JOURNAL OF COASTAL AND RIVERINE FLOOD RISK

Review and rebuttal of the paper

BREACH-METHOD: a new framework to generate event sets for financial flood risk assessment of the Dutch Delta

B. Kolen and R. Nicolai

Editor handling the paper: Jeremy Bricker

The reviewers remain anonymous.

Round 1
Reviewer A:

The paper presents a new methodology for flood risk assessment in the Netherlands, arguing that current flood risk estimates are overly conservative. The authors claim that their approach leads to less conservative flood probability estimates by addressing three factors that, in their view, inflate current risk estimates: (1) the assumption of independent failures, (2) overestimated failure probabilities, and (3) exclusion of emergency measures. The methodology prominently relies on expert judgment to adjust risk estimates, particularly in scenarios involving multiple breaches.

Major concerns

1. The paper is grounded on the premise that current flood risk assessments in the Netherlands overestimate the level of risk, but this claim is not substantiated by logically sound arguments, mathematical proofs, or sufficient evidence. Instead, the authors emphasize this claim through strong assertions that lack rigorous justification. For the reader to be convinced of the need for the proposed methodology, the authors must provide detailed analytical support, which is currently absent. The authors need to provide clear evidence or a mathematical foundation to substantiate the claim that current assessments are overly conservative before proposing their less conservative approach.

   We have improved the structure of the paper and supported the analyses with evidence. We have improved the description of the current flood risk management approach and the assumptions made in this approach (see 2.2). The current method was developed to define the safety standards for levees, and supported by literature for the impact of system behavior (see Dupuits 2019). This method was also used for several maps which are used by the financial sector. In section 2.2.4 table 5 we present the calculated probability of flooding in The Netherlands if the risk analyses is applied using this method. Below this table we give clear arguments why this flood probability of NL is too high with references to past events and levee reinforcements. Also supporting statements of other literature (De Bruijn et al 2014 and Curran 2020) is presented.

   For the Netherlands there is no evidence based on a historical record of flood events. The only statement that can be made is that the current estimation of flood risk is too high, and in a later stage the results of the new method are plausible.

2. The rationale behind the suggested steps in the new methodology is not provided. It is unclear how the authors developed the methodology or why specific steps were included. The reader is left with no understanding of the theoretical foundation or logic behind the proposed approach. The paper should explicitly explain the reasoning behind the suggested methodology. Each step should be justified and related to the existing body of knowledge in flood risk assessment.

   We added a new section in which we describe the philosophy of the model (4.1). We also improved and restructured the description of the current method in section 2 and impact of application of this method in other fields in section 3. This will improve the understanding.

3. The authors argue that three factors inflate flood risk estimates in the Netherlands; the assumption of independent failures, overestimated failure probabilities, and the exclusion of emergency measures. However, only the assumption of independent failures appears to be explicitly addressed in the methodology, primarily by considering multiple breaches. The other two factors are either entirely excluded from the analysis or are implicitly incorporated via expert judgment, which is not sufficiently explained. The authors need to provide clarity on how overestimated failure probabilities and exclusion of emergency measures are accounted for. If they are included implicitly through expert judgment, this needs to be explicitly stated and properly explained.

   We have added a better explanation in the paper, and restructured the paper. We combined the impact of system behavior (interdependency), emergency measures and overestimated failure probabilities (because of the failure definition) in one expert judgment approach to define the probability of 0, 1, 2, 3 or no breaches per class of return period (see section 4.5 for details).

4. The authors claim that including multiple breach scenarios results in less conservative flood risk estimates. This claim is counterintuitive, as one would expect that multiple failures would lead to higher consequences, and the probability of multiple breaches would be lower or equal to the probability of a single failure. Despite this, no mathematical proof or theoretical explanation is provided to support their argument. The authors need to provide a mathematical proof or a sound explanation of how incorporating multiple breach scenarios can result in less conservative estimates of flood risk. Otherwise, this claim remains speculative and unconvincing.

   The statement of multiple breach scenarios in the original paper was not correctly described. We have improved this. The multiple beach scenarios do not influence the probability of flooding in The Netherlands or in a part of the Netherlands. Adding multiple beaches does influence the way the probability mass is subdivided over all possible scenarios. We also improved the description of the current method and the use of multiple breach scenarios in the current method.

5. The methodology heavily relies on expert judgment to estimate flood risk. However, the process for eliciting expert input is unclear. The Delphi method is mentioned, but key details about the expert elicitation process are missing. What specific questions were posed to the experts? How were biases and subjectivity minimized during the elicitation process? How were the experts chosen, and what were their qualifications? Why were only five experts deemed sufficient, and what was the variability in their responses? Without addressing these critical points, the use of expert judgment remains questionable and could lead to subjective, biased, or unreliable estimates. The authors must provide a transparent and thorough description of the expert elicitation process, including the selection of experts, the questions posed, and how biases were addressed. A larger pool of experts or more detailed analysis of variability in their responses should also be considered to improve the robustness of the estimates

   It is correct that the method relies on expert judgement, however no other (more physics based) methods are available. We have made a reference to the questions asked to the experts in appendix 1. We also described the selection of experts and the process in more detail. Biases are addressed during the Delphi session to discuss and adjust the estimation.

   We agree on advise to increase the number of experts, maybe even apply other expert judgements

methods and improve the understanding about the physics. In the concluding remarks we reflect on these elements. In our opinion this is next step for future research which can improve the method. The new method however is already adopted by some financial institutes and therefore important to publish in scientific journals so new improvements can be added.

