

Author Response to Referee #1 (RC1)

We thank Anonymous Referee #1 for their thorough and constructive review. We are grateful for the positive assessment of our study's originality and significance, and we appreciate the detailed feedback on how to improve the manuscript. The reviewer raised two principal concerns: (1) a conceptual issue regarding our treatment of robustness vs. adaptiveness (and potential confusion with the concept of regret), and (2) extensive language and clarity problems in the writing. We address each point in detail below, with planned revisions indicated (including specific sections and line numbers from the original manuscript where changes be made). We have also carefully noted all the detailed line-by-line comments in the annotated PDF and incorporate those corrections. Once again, we thank the reviewer for these invaluable comments, which help us significantly improve the paper.

Comment 1 (Conceptual): Distinction between Robustness and Adaptiveness, and confusion between Robustness and Regret

RC1: *"The authors characterize their study as not solely relying upon robustness criteria... but also engaging adaptivity considerations as well. This is curious... Adaptation and flexibility have long been understood as being some of the principal tools for ensuring the robustness of a planned course of action... Yet, in this paper they are presented as alternative approaches... There may be several misunderstandings. One is that 'robustness' is a much abused term... Another possibility is that the authors are confusing the concept of robustness with that of regret in terms of decision analysis. They discuss at one point different approaches to measuring 'robustness' when in fact they are citing different methods for calculating regret... They introduce alternative metrics such as adaptability and do so convincingly. The problem is that they do so in a false opposition to a straw man that they term 'robustness'."*

Response: Thanks for the suggestion and we apologize for the confusion caused by our framing of **"robustness"** and **"adaptiveness."** We agree with the reviewer that in the context of decision-making under deep uncertainty, **adaptiveness (flexibility)** is not a separate or opposing concept to **robustness**; rather, it is often an integral means of achieving robustness over time. Our intention was not to set up a false dichotomy. Instead, we aimed to highlight that focusing **solely** on near-term robustness (in terms of performance across scenarios) can overlook the long-term benefits of adaptability. We realize that this intent was not clearly communicated. In the revised manuscript, we clarify our conceptual framework to eliminate any suggestion that robustness and adaptiveness are mutually exclusive approaches. Specifically, we make the following clarifications and changes:

- In the Introduction and framework description (see Section 1, lines 83–89 and 108–112; and Section 2.1, around lines 173–180 of the original manuscript), we define clearly of

“robustness” and “adaptiveness.” For example, we state that robustness in our context refers to *the* we define ‘robustness’ as the ability of a strategy to maintain acceptable performance across a wide range of plausible future scenarios, and ‘adaptiveness’ (flexibility) as the ability to adjust or augment the strategy over time in response to how the future unfolds. We emphasize that flexibility and adaptation are key tools to achieve robustness under deep uncertainty, rather than a separate objective. This addition ensures readers understand that our use of “robust” and “adaptive” are complementary aspects of a strategy’s performance.

- We reframe the text wherever it implied that robustness and adaptiveness were alternative or opposing approaches. In particular, the sentences in the Introduction (lines 105–112) that currently discuss a trade-off (suggesting that maintaining a high level of robustness may compromise adaptiveness) is revised. We clarify that while there can be a tension in practice (e.g., an initially over-engineered solution might reduce future flexibility), **our framework is designed to balance both** – ensuring a strategy is robust **and** remains adaptable over time. The revised narrative stress that **our contribution lies in integrating adaptiveness into robust decision-making**, not treating it as a separate paradigm. For instance, we explain that our framework builds on Robust Decision Making (RDM) principles combined with Dynamic Adaptive Policy Pathways (DAPP) concepts, so that short-term performance and long-term flexibility are both evaluated. This addresses the reviewer’s point that adaptation and flexibility should be seen as part of robustness: we make it clear that our approach explicitly combines them, rather than introducing adaptiveness as an opposed idea.
- Clarifying Use of “Robustness” vs. “Regret”: We acknowledge the reviewer’s observation that our manuscript in places may have conflated robustness with regret-based metrics. In the original text (Section 2.1; see lines 225–233), we discussed “different approaches to measuring robustness” and cited methods that are essentially different ways of calculating regret (e.g., using satisficing criteria or minimax regret as performance measures). We agree that regret is a distinct concept: it measures the difference between the outcome of a strategy and the best possible outcome in each scenario (i.e. the opportunity loss), and is often used in decision analysis to evaluate robustness, but regret itself is not synonymous with robustness. To avoid confusion, we correct our terminology in those sections:
 - We revise statements like “*robustness metrics (i.e. satisficing and regret)*” to more precise wording. For example, we replace that phrase with “*decision-making performance metrics (such as satisficing criteria or regret-based criteria)*”. This makes it clear that regret is one way to evaluate outcomes under uncertainty, rather than a definition of robustness itself.

