with ranking results based on scores on cost, feasibility and effectiveness. The top-5 and top-10 are highlighted using green circles and blue crosses, respectively. Background colours indicate whether performance improvements are promising (green) or unattractive (red). Dotted lines are isolines of total scores.

**Table 2.** Top 3 critical items.

| Item | 1st | 2nd | 3rd |
|---|---|---|---|
| Failure rates | OS1: Measurements | OS1-1-1-2 | OS5-1-4: CCF |
| Failure on demand | Dependency SW | SW type 1 | SW type 2 |
| Repair time | OS5: 24h | OS10: 168h, | OS1: 5h |
| Test interval/ Mission time | OS1: 1 month | OS5: 4 months | OS1: 33h |
| Single events | Dependency SW 1 | Dependency SW 2 | Ex Inf 1 |
| Failure scenarios | Ex Inf 1 | Ex Inf 2 | CCF Inf cable |
| Object systems | OS2 | OS1 | OS5 |
| Subsystems | OS1-1 | OS5-1 | OS10-1 |
| Components | OS5-1-4 | OS3-1 | OS6-1-3-1 |

Object systems (OS) and their sub-elements are anonymously numbered. Other abbreviations are: CCF (Common Cause Failure), SW (Software), Ex (External) and Inf (Information).

Maeslant barrier. This process helped to contextualise the generically formulated improvements and ensure their applicability to the specific case study.

At the end of each category discussion, the experts were asked whether any important performance improvements had been missed. Six additional performance improvements were found in this manner. In fact, all the newly found performance improvements were already on the checklist. However, the formulation of the PPI did not trigger the according performance improvement. For instance, continuous monitoring was found unpromising, but 'making dormant failures noticeable' was identified as an additional improvement, while being essentially the same.

In total, 44 performance improvements (38 from the original checklist and six additional) were considered promising enough to progress to the next screening stage. The results of this session, including the rationale for including or excluding each improvement, were documented and shared with the experts for verification.

### 5.3. Screening stage 3: Ranking

In this final screening stage, the expert panel convened to score the remaining performance improvements identified in stage 2. The session followed the method outlined in the paper, using four scoring criteria: maximum theoretical effect, effectiveness, costs, and feasibility. Each criterion was scored on a scale from 1 to 5. Performance improvements with a large effect on the probability of failure on demand at low cost scored highest.

The experts carefully considered each performance improvement, discussing its potential impact, practicality, and associated costs. They used the Fussell-Vesely Importance Measures from the Importance Analysis to inform their judgments on the maximum theoretical effect. For effectiveness, they estimated the improvement percentage given the maximum effect. Costs were evaluated based on life cycle costs relative to the current situation, using the case-specific cost classes established in the previous session. Feasibility scores accounted for factors such as the need for specialised staff, the difficulty in defending less conservative assumptions and organisational/political acceptance.

Figure 7 presents the scores of the experts in a quadrant. The vertical axis shows the score for the expected effect on performance. The horizontal axis indicates costs and feasibility, where a high score represents a low cost and high feasibility. Those performance improvements that score high on both aspects receive the highest overall score and can generally be found in the top right quadrant, which is indicated in green. Other quadrants are coloured orange and red to indicate medium-favourable and unfavourable performance improvements, respectively. As these quadrants do not accurately reflect the total score, isolines were included in the figure. More information on the names and scores of the performance improvements can be found in Supplementary Appendix B. The scoring process resulted in the following top-5 performance improvements:

1. Implement human recovery actions for software failures.
2. Test object system 5 every month in stead of every four months

3. Jack-up the ball joint earlier in the closure procedure
4. Measure and update repair times
5. Establish more realistic common-cause failure probabilities

It is highlighted that the performance improvements found are a result of an old RA analysis and are not applicable to the current situation of the Maeslant barrier as most performance improvements were already implemented.