Reviewer B:

I enjoyed reading the manuscript 'New framework to generate event set for risk based spatial planning and financial decisions for Dutch Flood' (something missing at end of title though?). I believe it is an important topic as determining the flood risk of the Netherlands as a whole is surprisingly complicated and dependent on many nuanced factors. As such, I fully support eventual publication of this study. However, as it stands I have various comments that need to be addressed in my opinion. Some are on clarity, but some are fundamental on the line of reasoning and approach. I put many comments in a separate word document but will state some overarching things here:

1. The authors state that the current estimates of Dutch flood risk are too high because of three reasons: 1) Independence of failure events; 2) Probability of failure; 3) Exclusion of emergency measures. Reading through the methods, I only find the method going into the development of a database of joint breach scenarios, so addressing point 1. I don't see how point 2 and 3 are addressed in the new methodology.

   We have restructured the paper on this. The expert judgement combines biases for 1) Independence of failure events; 2) Probability of failure; 3) Exclusion of emergency measures. We have better explained the current method (section 2) and the impact of this method for other fields of application (section 3). In section 4 we described the new method and improved the explanation about the followed steps.

2. If and how climate change is incorporated is very unclear to me. This includes what information it is based on, what time horizon (I read in 2050, so 25 years, but also 50 years and 100 years time), and on what variable it is applied (on hydraulic load, on probability, and how?), how for rivers and how for coast (as that would probably require different source data).

   We have improved the explanation in the manuscript on this. For the situation in 2050 we took climate change into account. The frequence of flood events increase, eg a 1/1.000 py event in the current situation could be a 1/500 py event in 2050 because of climate change. This increase in frequency is based on hydrographs which describe the impact of climate change. This is applied for rivers, lakes and coastal areas.

   This approach is used because no flood scenarios for 2050 are available. Therefore we adjusted the probability of the flood events for climate change.

   However more important is the levee reinforcement program based in the new safety standards. This results for many levee sections in a large reduction of the failure probability. The factor in reduction of the failure probability is far more than the increase of the consequences.

   To make it complex, for 2050 we assumed for all levees that the failure probability is equal to the

safety standard. This is also a bit pessimistic because of several reasons mentioned in the paper. Most important is that some levees are very strong and high now and will not be lowered in the future. That is why for some location the risk can go up compared to the current situation. However in general the risk goes down because of the levee reinforcement and despite climate change.

3. The authors state that assuming that all levee sections is 'conservative', implying that it overestimates the real probability of a flood w.r.t. including multiple breach scenarios. However, I don't see how including multiple breaches (or even conditional probabilities) would result in lower actual probabilities of flooding. I would say that the main thing that changes is the potential extent associated with a flood event (as multiple breaches are possible). This would impact the flood risk (due to larger extent), but not the occurrence of 'a flood'. As such, this first point (independence of failure events) is not 'conservative' (a term I would avoid anyway as it is ambiguous as it depends very much on what it applies to); if anything it underestimates risk (in euro/yr) but does nothing for the probability of 'a flood'.

   We improved the explanation of multiple breaches because is was not clear (see also comment of reviewer A). We describe the use of multiple breaches in the current method as well (so the use of multiple breach scenarios is not new). The impact on the failure probability of flooding in the Netherlands is because we used classes for different return periods in which possible scenarios are combined. In these classes unique events are defined based on the expert judgment estimation of 1, 2, 3 (and 0) breach events. The mass of the probability is distributed over all possible scenario in a class. Therefore the use of multiple breach scenarios does not result in an increase in flood probability in the Netherlands. However this does result in more insight about correlations that different properties are exposed during the same event.

4. The authors state that they consider 'system behaviour' in the method. I do not think that is the case. They account for the possibility of multiple breaches occurring (Table 2). But this is not the same as system behaviour. The crux of system behaviour is that if one breach happens, the probability of breaching for other locations changes. I did not see how this was accounted for in their method.

   We have better structured the paper so this is more clear. We combined the impact of system behavior, emergency measures and failure definition. This is translated in an impact to the flood probabilities. In table 5 the impact is shown.

   We did not take the impact of the flood extend because of system behavior into account (as well as emergency measures and the pathway of failure). This could results in (in general) less extreme flood extents. Also for the multiple breach events we combined the maximum depths based on individual scenario's. For some cases this could result in higher water depths. We did not take this into account because these scenarios are not available. In the concluding remarks we reflected in these topics.

5. The core of the results (Table 5) is very interesting, but I believe some interpretation is missed, which leads to some important naunces/questions. Probability of a flood anywhere in the Netherlands goes down (1/3 to 1/50). The reason for this is not explicitly explored. Just considering multiple breach events can not result in this (it would not reduce the amount of flooding, just the poten-

tial extent if there is more than 1 breach). Rather, I believe this is the result of the probability of 0 breaches as indicated by the experts (resulting in a setting where the safety level is reached in terms of hydraulic load, but no breach occurs). This is key to the interpretation of the results as it (potentially) identifies a very specific sensitivity (e.g. the estimated 0 breach probability) that should be explored.

Thanks for the feedback. We have improved the explanation in the paper. Because these areas are assumed to be fully independent the flood probability of the Netherlands is the combination of all these areas. The assumptions (and biased) about the current (conservative) method en the new method are better described in the paper.

6. Related to the previous point. The authors also state that when they aggregated areas (e.g. all three coastal areas), the probability of flooding (at least one flood anywhere in the Netherlands) reduced (to 1/89). This should not be possible. More aggregated areas should just mean that breaches can occur at the same time in places further apart. This should not affect the probability of at least one breach in the Netherlands (in general multiple breaches should not affect this metric; it would affect flood risk in euro/yr though as extents can be larger). My guess is that this is the result of a different (independent) set of tables that were provided by the 5 experts for the more aggregated situation. And that a different probability of 0 breaches reduced the overall probability of at least one flood event in the Netherlands metric. This relates back to the sensitivity of the method to the expert estimates; something I feel should definitely be addressed (preferably by a sensitivity analysis which allows ranges of probabilities to be reported).