- Where we wrote “*different approaches to measuring robustness*” while referring to regret calculations (Herman et al., 2015; McPhail et al., 2018), we rephrase to “*different approaches to quantifying a strategy’s performance under uncertainty*”. This clarification ensures readers understand we are using regret as a tool within robustness analysis, not confusing the two concepts. These changes are made in Section 2.1 (Framework development) around lines 223–235, where we introduce how robustness is assessed in our study.
- Additionally, we ensure our use of the term “**robustness**” throughout the manuscript aligns with how it is used in policy and decision science literature (i.e., focusing on strategies that perform satisfactorily across uncertainties). The reviewer is correct that the term can carry different meanings in other fields (statistics, engineering, etc.), so we add a brief note in the Introduction (near line 85) acknowledging this and stating that in the context of DMDU (Decision Making under Deep Uncertainty), robustness is about insensitivity to uncertainty rather than, say, over-designing to a safety factor.
- **Addressing the “expense of robust solutions”:** The reviewer noted our reference to the “considerable expense of robust solutions” and pointed out that a purely engineering notion of robustness (overbuilding to withstand worst-case forces) might have influenced our wording. We clarify this to avoid misunderstanding. In the **Discussion**, we explain what we meant: a strategy that is **robust** against extremely adverse scenarios (e.g., constructing the highest possible flood defense immediately) can indeed require high up-front cost or lead to over-design if the worst-case never occurs. In policy terms, a **robust** strategy need not always be the most expensive – especially if it is combined with adaptive staging. We adjust the text (for example, in Section 4 Discussion, around lines 590–599) to make it clear that **our approach seeks robust outcomes in a cost-effective way by using adaptiveness**. In other words, rather than building an immensely over-sized infrastructure now (high-cost “robust” solution), it may be better to start with a moderate solution and then adapt (expand or add measures) if needed – achieving robustness through flexibility. By revising this explanation, we reinforce that **robustness and adaptiveness work in tandem**: adaptiveness can reduce the cost of achieving robustness by allowing adjustments over time instead of one costly irreversible investment. This directly addresses the reviewer’s point and removes the implication that we view “robust = expensive” as a general rule.

In summary, we substantially revise the introduction and conceptual framework sections to clearly convey that our “robust adaptive pathways” approach is a synthesis where robustness is pursued in conjunction with adaptiveness (ability to adjust strategies over time). All mentions

of these concepts be reviewed to ensure consistency. We eliminate any language that sets up robustness versus adaptiveness as a straw man or false opposition. Instead, the revised manuscript consistently present adaptiveness as an essential component of robust long-term planning under deep uncertainty. We believe these clarifications address the reviewer's concerns and help readers better understand our framework. Thank you for highlighting this important conceptual issue – our revision makes the manuscript's framing much clearer.

Comment 2 (Stylistic): Grammar, Syntax, and Clarity of Writing

RC1: *“The second problem is one of syntax, grammar, and English usage... There are syntactical problems with the use of articles, verb agreement, incomplete and awkward sentences and other issues that it simply was not within the resources of this reviewer to correct... Once the manuscript has been revised in light of my comments..., it would be wise for the authors to engage the services of a native English speaker or copy editor to provide the light edit that would immensely improve the readability of this draft. Please find comments of a more detailed character in the attached PDF.”*

Response: We sincerely apologize for the language issues in the manuscript. We appreciate the reviewer's acknowledgment of our work's value despite the readability problems, and we fully agree that the paper must be polished to meet the journal's standards. We take this critique very seriously. In response, we thoroughly revise the manuscript's language for clarity, grammar, and style. Our plan for addressing the stylistic concerns is as follows:

- **Incorporating PDF Comments:** We have reviewed all the detailed line-by-line comments and corrections provided in the annotated PDF. These comments pointed out numerous specific issues (incorrect articles, verb tense/agreement errors, awkward phrasings, run-on or fragment sentences, etc.). We implement each of these corrections in the revised manuscript. For example, instances of missing articles (“the”, “a/an”) be fixed (e.g., changing “in urban area” to “in urban areas” or “the urban area” as appropriate), and grammar mistakes such as improper verb forms be corrected (for instance, where the original text might say “the method provide” we change it to “the method provides”). We also restructure any incomplete or awkward sentences noted by the reviewer to ensure they are clear and grammatically complete. Every single suggestion from the PDF annotations be addressed to improve readability on a micro-level.
- **Overall English Editing:** In addition to the specific fixes above, we conduct a comprehensive edit of the entire manuscript for English usage. This includes checking for consistent tense use, clarity of pronouns and references, and smooth flow of ideas. We simplify or break down overly long sentences for clarity. We have identified certain

sentences that were particularly clumsy or hard to follow, and we rewrite them. For example, a sentence like “*Options with high performance often mean higher cost input and potentially longer construction periods... The single-objective in either performance assessment... provided limited information for long-term planning, indicating a potential for lock-in...*” be reviewed and rewritten for clarity and conciseness so the ideas are easier to parse. We ensure that each paragraph conveys its point clearly before moving to the next.