This set of performance improvements will not result in a probability of failure on demand to meet the original design criterion; rather, it reduces this probability by approximately a factor of three. Three approaches could be considered to come closer to the design criterion: 1) implementing performance improvements lower on the ranking, 2) a new screening procedure based on a revised RA analysis with implemented performance improvements or 3) a combination of these approaches.

Another option is to reconsider the design criterion. In 2005, this option was chosen and resulted in a criterion of a probability of failure of 1/100 on demand. Further results of the case study are reported in Supplementary Appendix B.

### 5.4. Adjustments to the checklist based on the case study

Based on the case study, the list of PPI's was adjusted. PPI's were clustered, especially in the common cause failure category as the applied list is fairly time demanding to navigate through without creating different performance improvements. Some of the model conservatism reductions were repeated at the end of a category. For instance, adjusting the model variable was introduced at the maintenance sub-categories (repair time, dormant failures), as an important PPI (using more realistic repair times in the model) was almost missed. Based on this choice, it was decided to move some other model conservatism reductions to another category. For instance, adjusting the model used for software was transferred to the software reliability category. Finally, the wording of some of the PPI's was adjusted if their concept was not well understood during the expert session. Supplementary Appendix A presents the final version of the checklist.

## 6. Discussion

This study primarily shows that current RA analysis practices have a critical weakness: they are likely to fail at detecting important performance improvements and, moreover, tend to employ unstructured and non-transparent procedures for identifying such improvements. These limitations are deeply rooted in current practices such as FMECA, HAZOP and RA analysis methods, which are primarily component/event-driven. As a result, they do not naturally reveal system-level or cross-system performance improvements, such as operational adjustments and reduced model conservatism. Moreover, these practices only look at risk and ignore the costs and feasibility of performance improvements. As a result, they focus on component/events with a large contribution to overall risk and not at identifying where safety systems can be improved effectively.

For example, raising the test frequency of object system 5 was one of the most effective performance improvements in the case study. This enhancement could not be identified through conventional ranking of failure scenarios because it influenced numerous components and, consequently, multiple failure scenarios. From a risk-oriented perspective alone, neither object system 5 nor its test interval appeared particularly prominent (the Fussell–Vesely Importance Measure was 9% for the object system, ranking third, and 6% for its test interval, ranking second). The key factor was instead the relatively low cost and minimal operational burden of the improvement. Although ultimately implemented, this measure was adopted only after several years, and only one of the five experts was aware of its implementation. This example illustrates not only inefficiencies in current RA practice but also the lack of transparency in decision support.

The proposed method is more systematic approach than current practices. The list of PPIs supports a more structured identification of performance killers, and the categorisation of PPIs helps organise expert discussions, reducing the likelihood that important categories of improvements are overlooked. Because the method is more systematic, it is also more efficient: deriving a top-five list of case-specific improvements required two expert sessions of one day each, spaced over a month, whereas historically the same set of improvements emerged only after multiple iterations over several years. However, since only a single case study was conducted, broader claims about the method's credibility cannot yet be made.

The method's results highly depend on its input: the RA analysis. The twenty-year-old RA used in the case study exhibits several of the quality issues documented by Rae et al. (2014) and lacks transparency regarding uncertainties, as highlighted by Aven and Zio (2011). More fundamentally, there is limited scientific literature on RA applications for storm surge barriers, despite their unique combination of characteristics: low-frequency operation, substantial heterogeneity across barriers, and operation in storm conditions in hazardous marine environments.

Parts of the method can also be used to set-up or refine an RA analysis. Already at the scoping phase, the list of PPI's could be used to determine which types of improvements the RA analysis should be able to identify. The presented categorisation can assist in structuring the RA analysis, for instance, to define what is considered a component. In the case study, this was not always clear. Furthermore, the method could help in establishing the level of detail. With this method, experts can determine in which parts of the model more detail is needed to find relevant performance improvements. In the case study, for example, improving the modelling on common-cause failures was identified as the most promising modelling effort.