The 1/89 py is defined assuming fully dependency between the two river areas as well as between the three coastal areas. (The highest probability remains and we end up with four independent areas.) This is because these can be caused by the same extreme weather event in the river catchment of the same storm surge. In the assessment of the probability of flooding in NL for the river (as well as for the coast) the highest probability of flood per area is used for the entire river or coastal system. This reflection is added because the choice to consider all areas fully independent is of course not correct, but fully dependent is also not correct. Reality will be in between and the current state of the levees and the last flood of 1953 the new method is plausible.

The impact of the use of different classes, as well as the probability of 0 breaches, is described in a better way in section 4. This should be more clear now.

Round 2

Dear editor

Thanks for the second review. We have improved the manuscripts based on the input of the reviewers. In this letter we respond per item of the review how we improved the manuscript. We have uploaded 2 versions:

- *B-jcrfr-review-assignment-7782-Article+Text-28686_ review 2 before text correction with track changes. This version holds all the changes based on the input of the reviewers.*
- *B-jcrfr-review-assignment-7782-Article+Text-28686_ review 2 final. This version includes the check for English writing and grammar. In this versions also the formulas used are included.*

Kind regards

The Authors

Reviewer A:
I note that this is the same paper I rejected just two weeks ago, primarily due to significant gaps in its adherence to basic elements of scientific discourse, as well as issues with clarity and readability. Having quickly scanned the revisions, it appears that the authors have made some hasty changes, but I am not confident that these updates adequately address the fundamental issues I initially raised. I remain unconvinced that the manuscript can be significantly improved without a thorough restructuring of their approach.

Answer: In the previous revision, we made substantial changes to the manuscript, particularly in explaining the methodology. Unfortunately, Reviewer A did not recognize these changes and was surprised by the timeline in which the improvements had been made. Some of the suggestions from the other reviewers relate to Reviewer A's points of concern and have been further addressed in this revision. We have substantiated the premise that current flood risk assessments overestimate the level of risk, added the rationale behind the model, and provided a more detailed description, including additional analyses. Furthermore, we improved the manuscript regarding the contribution of multiple breaches and addressed Reviewer A's concerns about the expert elicitation process.

Reviewer B:
The authors made substantial changes to the manuscript, particularly in explaining their methodology which definitely improved the manuscript. There are still some points that I feel need to be addressed though. Some are minor/textual, but some are also about the argumentation put forward and are more fundamental. I'll start with the latter:

Review feedback
In several parts of the manuscript past improvements in flood defences (after 1953, after 1995) are quoted to support the argument that current failure probabilities are too high (1/3), and their improved

estimates (1/50) are more realistic. I agree their improved estimates are more realistic, but those past improvements are not a proper argument for that as it concerns the current probabilities. The core of the reasoning is that there has been no failure since 1953 (and thus 1/3 is way too high) and I agree, but whatever reinforcements were made is irrelevant. Actually, it may undercut their argument a bit. Their estimate is 1/50 for 'a flood in the Netherlands' . However, in the past 70 years (since 1953) there has been no failure, whilst the state of the levees was considerably lower than the currently estimated 1/50 (all the years before reinforcements were in place over these 70 years). One could argue that the 1/50 is thus also too high.

Answer
The current probability of flooding in the Netherlands (or in a specific region of the Netherlands) is uncertain, as no historical dataset is available. In Section 2, we now provide a more detailed explanation of why the results are not realistic for use in spatial planning and the financial sector.
In Section 2.1, we discuss the available information, including the improvements made to levees. In Section 2.2, we examine the choices made in the current approach and their impact (Section 2.2.5). This approach is referred to as the DPV method, named after the Delta Program for Water Safety (in Dutch: *Delta Programma Veiligheid*), where it was used to assess the safety standards for all levees.
Due to the uncertainty in the current flood probability of the Netherlands (or a part of the Netherlands), we introduce a bandwidth ranging from a "fully independent system" (as in the original manuscript) to a "fully correlated system." For the flood probability of the Netherlands, we assume that river areas are correlated, as well as coastal areas, while the other areas remain independent. This results in a range of flood probabilities for the Netherlands as a whole and for specific regions.
In the fully correlated system, the flood probability of a region is not influenced by system behavior. The bandwidth analysis also demonstrates that the calculated flood probability for the Netherlands is unrealistically high. Furthermore, it highlights that all elements in Sections 2.2.1–2.2.4 matter, not just the contribution of system behavior.
Due to a lack of knowledge, no complete physical model is available for levee failure and the impact of emergency measures. However, the financial sector still demands this data and continues to use incorrect data. To prevent an overestimation of risk, we developed a new method—the BREACH-METHOD—to estimate flood risk. The method produces a flood probability for the Netherlands that is not contradicted by the available historical data. As discussed in the conclusion of this paper, the expert judgment assessment can be improved as more knowledge becomes available.

Review feedback. I am still not in agreement with how the manuscript uses the term 'system behaviour'. To me what the method does is more the inclusion of multiple-breach events, which is an important contribution, but actual system behaviour (i.e. what happens at A influences situation at B; check e.g. van Mierlo et al., 2007[1]) is actually nicely described in the concluding remarks in the 4th paragraph (as something that was not considered): "In the case of a breach, the local water level will decline locally, possibly reducing the likelihood of nearby breaches by considering local hydrographs. This would mainly impact flood risk profiles because the scenario probability of other events in the same class would increase (keeping the overall probability of flooding in the area the same) and thus have little influence on the probability of flooding in The Netherlands or an independent area.".