- **Professional Proofreading:** As recommended by the reviewer, we seek the assistance of a native English speaker (a colleague) and/or a professional copy-editing service once we have incorporated all content revisions. One of our co-authors is indeed based at an English-speaking institution, and we leverage that resource; additionally, we are prepared to use an external proofreading service for an extra layer of quality control. This step help catch any lingering subtle issues in syntax or diction that we might miss. The goal is to bring the manuscript to a polished, publication-ready standard of English. We fully agree that this “immensely improve the readability” of the paper.
- **Acknowledging Specific Patterns:** The reviewer specifically mentioned issues with *articles*, *verb agreement*, and *incomplete/awkward sentences*. We have noted patterns such as omission of “the” in several places, inconsistent singular/plural usage, and some sentence fragments. We pay special attention to these patterns. For instance:
 - All technical terms or specific references be checked for proper article usage (e.g., “the RDM approach” instead of just “RDM approach” where appropriate for clarity).
 - We ensure verbs agree with their subjects (e.g., “approaches have” instead of “approaches has” if such errors exist).
 - Awkward constructions be rephrased. If a sentence was identified as incomplete or hard to read, we either connect it with the preceding/following sentence or add the necessary words to complete the thought. In some cases, splitting a long sentence into two clearer ones be the solution.
- **Consistency and Terminology:** As part of the language editing, we also improve consistency in terminology (some of which overlaps with the conceptual clarifications in Comment 1). For example, we use terms like “strategy” vs. “option” consistently, and ensure that terms like “robustness” and “adaptiveness” are used in their clarified sense throughout. This prevents any further confusion and enhance clarity.

After these revisions, we double-check the entire manuscript to make sure no new errors were

introduced and that the text reads smoothly for an international audience. We are confident that with these extensive language corrections, the manuscript's readability greatly improves.

In conclusion, we thank the reviewer again for pointing out the numerous language issues and providing detailed suggestions. We understand that the current draft's readability was not up to the mark, and we are committed to delivering a much cleaner revised version. The manuscript undergoes native-level English proofreading before resubmission, and all identified stylistic problems be corrected. We believe this address the reviewer's concerns fully and make the paper more accessible to the HESS readership.

Once again, we thank Referee #1 for their valuable insights and careful review. The feedback on both the conceptual framing and the writing has been extremely helpful. We implement all the above changes in the revised manuscript. We are confident that these revisions resolve the raised concerns – clarifying our methodology's positioning (robust vs adaptive strategies under deep uncertainty) and improving the overall clarity and quality of the text. We appreciate the reviewer's positive comments about our study's merit, and we are committed to improving the manuscript accordingly. Thank you for helping us strengthen this work.

Author Response to Reviewer Comments (hess-2024-391)

Reviewer #2 (RC2)

RC2-1: *"Lines 56–89; 114–118: In the introduction, a description of the mentioned DMDU methods (e.g., RDM, DAPP, ROA, MORDM, Adaptive Policymaking, Adaptation pathway, and the concept of tipping point) would be beneficial. In this way, the reader will be able to compare these different methods, see their limitations, and better understand how the proposed methodology tries to address them. Probably, a scheme outlining the most popular approaches (at least RDM and DAPP; L78) would facilitate the comparison. At this stage, the authors talk about these existing methods assuming that the reader is already familiar with the topic."*

Response: We agree with the reviewer. In the revised manuscript, we have **added concise descriptions of RDM** and DAPP to discussed their differences and limitations. This allows readers unfamiliar with these approaches to understand the context and how our work builds upon or diverges from them.

In line 78-80 we add: "RDM (Robust Decision Making) is effective at identifying strategies that perform well across a wide range of future scenarios, focusing on vulnerability analysis and stress testing, but it lacks explicit guidance on sequencing and path dependency; whereas DAPP (Dynamic Adaptive Policy Pathways) excels at planning flexible adaptation pathways to avoid lock-in, but is relatively weaker in quantitatively evaluating robustness across uncertainties (Haasnoot et al., 2013)".