Expert judgement plays a central role in selecting performance improvements, as experts must evaluate both the expected increase in performance and the effort required to achieve it. Consistent with observations by Kletz (1992), it is

expected that expert selection and session design strongly influence the resulting top-five improvements. In the case study, the reliance on a twenty-year-old RA has biased expert judgments. For instance, modifying the software reliability model received a low score because experts were now aware that this measure had been less effective than anticipated two decades earlier. Additional case studies are therefore needed to evaluate the influence of expert selection and session structure. Methods such as Delphi (Dalkey & Helmer, 1963) or Cooke's classical model (Cooke, 1991) could be explored to obtain more objective results, although these approaches may not necessarily address the interdisciplinary complexity of the present problem.

The scoring method also affects the ranking of identified improvements. The Fussell–Vesely Importance Measure (FVIM) plays a dominant role in the current approach. Alternative importance measures might be informative; however, many do not readily apply to groups of failure scenarios (e.g., the Birnbaum Importance Measure). Even newer measures such as the Differential Importance Measure (DIM), despite theoretical advantages, may not be ideal, as some experts already find the FVIM difficult to interpret. Given that the RA includes probabilistic distributions for failure events, incorporating uncertainty importance measures may be valuable. Uncertainty applies not only to risk estimates but also to the performance effects of improvements themselves. For instance, the benefit of a model enhancement is typically unknown a priori. Finally, some improvements primarily reduce uncertainty in expected probabilities rather than the expected probabilities themselves. Ideally, the scoring method would value this aspect of these type of improvements as well.

Although the method was demonstrated using a single case study, its underlying principles are not specific to the Maeslant barrier. The characteristics that motivated the development of the screening framework—low-frequency operation, complex human-technical interactions, and mixed hardware–software dependencies—are shared by storm surge barriers worldwide. Consequently, the checklist and screening stages are expected to be applicable to other Dutch barriers (e.g., Oosterscheldekering, Ramspolkering) as well as the Thames Barrier (London, UK) or New Orleans flood control structures (LA, USA). Likewise, the method might be valuable for safety-critical infrastructures beyond the flood-defense domain such as nuclear power, offshore energy, and chemical processing.

## 7. Conclusions

This paper proposes a method to systematically identify performance improvements for storm surge barriers using a checklist. The method consists of 3 screening stages: one desk study and two structured expert evaluations. This checklist contains 85 principal performance improvements (PPI's), which were subdivided into eight categories: Model conservatism, hardware, common-cause failure, human error, software error, maintenance, and operation. This checklist was developed based on a literature review, dialogues with co-authors and Rijkswaterstaat colleagues, systematic structuring, and the Maeslant barrier case study (in Rotterdam, the Netherlands).

The case study demonstrates that the checklist is, at the very least, nearly complete. Together with the screening method, it is effective in deriving case-specific performance improvements. No additional PPIs were identified in the case study, indicating that the steps taken to develop the checklist (literature study, dialogues, and systematic structuring) were thorough. The process of deriving a top-5 list of case-specific performance improvements required two expert sessions of a day over a period of one month. In contrast, the historical process to derive these improvements required multiple iterative steps over a period of several years.

Compared to current methods, this screening method is more systematic in finding and selecting performance improvements. Especially low-cost performance improvements above the component level with little organisational burden are more easily found. Although the checklist and its screening method were found promising for other barriers, this paper contains only one case study, and more case studies are required to improve the credibility of the method.

## Note

1. Dormant failure refers to a failure which was undetected during normal conditions.

## Acknowledgements

Tycho Busnach and the anonymous experts are thanked for their contribution to this research.

## Author contributions

CRediT: **Johan A. van den Bogaard**: Conceptualization, Methodology, Supervision, Writing – review & editing; **Alexander M.R. Bakker**: Supervision, Writing – review & editing; **Martine van den Boomen**: Supervision, Writing – review & editing; **Sebastiaan N. Jonkman**: Supervision, Writing – review & editing.

## Disclosure statement

The authors report there are no competing interests to declare.

## Funding

The research was funded by the Dutch Research Council (NWO) and Rijkswaterstaat.