[1] van Mierlo, MCLM., Vrouwenvelder, ACWM., Calle, EOF., Vrijling, JK., Jonkman, SN., de Bruijn, KM., & Weerts, AH. (2007). Assessment of flood risk accounting for river system behaviour. *International Journal of River Basin Management*, *5*(2), 93-104.

Answer:
We agree on the definition of system behavior. In Sections 2.2.1 and 2.2.2, we have provided a more detailed definition of system behavior. We also referred to Mierlo (2007), as well as to the work of De Bruijn et al., who built upon Mierlo's research. In the manuscript, we further demonstrate that system behavior cannot result in a higher flood probability for the Netherlands or any specific region.
In the new Breach-method, we classify levee sections into different categories (such as T100). For each class, the combined failure probability of all possible flood events determines the failure probability of that class (1/100 minus the failure probability of more extreme classes). This approach ensures that system behavior, including hydraulic interdependencies, is accounted for in such a way that the flood probability of a region can never exceed that of its weakest link.
However, we also demonstrate that the current definition of failure, incomplete knowledge of the failure process, and the absence of emergency measures influence the results. Additionally, the duration of hydrographs is not considered, which may lead to different system behavior in the case of more extreme flooding. Therefore, we incorporated expert judgment to account for these processes.
We have provided a clearer explanation of the approach taken and the background of the method, primarily in Section 3, following a more structured introduction to the various choices in Section 2.2.

Review feedback
More importantly, in the results the authors attribute the lowering of the probability of 'a flood somewhere in the Netherlands' to the system behaviour they incorporated. Such as in the first sentence of 5.2.1: "The impact of the expert estimates on the probability of a flood somewhere in The Netherlands is greatest for the areas behind primary flood defences where hydraulic dependencies play a role." And in the last sentence of 5.2.2: "Along the coast in general the impact is less than on the other areas because the water level on sea is less impacted by system behavior than breaches along the river system". I have two issues with this:

1. Such a statement cannot be made from a methodological perspective as it is based on the difference in probabilities of LIWO (old method) and the tables estimated by the 5 experts (new method). However, in the new method the experts include in their flood probability estimates (Table 2) all three factors at the same time (system behaviour, incomplete failure probability and emergency measures). With no additional information/analysis, one cannot methodologically make a distinction as to which of those 3 factors the experts considered most in their estimates.
2. The system behaviour / multiple breach events would not affect the overall probability of 'a flood in the Netherlands' as, by definition, the overall probability of flooding in the area is kept the same when generating the multiple breach events (Table 3).

Answer to point 1:
This question is partly addressed in the response to the previous question. We combined the correction of all choices in the current risk approach into a single factor, estimated by experts. We improved the manuscript by providing a more detailed description of the Delphi process and expanding the discussion on the impact of system behavior in the new BREACH-METHOD. Additionally, we included a sensitivity analysis (Section 4.4) to examine the impact of class definitions in our model and the conditional probabilities for zero to multiple breaches.
It is well known that system behavior is particularly relevant for the (tidal) river area and the lake area. This aspect was considered by the experts involved in the Delphi method, as they had a state-of-the-art understanding of the available data and methods.

Answer to point 2:
We fully agree and have improved the manuscript to provide a clearer explanation. We also elaborate on the role of emergency measures and the methods used to define failure probabilities. Due to the choices made in these methods, the flood probability is lower than the outcomes produced by the models.

Review feedback:
My gut feeling is that consideration of the full failure mechanism (by the experts) in their new method is mostly responsible for the lowering of the probability of 'a flood somewhere in the Netherlands'. Resulting methodologically from the experts giving higher probabilities to 0 breaches as opposed to the LIWO data. However, the text of the manuscript currently reads mainly that this would be the result of system behaviour. Linked to the above, I strongly feel that something on the sensitivity of the approach should be mentioned in the manuscript. Given the low flood probabilities that are considered, I think the method is very sensitive to the estimates of 0 breaches by the experts. This is not discussed in the manuscript. Ideally, the range of the experts (if they made individual tables) can be used to illustrate the uncertainty around their estimates. Or a small sensitivity analysis can be done by slightly adjusting the 0 breach estimates and reporting on the consequent changes in the probability of 'a flood somewhere in the Netherlands' (should be possible to calculate just with the numbers of 0 breaches for the Table 2's of the different areas; no need to make new entire sets of 5 million events).

Answer:
This is an important element. We have added a sensitivity analysis to demonstrate the robustness of the model with respect to different choices (i.e., the impact on model outcomes). We have also improved the sections describing the new model and have made the model choices more explicit.
It is important to note that the impact of model choices on the flood probability of the Netherlands is significantly smaller than their impact on local exceedance frequencies of objects. For risk assessment in the financial sector and building codes, these local exceedance frequencies—as well as correlations and coincidences—are the most critical factors.

Reviewer: Some smaller items:
Review feedback: As also remarked by the other reviewer, I feel that some more information on the delphi approach with the experts is needed. For instance what the background of the experts was, but also on the process itself (i.e. were they put together to discuss and come with one set of tables, or did they make tables per individual expert and were these combined using equal weighting or something?)
Answer: We have provided a more detailed explanation of the Delphi process and the expertise of the involved stakeholders

Review feedback: Paragraph 2.2.3 is only a single sentence. Please give this some body by naming explicitly what is not taken into account, and give the reader an indication of how important those missing steps are (i.e. emergency measures that are often taken, difference in time/probability between backward erosion and enlargement and eventual collapse).
Answer: We have improved this section by providing a more detailed description. In the assessments of failure probabilities in the basic data, these measures are not considered at all. The local impact also depends on the failure mechanisms, which are critical and vary across the system.