We then explicitly state how our proposed framework relates to these: in line 86, we highlight that our work synthesizes elements of RDM and DAPP to address their noted limitations (as per Line 96, we mention that integration of RDM & DAPP is promising but has been rarely implemented, which our study attempts to do).

RC2-2: *"The way flexibility is defined in this work—i.e., as the number of combined flood-control measures gradually implemented in a given flood-control plan (L453-456; Fig. 4; L276-286) —provides a misleading assessment of the real adaptability of a given flood control implementation. This is because implementations based on a single flood control measure—hence, with the lowest flexibility rank, according to this adopted metric—do not necessarily preclude future, not-originally-planned updates by adding other control measures at later stages (e.g., if the original intervention was drainage improvement, by later adding green areas and/or the tunnel, if we consider candidate options from the specific case study; L455), as worse conditions are predicted for the years to come. Therefore, the actual difference between 1) solutions considering only one flood control strategy in the short term and 2) adaptive solutions where different strategies are all planned ahead and gradually implemented is just in the time of planning, not in their inherent flexibility. It might be interesting to study the effects of making plans for upgrades “on the fly” (i.e., when only one flood control approach is planned in the*

short run and additions are planned at later stages) vs. planning ahead all the different, gradual implementations. However, measuring flexibility using the proposed approach does not even help studying this. There are two major consequences related to this way of measuring flexibility, as detailed in the following two comments #3 and #4. To address these issues, the authors should either reframe the purpose of the manuscript, avoiding considering flexibility and centering the adaptability analysis around flexible pathways only (Fig. 5 and 6), or else re-run their analysis considering a more appropriate flexibility metric."

Response: We appreciate this detailed critique of our flexibility (adaptiveness) metric. The reviewer correctly observes that our original definition – “*flexibility = number of measures in the plan*” – can be misleading, because a strategy with only one initial measure can still be adapted later (even if not pre-planned). We have taken a two-pronged approach to address this:

(1) We have reframed the purpose and presentation of our adaptiveness analysis to focus on pre-planned adaptive pathways. In the revised manuscript, we explicitly acknowledge the limitation of our flexibility metric. We state that our metric captures *planned* adaptiveness (i.e., how many measures are included from the outset in a strategy), and we clarify that this does **not** imply that a single-measure strategy cannot be adapted in the future – only that such adaptations were not planned within our framework. To make this clear:

- We removed any suggestion that we are comparing “adaptive” vs “non-adaptive” strategies in absolute terms. Instead, we explain that all strategies could be made adaptive, but our framework distinguishes those that were explicitly planned as multi-stage pathways versus those that were not.
- We added text in Section 2.2 to emphasize this point: “It is important to note that a strategy initially implemented as a single measure (e.g., only drainage improvement) does not preclude future augmentation (such as adding green areas or a tunnel later) if conditions worsen. In our framework, however, such augmentations were not pre-planned in single-measure scenarios. Therefore, our ‘flexibility’ metric should be interpreted as the degree of planned adaptability, rather than an absolute limit on a strategy’s potential to adapt.”
- Following the reviewer’s advice, we center the adaptiveness discussion around the multi-measure pathways (the inherently flexible pathways). Figures 6 and 7 (which was figure 5 and 6) are now discussed with the understanding that they pertain to strategies that had flexibility built-in from the start. We explicitly avoid over-generalizing those results to claim something about one-step strategies beyond our experimental setup.

(2) We partially accept the suggestion to consider a different metric. Due to time and data constraints, we did not re-run a completely new analysis with an alternative flexibility metric (such as dynamic on-the-fly adaptation modeling). However, we have substantially improved our explanation of results to remove biases introduced by the metric. Specifically, we **no longer**

portray the strategy with the highest “flexibility rank” as unequivocally the best (see responses to comments #3 and #4 below). We also added a brief discussion in the Conclusion about how future work could explore truly dynamic adaptation (“on-the-fly” upgrades versus planned pathways) to provide a more direct assessment of flexibility in the real-world sense.

To summarize, we **accept the core of the reviewer’s critique** (our flexibility metric is limited) and have adjusted the manuscript accordingly by reframing our analysis and acknowledging this limitation. We did not execute a new metric analysis for this revision (hence partially accepted), but we believe the changes sufficiently mitigate the issue:

- The manuscript’s objective is now framed around robust adaptive pathway planning (as opposed to claiming a general robustness-vs-flexibility optimization).
- We explicitly caution that flexibility results are conditional on our metric and experimental setup.
- We suggest that evaluating unplanned adaptations is an interesting avenue (and indeed, outside the current paper’s scope, aligning with the reviewer’s insight).