## References

Apostolakis, G. E. (2004). How useful is quantitative risk assessment? *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *24*(3), 515–520. https://doi.org/10.1111/j.0272-4332.2004.00455.x

Aven, T. (2016). On the use of conservatism in risk assessments. *Reliability Engineering & System Safety*, *146*, 33–38. https://doi.org/10.1016/j.ress.2015.10.011

Aven, T. (2017). Improving the foundation and practice of reliability engineering. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *231*(3), 295–305. https://doi.org/10.1177/1748006X17699478

Aven, T., & Heide, B. (2009). Reliability and validity of risk analysis. *Reliability Engineering & System Safety*, *94*(11), 1862–1868. https://doi.org/10.1016/j.ress.2009.06.003

Aven, T., & Renn, O. (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, *12*(1), 1–11. https://doi.org/10.1080/13669870802488883

Aven, T., & Zio, E. (2011). Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering & System Safety*, *96*(1), 64–74. https://doi.org/10.1016/j.ress.2010.06.001

Aven, T., & Zio, E. (2013). Model output uncertainty in risk assessment. *International Journal of Performability Engineering*, *9*(5), 475–486.

Barnes, M. J., Bley, D., & Cooper, S. (2000). *Technical basis and implementation guidelines for a technique for human event analysis (ATHEANA)* (TechRep). http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Technical+Basis+and+Implementation+Guidelines+for+A+Technique+for+Human+Event+Analysis#0

Bier, V. M. & Cox Jr., L. A. (2007). Probabilistic risk analysis for engineered systems. In *Advances in decision analysis, from foundations to applications*. (1st ed.). Cambridge University Press.

Birnbaum, Z. (1969). On the importance of different components in a multicomponent system. In P.R. Kreshnaiah (Ed.), *Multivariate analyses II* (pp. 581–592). Academic Press.

Borgonovo, E., & Apostolakis, G. E. (2001). A new importance measure for risk-informed decision making. *Reliability Engineering & System Safety*, *72*(2), 193–212. https://doi.org/10.1016/S0951-8320(00)00108-3

Cheok, M. C., Parry, G. W., & Sherry, R. R. (1998). Use of importance measures in risk-informed regulatory applications. *Reliability Engineering & System Safety*, *60*(3), 213–226. https://doi.org/10.1016/S0951-8320(97)00144-0

Cooke, R. (1991). *Experts in uncertainty: Opinions and subjective probability in science*. Oxford University Press.

Dalkey, N., & Helmer, O. (1963). An Experimental Application of the DELPHI Method to the Use of Experts. *Management Science*, *9*(3), 458–467. https://doi.org/10.1287/mnsc.9.3.458

Dummer, G., Tooley, M., & Winton, R. (1997). *An elementary guide to reliability* (5th ed.). Elsevier.

Dutuit, Y., & Rauzy, A. (2015). On the extension of Importance Measures to complex components. *Reliability Engineering & System Safety*, *142*, 161–168. https://doi.org/10.1016/j.ress.2015.04.016

Goerlandt, F., Khakzad, N., & Reniers, G. (2017). Validity and validation of safety-related quantitative risk analysis: A review. *Safety Science*, *99*, 127–139. https://doi.org/10.1016/j.ssci.2016.08.023

Gomes, W. J. D S., & Beck, A. T. (2021). A conservatism index based on structural reliability and model errors. *Reliability Engineering & System Safety*, *209*, 107456. https://doi.org/10.1016/j.ress.2021.107456

Haimes, Y. Y., Kaplan, S., & Lambert, J. H. (2002). Risk filtering, ranking, and management framework using hierarchical holographic modeling. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *22*(2), 383–397. https://doi.org/10.1111/0272-4332.00020

Hallegatte, S., Green, C., Nicholls, R. J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. *Nature Climate Change*, *3*(9), 802–806. http://www.nature.com/doifinder/10.1038/nclimate1979 https://doi.org/10.1038/nclimate1979

Hegseth, J. M., Bachynski, E. E., & Leira, B. J. (2021). Effect of environmental modelling and inspection strategy on the optimal design of floating wind turbines. *Reliability Engineering & System Safety*, *214*, 107706. https://doi.org/10.1016/j.ress.2021.107706

Henley, E. J., & Kumamoto, H. (1981). *Reliability engineering and risk assessment*. Preston Hall.