Review feedback: Step numbers quoted in first sentence of 4.6 not in line with the steps defined in 4.2.
Answer: This has been improved.

Review feedback: Note that with the substantial re-writing I came across various typos and incorrect sentencing. I did not go in detail to mark all these things but it would be good to have someone do a final close read for these things.
Answer: We did. Sorry for the inconvenience.

Reviewer C:

JCRFR 7882

This paper aims to propose a new flood risk assessment framework and model for flooding, to be used for spatial adaptation and the financial sector, using existing flood risk information, statistical methods and expert judgment. The topic is relevant and of interest for JCRFR.

Answer: Thanks for your detailed checks! We have enhanced the manuscript to ensure it is well-written. In case of questions or remarks—except for grammatical errors—we have explained below how we improved the manuscript.

Authors should correct/look for:

- Comas after e.g. however, et al., hypothetical sentences, and other situations – and before e.g.
- Double dots ('..')
- 'this' without a noun after
- Composed words like "risk-based" needs the "-"
- 'analysis' is singular, 'analyses' is plural
- Numbers before 10 should be in words (e.g. two), numbers after 10 should be numbers (e.g. 11)
- All cross-references to tables/figures have failed
- Figures with multiple pictures should have a, b, c, etc and not 'left/right'
- There are also various spelling mistakes

The paper would also benefitted from re-structuring: the 'literature part' (up to page 8, line 20) is very long. Also, aim and objectives (or research questions) are not clearly stated. The methodology would improve if a methodological flowchart is included; a clear table of input and output data is missing too. The main remark, however, is about the core of the paper, which I understand to be the "new model". This model is said to be more realistic, although a validation is not offered in the paper.

Answer: Reviewer B had a similar question, which we answered above. ***In brief,*** *we have enhanced the paper by describing a bandwidth for the current flood probability and relating this to the available data.*

I would also suggest to give a clear name to both the 'old' model (e.g. with a reference, Smith's model or the NL national model) and the 'new' model (maybe an acronym – this would be helpful also for people using and citing it). Sec. 4.2 should have at least 'new' in the title, or indeed the name of this model.
Answer. We introduced names for the different models. The current approach is called the DPV method, named after the Delta Program for Water Safety (Delta Programma Veiligheid in Dutch), where it was used to assess the safety standards for all levees. The new approach is called the Breach Method.

The paper also stated that the method/model is 'tailored to delta areas with protection infrastructure and hydraulic interdependencies', but there is no evidence of this in the paper.
Answer. The same challenges apply to all delta areas. System behavior and uncertainty in failure probability arise due to a lack of knowledge about failure processes and emergency measures.

Finally, it is not explained well (just a paragraph) how the expert judgment is included, in such details to be reproduced by other researchers.
Answer. We have described the Delphi process in more detail, including the selection of experts.

More suggestions are offered below (P is page, L is line).
P1L5: "A" new framework to generate event "sets"
P1L11: 'available' instead of ' developed'
P1L14: 'In this assessment': the authors' paper of reference's?
P2L18: … and increased exposure, e.g. rising population
P2L29: remove ')' after full stop
P2L29: 'People…houses' sounds speculative
P2L34: it's never realistic
P3L10: not sure what 'open windows' are
P3L18-29: summarise, not all info are needed
P3L32: explain what the Room for the River is (with source), or remove the reference
P5L6: instead of 'below', specify the section no.
P5L25: delete 'conform', not needed
P5L26-27: ref needed
Answer. This is an example of our own

P6L13-18: out of place here, maybe in discussion.
Answer. Section 2.2.3 discusses one of the three main assumptions made in the current DPV method. Reviewer B requested a more detailed explanation, which we have provided.

P6L19: Sec. 2.2.2 is very short it could be combined.
Answer.
This section also addresses one of the main assumptions. Do you mean Section 2.2.3? Reviewer B asked us to elaborate on this, and we have done so.

P7L32: remove sub-title (equal to title), plus there is not Sec. 3.2
P7L36: the subject of the sentence is missing
P7L38-39: the main verb is missing
P9L5: not sure why data has ' '
P9L10-14: include for each bullet point the section of reference (e.g. Sec. 4.3 for the first)
P9L39-43: all terms like RCP8.5, W+, SSp5-8.5 need to be explained. Be inclusive of all readers.
P10L25: 'which' without capitals
P10L33: what 'finer' means? What is the resolution needed in spatial planning?
P10L34: 'the' instead of 'de'
P11L5-4: ref needed
P12L11: sentence not finished
P13L7: what is the 'new flood event set'? how is new?
P13L14: spelling of 'Hower'
P13L18: what is meant for 'impact for continuity analyses'?
P13L20: what is exactly the novelty of the 'new flood risk model'?
P14L7: spelling of 'cas'
P14L12: Room for the River with capitals
P14L32: the term 'impact' is not clear, meaning the difference?
Figure 5: the legend has not object
P15L23-28: the bullet list is not needed
P16L9: indicate figures number; again 'impact' is not a clear term

Figure 6: the legend should be smaller, same font of axis for example
Answer. We did not change this due to visibility. If needed, we can update it during the publication process.