These revisions should prevent readers from drawing the wrong conclusion that “single-measure plans can’t be adaptive,” and instead understand that our study compares different planning approaches (pre-planned vs. not pre-planned adaptiveness). All corresponding sections (Introduction, Methodology, Results, Conclusion) have been updated to reflect this clarified focus.

RC2-3: *"In the light of my previous comment #2, comparing solutions with different numbers of combined flood control measures is misleading (e.g., Fig. 4), because strategies with more combined options (e.g., D+G+Tun30) will be systematically ranked as more flexible, although single-option solutions (such as GA and Dr) can also ensure high flexibility, in principle."*

Response: We agree and have implemented changes to **avoid misleading comparisons in our results**. In the original manuscript, Figure 4 and the related discussion might have given the impression that strategies with more measures (like D+G+Tun30) are inherently better simply due to a higher “flexibility” score.

- (1) In the revised version, we have made the following adjustments in Figure 5 interpretation: We rewrote the description of Figure 5’s findings to remove any implication that the number of measures alone makes one strategy superior. We now emphasize a more nuanced interpretation. For instance, instead of stating “D+G+Tun30 is more flexible and thus better than single-measure options,” we say: “However, we found that the D+G+Tun30 pathway achieved a well-balanced performance across risk reduction, cost-effectiveness, and our

flexibility metric. In our initial analysis, this made D+G+Tun30 appear as the most promising option overall. However, as discussed, this planned flexibility advantage does not necessarily mean a single-component strategy cannot be adapted later. Therefore, rather than directly comparing flexibility values across all options, we focus on how each pathway performs on multiple criteria.” (in Results Section 4.3)

(2) No direct ranking purely by flexibility: We stopped short of any rank ordering of solutions solely based on the flexibility metric. If in the original text we had a sentence like “Strategy X is the most adaptive,” we have either removed it or qualified it heavily. Instead, we present the trade-offs: e.g., Strategy A has higher initial cost but was planned to adapt (multiple measures), Strategy B is simpler initially but could be expanded later if needed, etc.

(3) We strengthen the statement in Section 4.4, we add “However, its lack of initial flexibility – requiring a large up-front investment in gray infrastructure – could lead to path dependency if future conditions turn out to be mild. In contrast, strategies that start with smaller measures (like Dr or GA) and can add on bigger projects later avoid that risk of over-commitment. This underscores the classic tension in planning: a strategy like Tun70 is robust but inflexible, whereas a phased approach is flexible but may initially be less robust. Our framework attempts to balance these aspects by evaluating both.”

By making these changes, we hope we can address the reviewer’s concern that our comparison was systematically biased in favor of multi-measure solutions. We now highlight that single-measure solutions (like GA or Dr alone) remain viable and potentially flexible in a real-world sense, even if our metric gave them a lower numerical flexibility score. This revised approach ensures the comparisons in the paper are fair and not misleading to the reader.

RC2-4: *"Another implication of my previous comment #2 is that the proposed methodology does not really address one of its declared main purposes, i.e., finding trade-offs between robustness and adaptiveness (L195-196; L129-130; L146-147; L645-646; L653). In general, by combining a larger number of flood control measures over time, it should be expected that the resulting (gradually increasing) mitigation effects will be greater, during the infrastructure lifespan under the considered dynamic operating conditions (precipitation, sea level, etc.), compared with mitigation strategies using fewer gradual interventions; those solutions with many gradual interventions will be also ranked as more “flexible”, because of the ill-defined metric for flexibility. Because of this, solutions that implement many gradual updates (and hence that tend to achieve greater performance over time and more convenient cost-benefit ratios, through deferring expenses over time) will always tend to be preferred by this methodology, as they will be systematically regarded as more flexible too, with the risk of introducing bias. This is evident from Fig. 4, where the authors take D+G+Tun30 (brown curve)*

as the best flood mitigation approach (L473-475), due to its “well-balanced overall risk control performance and high value of flexibility”. However, if it were not for its high value of “flexibility”, that solution would be regarded as sub-optimal. Based on all the other objective values shown in Fig. 4, the actual best solution is instead Tun70 (blue curve), since it displays better values than D+G+Tun30 for average risk reduction rate (ARRR), Cost-Benefit ratio, and “valid period” (i.e., the period during which that flood control approach is effective). The low flexibility rank of Tun70 is only due to how flexibility is measured and does not reflect the fact that Tun70 would not preclude, in principle, combining other flood control strategies in the long period, if someday in the distant future upgrades were deemed necessary.”