Heslinga, G. (2004). *Report failure probability model human error (Dutch: Rapport faalkansmodel menselijke fouten), ref: OKE-2005-258-T* (TechRep.). KEMA.

Horvat, P. (2006). *Second Opinion Failure Probability Maeslant barrier (Dutch: Second opinion Faalkans Maeslantkering)* (Tech. Rep.)

Iaea. (1992). *Procedures for Conducting Common Cause Failure Analysis in Probabilistic Safety Assessment: IAEA TECDOC Series No. 648* (Tech. Rep.).

International Electrotechnical Commission. (2016). *Hazard and operability studies (HAZOP studies)*. Application guide (IEC 61882) (Tech. Rep.).

International Electrotechnical Commission. (2018). *Failure modes and effects analysis (FMEA and FMECA) (IEC 60812)* (Tech. Rep.).

International Electrotechnical Commission. (2021). *Safety of machinery: Functional safety of safety-related control systems (IEC Standard no. 62061)* (Tech. Rep.). https://www.nen.nl/nen-en-iec-62061-2021-en-fr-285780

Jonkman, S. N., Voortman, H. G., Klerk, W. J., & van Vuren, S. (2018). Developments in the management of flood defences and hydraulic infrastructure in the Netherlands. *Structure and Infrastructure Engineering*, *14*(7), 895–910. https://doi.org/10.1080/15732479.2018.1441317

Kamps, M., van den Boomen, M., van den Bogaard, J., & Hertogh, M. (2024). Intergenerational transfer of engineering expertise: Knowledge continuity management in storm surge barrier engineering. *Built Environment Project and Asset Management*, *14*(6), 874–891.), https://doi.org/10.1108/BEPAM-10-2023-0179

Kharoubi, Y., van den Boomen, M., van den Bogaard, J., & Hertogh, M. (2024). Asset man-agement for storm surge barriers: How and why? *Structure and Infrastructure Engineering*, *2024*, 1–15. https://doi.org/10.1080/15732479.2024.2391034

Kiel, E. S., Catrinu-Renstrøm, M. D., & Kjølle, G. H. (2023). A Transformer Outage Duration Model with Application to Asset Management Decision Support. In *Esrel 2023*, pp. 2215–2222.

Kletz, T. (1992). *HAZOP and HAZAN* (3rd ed.). Institution of Chemical Engineers.

Knochenhauer, M., & Louko, P. (2003). *Guidance for External Events Analysis* (Tech. Rep. No. February). Swedish Nuclear Inspectorate (SKI).

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, *42*(4), 237–270. https://doi.org/10.1016/S0925-7535(03)00047-X

Lewin, J., Ballard, G., & Bowles, D. S. (2003). Spillway gate reliability in context of overall dam failure risk. In *USSD annual lecture* (pp. 1–17).

Li, Y., Zhang, W., Liu, B., & Wang, X. (2024). Availability and maintenance strategy under time-varying environments for redundant repairable systems with PH distributions. *Reliability Engineering & System Safety*, *246*, 110073. https://doi.org/10.1016/j.ress.2024.110073

Liu, X., Fang, Y. P., Ferrario, E., & Zio, E. (2021). Resilience Assessment and Importance Measure for Interdependent Critical Infrastructures. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, *7*(3), 4051196. https://doi.org/10.1115/1.4051196

Lokuge, W., Wilson, M., Tran, H., & Setunge, S. (2019). Predicting the probability of failure of timber bridges using fault tree analysis. *Structure and Infrastructure Engineering*, *15*(6), 783–797. https://doi.org/10.1080/15732479.2019.1569069