P17L19-20: has the 'new' model being validated? How can authors say it is realistic?
*Answer. We have improved this discussion. The results cannot be rejected based on the available data, rather than on the current risk approach.*P17L13-27: revise the term 'impact'
P18L3: 'of' before 'other structures'

Round 3

Dear Reviewers

Again, thanks for your time and effort to review our manuscript. Based on the valuable feedback, and discussion with the editor, we fully restructured the manuscript and worked on the improvements. In this review letter we inform you point by point how the review has resulted in an improvement of the manuscript.

Our point-to-point reply is in the 'red' text.

Kind regards

We have reached a decision regarding your submission to Journal of Coastal and Riverine Flood Risk , "New framework to generate event set for risk based spatial planning and financial decisions for Dutch Flood".

Our decision is to:
Please revise your paper considering the few comments. Additionally, a list that includes each reviewer's remark and your reply indicating how you have considered the comment is needed in an extra file. It should be given per reviewer, and in the same order as the remarks were given. Additionally, please provide a second copy of the paper where the changes are clearly indicated (e.g. 'track changes' in Word).

Reviewer B:
My main concern is that a part of my previous feedback has been ignored (see for example "The paper would also benefitted from re-structuring: the 'literature part' (up to page 8, line 20) is very long. Also, aim and objectives (or research questions) are not clearly stated.

The methodology would improve if a methodological flowchart is included; a clear table of input and output data is missing too.").

Based on the review comments, we have added a flowchart to the manuscript. This flowchart illustrates the steps of the BREACH method, including its inputs and intermediate results at each stage. A full explanation of the method is provided in the appendix. In addition to the textual description in the main method section, we believe the flowchart offers a clear and accessible overview of the approach, as requested by the reviewers.

The structure still needs improvement. The paper is 20 pages (excluding reference) and it takes 9 pages to arrive to something new, i.e. the "BREACH method" explanation. Precedent sections need a 50% cut; for example, I am not sure that Sec. 2.3 is relevant for the paper in such a long form. See previous points above aim and objectives. After the results, I would add a clear "Discussion" section, which should also include future works. "Concluding remarks" is a very long section: usually "Conclusion" is about key points (similar to the abstract), with no citations, without reporting any new information.

About the structure of the manuscript

We also discussed the structure of the paper with the editor and we have decided to restructure the manuscript to improve clarity and focus. The revised structure follows a conventional scientific layout: introduction (including a clear aim and research questions), literature review, new method, results, discussion, conclusion, and recommendations. We also shortened each section to enhance readability. Detailed explanations have been moved to the appendix.

- The introduction is now limited to two pages and explicitly outlines the research goals.
- The review of the current DPV method and data is also two pages; additional background information has been moved to the appendix. This section now also includes the description of the available data, as all methods are based on the same dataset.
- The method section describing the new model has been reduced to two pages.
- The results section spans four pages. The discussion and conclusion together cover two pages.

The main body of the manuscript is now considerably shorter than in the previous version. Some of the background information has been placed in the appendices, in response to reviewer comments that the introduction was too extensive but necessary to understand the context.

We now clearly state the aim, objectives, and research questions of our study. The final paragraph of Section 1 now reads as follows:

The central question in this study is whether the available flood risk data and existing risk assessment methods can be applied within the financial sector. The objective is to evaluate the currently available flood risk data for the Netherlands, clarify the underlying assumptions, and assess the implications of applying the DPV method in a financial context. To this end, we introduce a new method 'BREACH' which generates a large flood event set that explicitly incorporates the assumptions embedded in the underlying flood risk data. This method enables a more realistic and transparent risk assessment tailored to the needs of the financial sector, as it also captures the tail of the distribution.

The paper still contains mistakes in terms of punctuation and "style" (e.g. "this" without a noun, already said in previous review). I am not going to point them out, authors should proofread the manuscript, or use a dedicated service.

Style, Spelling and Grammar

The manuscript is fully checked on this. We added a version with all the methodological and structural changes in the manuscript, and a version with all the grammar changes.

Minor things

General: uniform the capital (or non-capital) letters for "Section/section", as well as "The/the" for the Netherlands

Title: "risk-based"

P1, L15: the DPV acronym is not explained

It is an abbreviation which we mention in the abstract first. In the main text it is first introduced on page 3, where we clarify why the letters DPV are used.

P3, L6: delete the space after the bracket

We have not found the space after the bracket on Page 3 Line 6 of the previous version, and we have checked spaces after the brackets in the new version.

Fig. 2: there is not legend for black dots, red lines and blue areas

This figure has moved to an appendix, the legend is explained the subscript of the figure.

P6L15: again DPV is not explained

See above.

Fig.3: spelling mistake in "Simpel"

We have changed this typo.

P7, L4: instead of "above example", give the figure number

We now refer to the figure number.

Fig. 5, caption: if direct citation, the correct style is "Name (year). So "(Pol et l. 2024)" need corrections

I think the remark is about Figure 4 and we have changed the reference to Pol et al. (2024).

Fig. 7: a/b should be smaller and non-capital (please, look how other papers do). Edit the caption too

We have changed the plots and the captions of this figure.

Fig. 8: there is an "o" close the legend of top right figure

We have removed the 'o'.

P9, L1: the number in brackets in the sub-sections of Sec. is not needed

We are not sure what the referee means by the number in brackets (step 1, 2, 3?). We think it is important to order the steps, so we have numbered them from 1 to 6 (and deleted the brackets).

Reviewer C:

Recommendation: Reject

The manuscript introduces the BREACH method, a new approach for flood risk assessment in the Netherlands, aiming to show that current flood risk estimates are overly conservative. While the authors have made notable efforts to incorporate feedback from previous review rounds, in its current form the study remains underdeveloped. Significant methodological and conceptual weaknesses persist, particularly in how expert judgment is used and justified.