Response: The reviewer is correct that, in the original manuscript, we concluded that D+G+Tun30 was the “best” solution largely because of its high flexibility metric, whereas another option (Tun70, the deep tunnel with 70% capacity) actually outperformed D+G+Tun30 on most other metrics. This pointed out a bias in our trade-off analysis. We have taken the following steps to address this:

(1) Revise the claim of D+G+Tun30 being unequivocally “best”: In the Results (and Abstract/Conclusions), we no longer use language suggesting that D+G+Tun30 is the optimal pathway in a broad sense. Instead, we now say it was the most *robust under our multi-objective criteria*, but we immediately caveat that this finding is sensitive to the inclusion of our flexibility metric. For instance, the statement at Lines 473-475 has been revised to: “Our multi-objective analysis initially identified the D+G+Tun30 pathway as the most promising, as it achieved a good balance across risk reduction, cost-effectiveness, and the (planned) flexibility metric. However, if flexibility is evaluated differently or not given as much weight, a different option (e.g., the larger single tunnel, Tun70) could emerge as preferable for long-term risk control.” In section 4.3.

(2) Emphasized the conditional nature of the trade-off: We adjusted our discussion of “robustness vs adaptiveness trade-offs” to clarify that our framework *formulates* such a trade-off but, in this case, did not fully explore it because of how adaptiveness was quantified. We have inserted a note that no true multi-objective optimization was performed (see also RC2-20), and thus we manually compared solutions. In doing so, we now highlight that different weighting of objectives would lead to different “optimal” choices.

(3) Tun70 vs D+G+Tun30 discussion: We added a specific discussion (in Section 5, Discussion) directly referencing the reviewer’s insight. We now mention that Tun70 outperforms D+G+Tun30 in several aspects and that if one were to prioritize those aspects (robustness metrics like ARRR and long-term efficacy) over the flexibility count, Tun70 would be a strong candidate for the optimal solution. We explain that our framework, as applied, favored the multi-component pathway because we explicitly valued planned flexibility – a methodological choice rather than an inherent truth. This level of transparency addresses the potential bias.

(4) Conclusion tempered: In the Conclusion section, we have tempered any statements about “the best strategy”. We instead focus on the methodological finding (that integrating robustness and planned adaptiveness can reveal trade-offs and potentially different choices than using robustness-alone). We also explicitly mention that the identified optimal solution might change if different metrics or assumptions were used, which is why robust decision frameworks are useful to decision-makers (they can stress-test such assumptions).

In summary, we fully accept the reviewer’s point. We have removed overstatements and given a more nuanced interpretation of our results, making it clear that D+G+Tun30’s selection was conditional on our approach. We hope this revised could be accepted by the reviewer.

RC2-5: *"The entire methodology section is vague and unclear, in the sense that it does not give a sense of logical flow starting from raw data processing to the final product. Different subsections describe different parts of the methodology, but they look somehow disconnected from each other. Given that the purpose of the paper is to provide a novel methodological framework for adaptive decision making, I strongly recommend revising this section in a way to guide the reader step by step through the proposed analysis framework. It may be beneficial to move Fig. 2 to the methodology section and organize the section around that Figure."*

Response: We have substantially revised Section 2 (Methodology) to improve its clarity, logical flow, and cohesion. The following specific changes have been made:

(1) We add a new Figure 1 (which illustrates our framework flowchart) right at the beginning of the methodology. We reference this figure as a roadmap for the analysis.



Figure1 Integrated framework of robust adaptive pathways for long-term flood control

(2) We added overview text at the start of Section 2 that walks the reader through the methodology at a high level. *Revised text (Section 2 Intro)*: “The proposed decision-making framework consists of eight sequential steps. (1) The process begins with the definition of flood adaptation objectives and relevant constraints. (2) A diverse set of candidate flood control options—both structural and non-structural—is identified. (3) An ensemble of plausible future scenarios is generated to represent key sources of deep uncertainty (4) Each individual option and selected combinations are evaluated across all scenarios using a simplified hydrological simulation model. (5) Performance metrics are calculated for each strategy. (6) Robustness are assessed using metrics like regret, and failure-prone scenario clusters are identified. (7) Adaptive pathways are formulated by sequencing options based on performance and scenario-specific tipping points or signposts. (8) A multi-criteria analysis is then conducted to support the selection of strategies that perform reliably under deep uncertainty.”

(3) At the end of each subsection, we added a sentence or two that links it to the next. For example, after the robustness evaluation subsection, we added a line like “This robustness assessment provides the baseline upon which we evaluate adaptiveness in the next step.”