Mooyaart, L. F., Bakker, A. M., van den Bogaard, J. A., Jorissen, R. E., Rijcken, T., & Jonkman, S. N. (2025). Storm surge barrier performance—The effect of barrier failures on extreme water level frequencies. *Journal of Flood Risk Management*, *18*(1), 1–17. https://doi.org/10.1111/jfr3.13048

Mooyaart, L. F., Bakker, A. M. R., van den Bogaard, J. A., Rijcken, T., & Jonkman, B. (2023). Economic optimization of coastal flood defence systems including storm surge barrier closure reliability. *Journal of Flood Risk Management*, *16*(3), e12904. https://onlinelibrary.wiley.com/doi/10.1111/jfr3.12904 https://doi.org/10.1111/jfr3.12904

Mooyaart, L. F., & Jonkman, S. N. (2017). Overview and Design Considerations of Storm Surge Barriers. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *143*(4), 06017001. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000383

Mutar, E. K., & Hassan, Z. A. H. (2025). Survival Signature-Based Structural Importance Analysis of Multistate System with Binary-

State Components. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, *11*(2), 4067826. https://doi.org/10.1115/1.4067826

NRG. (2003). *Reliability analysis OKE (Dutch: Betrouwbaarheidsanalyse OKE)* (Tech. Rep.).

Nuclear Energy Institute. (2018). *NUMARC 93-01 revision F, Industry guideline for monitoring the effectiveness of maintenance at nuclear power plants* (Tech. Rep.). https://www.nrc.gov/docs/ML1812/ML18120A069.pdf

NUREG. (2004). *Good Practices for Implementing Human Reliability Analysis (HRA)* (Tech. Rep.). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.194.8511&rep=rep1&type=pdf

Parry, G. W. (1991). Common cause failure analysis: A critique and some suggestions. *Reliability Engineering & System Safety*, *34*(3), 309–326. https://doi.org/10.1016/0951-8320(91)90106-H

Pasman, H. J., Rogers, W. J., & Mannan, M. S. (2017). Risk assessment: What is it worth? Shall we just do away with it, or can it do a better job? *Safety Science*, *99*, 140–155. https://doi.org/10.1016/j.ssci.2017.01.011

Paté-Cornell, E. (2002). Finding and fixing systems weaknesses: Probabilistic methods and applications of engineering risk analysis. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *22*(2), 319–334. https://doi.org/10.1111/0272-4332.00025

Paula, H., Campbell, D., Parry, G., Mitchell, D., & Rasmuson, D. (1990). *A cause-defense approach to the understanding and analysis of common cause failures, NUREG/CR-5460* (Tech. Rep.). Springfield: NUREG. http://inis.iaea.org/search/search.aspx?origq=RN:34002229

Pitblado, R. (1994). Quality and offshore quantitative risk assessment. *Journal of Loss Prevention in the Process Industries*, *7*(4), 360–369. https://doi.org/10.1016/0950-4230(94)80050-2

Rae, A., Alexander, R., & McDermid, J. (2014). Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment. *Reliability Engineering & System Safety*, *125*, 67–81. https://doi.org/10.1016/j.ress.2013.09.008

Rijcken, B. T., Mooyaart, L., Oerlemans, C., Ziel, F. V. D., Borghans, J., Botterhuis, T., & Kok, M. (2023). It takes two to tango with Sea level rise. *TU Delft DeltaLinks*, 1–16. https://flowsplatform.nl/#/it-takes-two-to-tango-with-sea-level-rise-1682542508929

Rijkswaterstaat. (2011). *Leidraad Risicogestuurd Beheer en Onderhoud (conform de ProBo werkwijze)* (Tech. Rep.). https://puc.overheid.nl/rijkswaterstaat/doc/PUC14394931/

Rijkswaterstaat. (2018a). *Handreiking prestatiegestuurde risicoanalyses (PRA)* (Tech. Rep.). https://www.leidraadse.nl/assets/files/latestversion/Downloads/RWS/RAMS/1 Handreiking PRA/LeidraadPRA vigerend.pdf

Rijkswaterstaat. (2018b). *TOPAAS. Deel 1 Handleiding (Kader)* (Tech. Rep.).