Thank you for recognizing the improvements made. We have enhanced the description of the expert judgment process and explicitly acknowledged that expert choices are already embedded in the underlying data. A detailed account of the expert judgment procedure has been added, including the formulation of questions, expert selection, multiple rounds of consultation, and the final findings.
We also discuss potential avenues for improvement. While we fully agree that the process can be further refined—as we explicitly recommend—we do not agree with the assertion that the process is incorrect.

Justification of the method: conceptual and analytical gaps

The need for this new method is insufficiently demonstrated. Although the paper argues that current flood risk assessments significantly overestimate risk—leading to property devaluation and higher insurance costs—this claim is not supported by solid mathematical or empirical evidence. The evidence is presented in the paper. Applying the DPV method at a national scale results in flood probabilities for the Netherlands that are not supported by empirical data, nor by the extensive flood protection measures implemented over recent decades. Consequently, the local failure probabilities of individual properties are, by definition, overestimated.
In contrast, the new BREACH method yields lower flood probabilities, which are more consistent with historical events and the numerous improvements made to the flood defense system. As discussed, further empirical evidence is currently not available. Given the substantial influence of the method on estimated flood probabilities—and the implications for solvency and insurance assessments—it is evident that the DPV method is not suitable for this purpose, as demonstrated in the paper. Specific concerns include:

- **Section 2.2.1**: The discussion conflates annual flood probabilities with conditional flood probabilities under specific scenarios. While levee breaches and resulting flood depths may be dependent within a given event, flood event scenarios themselves can be constructed to be independent. This undermines the necessity of the method's treatment of system behavior. This section focuses on the probability of flood scenarios arising from system behavior, rather than on the flood scenarios themselves (which are described in Section 2.2.2). Flood events can indeed be defined independently; the underlying dataset already includes multiple flood scenarios per breach location, with return periods of water levels often varying by a factor of 10 to 100, and covering different breach locations.
  The existing flood scenario database developed over more than 10 years does not account for system behavior (as previously discussed in former Section 2.2.4). However, to perform a risk analysis, the probability of these scenarios must also be known, which is the focus of Section 2.2.1. The improved and restructured manuscript clearly outlines the challenges involved in defining these probabilities.
- **Section 2.2.2**: The main point is not clearly articulated, leaving the reader confused about what the authors are trying to demonstrate. We improved the description of the foundational choices in the manuscript, and restructured the manuscript.

- **Section 2.2.3**: Claims about the limitations of current methods lack substantiation through uncertainty quantification. A proper reliability analysis could provide the missing importance coefficients. Also, emerging threats to levee integrity—such as damage from protected species like beavers and badgers—are not addressed, despite their growing relevance. We used the officially reported flood probabilities provided by the regional water boards and Rijkswaterstaat. These include all failure mechanisms and explicitly address issues such as those caused by beavers and badgers. However, applying these probabilities within the DPV method leads to a flood probability for the Netherlands that is inconsistent with historical data and observed performance. The knowledge gaps in defining accurate failure probabilities are acknowledged in the manuscript.
  In principle, a proper reliability analysis could be used to derive the required importance coefficients—but the challenge lies in defining what constitutes a 'proper' analysis. Moreover, the relationship between factors such as beaver activity and failure probability is not yet well understood, nor is it implemented in current levee assessment frameworks. The same applies to various other factors, including time dependence.
  While we reference literature that could help close some of these knowledge gaps, this work has not yet been carried out. The next national assessment of levees is scheduled for more than 12 years from now, whereas financial institutions require reliable risk estimates to support their current decision-making.
- **Historical data misinterpretation**: The use of a short historical record (<100 years) as evidence for overestimation of flood probabilities reveals a lack of understanding of extreme event dynamics, particularly regarding non-linearity and fat-tailed distributions. These are central to flood insurance and risk modeling, and their absence is a serious flaw. The model has been developed specifically to describe the tail of the distribution. The limited number of historical flood events is a fact; as shown in the paper, flooding due to breaches of primary flood defenses in the Netherlands has been rare. However, applying the DPV method results in a flood probability for the Netherlands that appears extremely high when compared to the historical record and the extensive flood protection measures in place. This discrepancy is thoroughly discussed in the paper.

Methodological weaknesses in the use of expert judgment

The methodology is heavily reliant on expert judgment, yet the design and execution of the expert elicitation process raise significant concerns:

We acknowledge that expert judgment is involved, but also all the available flood risk data; however, the DPV method also relies on expert judgment, although its use is often less explicitly stated. The inclusion of expert choices in risk analysis is common practice. In our manuscript, we aim to be as transparent as possible about the expert judgments used, to enable future improvements and refinements.

- The paper does not convincingly argue why expert judgment is preferable to advanced modeling, especially in a context like the Netherlands where rich quantitative data and sophisticated models are available (see also points raises in the first part of this review). We recommend improving both advanced modelling techniques and the use of expert judgment. The foundational data used in the DPV and BREACH methods—already incorporating expert choices—are the result of decades of research. Further improvement of these data through advanced research will take many years, whereas risk analyses for financial stakeholders are needed now. Therefore, we advocate for continued development of advanced modelling to eventually reduce reliance on expert judgment. However, expert judgment and expert choices will continue to play a critical role. We therefore also recommend efforts to further improve the quality and transparency of expert judgment.