RC2-6: *"Much information contained in the supplementary material is critical for understanding the proposed methodology and the presented case-study implementation; it is therefore suggested to suitably include parts of it into the revised methodology section, when addressing my previous comment #5."*

Response: We agree that certain details from the Supplement were important for understanding our methodology and should be included in the main text. In revising the Methodology (Section 2), as described above, we have integrated the critical pieces of information from the Supplement into the manuscript:

We identified which parts of the supplement were essential for comprehension. For example, our Supplement contained definitions of some parameters (like in the adaptation tipping point analysis), additional explanation of the PRIM algorithm settings, and decision criteria that were only summarized in the original text. We then incorporated those details into Section 2 at the appropriate places. For instance:

- When discussing scenario generation and model input data, we brought in a brief summary of data sources and any specifics (previously in Supplement) that are necessary to follow the case study.
- In describing the PRIM analysis, we included the definitions of “coverage” and “density”.
- For the adaptiveness analysis, we introduced the mathematical formulation (e.g., how

we compute the “adaptiveness score” or flexibility metric) directly in the text, rather than referring the reader to the Supplement.

RC2-7: *"Lines 176-179: candidate hydrologic-hydraulic models suggested for use in the proposed decision-making procedure are quite different from the simpler SCS-CN model that is ultimately adopted in the case-study implementation (L395). I would state upfront that the methodology can leverage either sophisticated 1D-2D or simpler models, with likely major effects on the computational times. I would also state in the methodology section what model was ultimately selected for the case-study application, instead of mentioning it in the results (L395)."*

Response: We have made two clear additions to address this comment:

(1) Upfront statement about model choices: In the Methodology section (specifically where we discuss the modeling step, which is Section 2.1 in the revised structure), we now explicitly note that our framework is model-agnostic in the sense that it can use either complex 1D-2D hydrodynamic models or simpler conceptual models, depending on the case. We also acknowledge the trade-off in computational time and detail between these choices. For example, we inserted: “The framework can accommodate a range of hydrologic/hydraulic models. For instance, high-fidelity 1D-2D models (e.g., SOBEK, MIKE21) could be used for more detailed flood simulations, at the cost of greater computational effort, whereas simpler conceptual models can provide faster estimates of flood response.” (*in Section 2.1*). This sets the expectation that model complexity is a variable.

(2) State the chosen model in methodology: We moved the information about which model we actually used (the SCS-CN based rainfall-runoff model) from the Results into the Methodology. Early in Section 2.1, we now state: “In our case study, we employ a conceptual hydrological model based on the SCS-CN method to simulate runoff and inundation. This simpler model was chosen to keep computational requirements manageable, given the thousands of plausible futures simulations, although the framework could integrate more complex models if needed.”

RC2-8: *"(related to the previous comment) It important to note that model choice is expected to dramatically affect the overall computational times of the proposed procedure, but this is not clearly and exhaustively discussed at lines 618-620. In that paragraph, the authors emphasize the moderate computational and data requirements of the proposed methodology; however, this really depends on what hydrologic model is used within the methodology, and those considerations may not apply to cases where more sophisticated 1D-2D models (like those mentioned in the Methodology section; L176-179) are deployed."*

Response: We concur and have expanded our discussion regarding computational requirements model complexity:

(1) In the revised manuscript, when we mention the computational and data efficiency of our framework (originally at lines 618-620 in the Discussion), we now include a caveat that this efficiency was achieved using a relatively simple model (SCS-CN) and may not hold if a complex model were used. We effectively tie this point back to the model choice discussion in Methodology.

(2) Specifically, we modified the paragraph around line 800: “In our case study, the framework was able to evaluate thousands of scenario–strategy combinations within a few days, with moderate computational and data requirements. This was largely because we employed a simplified hydrological model (SCS-CN) for simulation. If a more sophisticated 1D-2D model were used, the computational time and data needs would increase substantially, and the ‘moderate’ resource demand we report would no longer apply.”

We then add: “Thus, the practicality of this framework in other settings will depend on careful selection of the modeling approach or on the use of advanced computational techniques (e.g., emulators or parallel computing) when high-resolution models are necessary.”

RC2-9: *"The term “valid period”, repeatedly used in the paper (e.g., L188; L200; L303; L308; L438; L489; L507; L520; L551; Table 1, etc.) to refer to the finite time period during which a given (single or combined) flood control measure is effective (L438-439), is confusing. I would change it to something easier to interpret, such as “effectiveness period” of a given flood control measure."*

Response: Changing to “effectiveness period” not only clarifies meaning but also aligns with common usage in adaptation literature (some publications use terms like “design life” or “adaptation tipping point year,” but “effectiveness period” is intuitive as well). We have adopted the reviewer’s excellent suggestion to improve this terminology:

(1) All occurrences of “valid period” in the manuscript have been replaced with “effectiveness period”, which we agree is much clearer. This change appears in the text, tables, and figure captions wherever applicable.