Rijkswaterstaat. (2022). *Legislated assessment Europoort barrier (Dutch Wettelijke beoordeling Europoortkering I Dijktraject 208)* (Tech. Rep.).

Rijneveld, B. (2008). *A study into failure probability reducing measures at the Maeslant barrier (Dutch: Een studie naar faalkansreducerende maatregelen voor de Maeslantkering)* (Tech. Rep.). Delft University of Technology.

Rusnak, P., Zaitseva, E., Levashenko, V., Bolvashenkov, I., & Kammermann, J. (2024). Importance analysis of a system based on survival signature by structural importance measures. *Reliability Engineering & System Safety*, *243*, 109814. https://doi.org/10.1016/j.ress.2023.109814

Scherb, A., Garre, L., & Straub, D. (2017). Reliability and component importance in net-works subject to spatially distributed hazards followed by cascading failures. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, *3*(2), 6.

Schüller, J., Brinkman, J., van Gestel, P., & van Otterloo, R. (1997). *Methods for determining and processing probabilities. Red Book* (Tech. Rep.). Committee for the Prevention of Disasters.

Swain, D., & Guttmann, H. E. (1983). *Handbook of reliability analysis with emphasis on nuclear plant applications* (Tech. Rep. No. August).

ten Veldhuis, J. A., Clemens, F. H., & van Gelder, P. H. (2011). Quantitative fault tree analysis for urban water infrastructure flooding. *Structure and Infrastructure Engineering*, *7*(11), 809–821. https://doi.org/10.1080/15732470902985876

Thekdi, S., & Aven, T. (2026). In What Cases Are Gaps in Risk Study Quality Unacceptable? *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *12*(1). https://ascelibrary.org/doi/10.1061/AJRUA6.RUENG-1602 https://doi.org/10.1061/AJRUA6.RUENG-1602

Trace-Kleeberg, S., Haigh, I. D., Walraven, M., & Gourvenec, S. (2023). How should storm surge barrier maintenance strategies be changed in light of sea- level rise? A case study. *Coastal Engineering*, *184*, 104336. https://doi.org/10.1016/j.coastaleng.2023.104336

United States Nuclear Regulatory Commission. (2011). *Development of Quantitative Software Reliability Models for Digital Protection Systems of Nuclear Power Plants* (No. May).

US Nuclear Regulatory Commission. (2006). *Evaluation of Human Reliability Analysis Methods Against Good Practices (NUREG-1842)* (Tech. Rep.). https://www.nrc.gov/docs/ML0632/ML063200058.pdf

Vader, H., Bakker, A. M., Jonkman, S. N., van den Boomen, M., van Baaren, E., & Diermanse, F. L. (2024). A framework for assessing the remaining life of storm surge barriers. *Structure and Infrastructure Engineering*, *20*(12), 2022–2034. https://doi.org/10.1080/15732479.2023.2177874

Van Der Weide, J. A., & Pandey, M. D. (2015). A stochastic alternating renewal process model for unavailability analysis of standby safety equipment. *Reliability Engineering & System Safety*, *139*, 97–104. https://doi.org/10.1016/j.ress.2015.03.005

Van Manen, S., Brandt, E., Van Ekris, J., & Geurts, W. (2015). TOPAAS: An alternative approach to software reliability quantification. *Quality and Reliability Engineering International*, *31*(2), 183–191. https://doi.org/10.1002/qre.1570

Vesely, W. E., Davis Thomas, C., Denning, R. S., & Saltos, N. (1983). *Measures of Risk Importance and Their Applications, NUREG/CR-3385, BMI-2103* (Tech. Rep. No. May).

Xu, Z., Ramirez-Marquez, J. E., Liu, Y., & Xiahou, T. (2020). A new resilience-based component importance measure for multi-state networks. *Reliability Engineering & System Safety*, *193*(2006), 106591. https://doi.org/10.1016/j.ress.2019.106591