- The reported number of experts is inconsistent (3 in the appendix, 5 in the main text). Either way, the panel is too small for robust elicitation. Additionally, all experts appear to come from similar backgrounds (flood risk modeling), lacking disciplinary and institutional diversity (e.g., emergency response, policy, insurance).
  The correct number is 5. All the disciplines were on the table, all experts were involved in flood risk analyses for these fields.
- The paper does not disclose expert credentials, affiliations, or potential conflicts of interest. The selection criteria are also unclear, raising questions about the rigor of the process.
  The selection is based on the expertise as mentioned. There is no conflict of interest, none of the experts is responsible for levees or the consequences.
- The framing of questions is not clearly linked to specific knowledge gaps that cannot be resolved through modeling. Only two Delphi rounds were conducted—insufficient for developing stable consensus or adequately testing assumptions. We do not agree with this comment. The approach followed is consistent with the standard Delphi method. The research gap is clearly defined. Although some reviewers suggest providing expert judgment for each individual foundation type, we have combined them into a single conditional probability. Exploring conditional probabilities per foundation type is indeed a valuable direction for future research.
- The connection between expert inputs and the outputs of LIWO (the national flood information system) is vague. There's no clear method showing how expert judgment complements, adjusts, or replaces model-based outputs.
  We added a flowchart to illustrate this better.
- Emergency measures and expert gaps: While the time required to implement emergency measures is rightly flagged as crucial, none of the experts involved in the Delphi process appear to represent emergency management or crisis response backgrounds. This omits a critical perspective and limits the credibility of the derived judgments. This is not true, one of the experts is a top expert in crisis management.

1. **Overstated relevance and conclusions**

The authors claim the BREACH method has strong relevance for the financial and insurance sectors. This is overstated and premature:

- The authors claim that the results are relevant for use in the financial sector, which is an overstatement. Extensive additional research is needed to bring the breach method to the level of sophistication of risk assessments in the financial sector. Notably, the first key weakness indicated of the current flood risk assessment methodology, system behaviour, was not addressed in this study, and evidence that non-linearity effects are sufficiently accounted for in the estimations needs to be provided. The method is already used in the financial sector, including a validation by the financial sector.
- In the concluding remarks, but also in section 2.2.5 the authors imply that the increase in hydraulic load and frequency due to climate change is less significant than the level of flood risk overestimation that is currently the case due to methodological weaknesses in the current risk assessment approach. This is a questionable claim. That is correct. As shown in the paper, the impact of the method is, in many cases, more than a factor of 5, whereas the impact on the consequences is only a factor of 0.35 for events with a hydraulic load ten times more extreme. For the former there is plenty of evidence that emergency situations will arise more often than in the past. For the latter, there isn't sufficient proof provided in this manuscript (see first part of this review). This point needs further clarification and delineation. Yes, climate change will lead to sea level rise and an in-

crease in extreme river discharges. However, Dutch levees will also be reinforced to reduce their probability of failure. (The Dutch Ministry of Infrastructure states: The protection standards are guiding for the reinforcement of levees, not climate change. But the design of levee reinforcements takes into account climate change, as levees are designed for a lifetime of 50–100 years and they will be reinforced afterwards if needed.) It is important to note that the current failure probability of many levees exceeds the safety standard; therefore, extensive reinforcements are planned up to 2050, which will significantly reduce failure probabilities. The revised manuscript provides a clearer and more structured description of the impact of these reinforcements.

- The authors also provide a number of recommendations without providing enough arguments of why these are needed: e.g. they recommend moving to a structured expert judgement where non-uniform weighting among experts is applied. This can introduce further bias in a study where the expert elicitation process hasn't followed the proper steps. The authors are advised to get their basic Delphi first right before moving to a more complex process. More examples like this can be found through the concluding remarks section.
  We do not intend to discourage adherence to established procedures. Earlier, we discussed the Delphi process and methods to enhance expert judgment and advanced modeling.

1. **Readability and structure**

The paper's structure and presentation hinder its effectiveness, even for readers with relevant expertise:

- Key concepts are introduced without prior explanation. For instance, the notion of "classes" in Section 2.2.5 is not previously defined, leaving the reader disoriented. In the same section Table 1 lacks clarity regarding its data source. If the figures are drawn from LIWO, this should be clearly stated. Similar examples can be found throughout the manuscript.
- Overall, the structure and sequence of sections feel disjointed. The paper does not guide the reader through the logic of the argument in a coherent way. Some sections delve into detail prematurely while others skip essential context. This affects the paper's accessibility and makes it difficult to assess the core contribution.

We asked the editor for advise on the readability and structure. We then fully restructured the paper. See also the above answers.

Round 4

Dear Reviewers

Again, thanks for your time and effort to review our manuscript and accept it for publication. We improved the last minor issues as pointed out by reviewer A. We replaced 'this' for the mentioned locations in the manuscript (and some more) by a proper description. We also improved the other remarks.

We uploaded a version with track changes so all changes can be seen. We also uploaded a 'clean' version in which all the track changes are accepted. The clean version can be seen as the final version.

Kind regards

The authors

Reviewer A:
I congratulate the authors for the substantial improvement.
Final suggestions, mainly about style and form:
P5, L18: delete the internal parenthesis of the year of the reference Ten Brinke et al. 2010
P5, L19 and 21: 'This' what? This limitation?
P5, L29 and 33: 'This' what? (this without a noun is always unclear, and also a symptom of careless wording)
P6, L38: 'This' what? à could be changed into 'This assumption is conservative'
P9, L1: 'This' what?
P10L2: it would read better as 'Figure 2 shows'
P11, L9: 'This' what?
P11, L13: improve the redundant expression 'This information can inform'
P12, L29: 'This' what?
P13, L22: 'These' à 'These consequences'
Recommendation: Accept Submission

Reviewer B:
Recommendation: Accept Submission