(2) In addition, to ensure the term is understood, we provided a definition at first use (which is around the end of the Introduction or beginning of Methodology when we first mention this concept). For example, we now say: *Revised text:* “...the effectiveness period of an option (the time span during which the flood control measure meets the specified performance targets).”

(3) We double-checked instances like Table 1 and the Results section lines (L438-439, etc.) to

ensure consistency in the updated term.

RC2-10: *"L661-665: I tend to disagree with the authors' claim that their work provides a theoretical foundation for decision-making methods in flood mitigation for coastal megacities. A "theoretical foundation" should imply the development of an entirely novel mathematical framework for addressing decision-making problems. While the resulting decision-making framework is novel, the authors primarily build upon existing metrics, models, and algorithms. Because of this, I would characterize the contribution of this work as a synthesis and application of established methodologies rather than the creation of a theoretical foundation."*

Response: We agree with the reviewer's assessment and have revised the wording in our Conclusions to accurately characterize our contribution:

(1) We no longer claim to have developed a wholly new "theoretical foundation." Instead, we now describe our work as an integration or synthesis of existing concepts applied to a complex problem. For instance, at the end of the paper we changed the phrasing to something like: "This work provides a novel decision-making framework for flood mitigation in coastal megacities by synthesizing and building upon established DMDU methods (such as RDM and adaptive pathways)." And we continue, "Rather than introducing a new theory, our contribution lies in the innovative combination and application of these methods to address the joint challenges of robustness and adaptiveness in flood risk management."

(2) By using words like "framework" or "approach" instead of "theoretical foundation," we make it clear that our work is practical and integrative in nature, not a fundamental new theory. We also explicitly credit existing metrics, models, and algorithms as the building blocks of our approach (with citations to where we got them).

Indeed we used known metrics (ARRR, BCR, regret), known algorithms (PRIM, etc.), and our novelty was in how we combined them and the context in which we applied them. Now we think the paper reflects that more humbly and accurately.

RC2-11: *"Eq. 1 (L221) and related text (L222-224) are unclear, because apparently the word "option" is used to refer to both individual flood control measures (a_i) and combinations of them (a or a_p) (L222-224). If the authors intend to determine the combination a_p of one or more options a_i that maximizes flood control performance across the range of scenarios w_j , chosen from the set of combinations a , then I suggest using different notations to indicate individual options and combinations of them (for example, keep a_i for individual options, and indicate combinations as c and c_p , respectively; also clearly distinguish between options and combinations of options in the text L222-224). Using the current notation, a_p could be*

mistakenly interpreted as a specific option, not as a combination of options. On the other hand, if the authors indeed intend to determine what individual option a_p among all candidates a_i shows the best performance across all operating condition scenarios w_j , then the current notation is formally incorrect, since the max function is applied to a single value, not a set of values. The correct notation should exclude the summation for $i=1$ to $i=m$ from the current version of the equation and the range of variability of i should be specified below the max operator (in place of a)."

Response: We appreciate this detailed feedback on our notation. In response to RC2-11, Equation (1) to (3) have been revised to eliminate ambiguity and to clearly express the intended decision rule. The revised formulation identifies the option (or combination of options) that achieves the highest average performance across all considered scenarios, in line with the Laplace criterion. Specifically, the equation (1) is now written as:

$$a^* = \arg \max_{a \in A} \left(\frac{1}{N} \sum_{j=1}^N f(a, w_j) \right)$$

where a^* denotes the optimal option or combination, A is the set of all alternatives (listed in table 2), N is the total number of scenarios. And $f(a, w_j)$ is the performance of option a under scenario w_j , which represents the expected flood risk associated with adaptation option a under scenario w_j , as generated from the flood-damage simulation model. This risk value forms the basis for evaluating robustness through indicators such as the ARRR.

The equation (2) is now written as:

$$P_i = \frac{1}{N} \sum_{j=1}^N \frac{|f(a, w_j) - f(a_0, w_j)|}{f(a_0, w_j)}$$

Where P_i is the performance value of average risk reduction rate (ARRR) of alternative options or combinations in all plausible futures N , $f(a, w_j)$ represents the performance value of pluvial flood risk of option a in future w_j , and $f(a_0, w_j)$ is the performance value of pluvial flood risk of the baseline option a_0 in future w_j .

The equation (3) is now written as:

$$P_i^* = \frac{1}{s} \sum_{j=1}^s \frac{|f(a, w_j) - f(a_0, w_j)|}{f(a_0, w_j)} \geq F_0$$

Where P_i^* represents the performance value of ARRR of the option or combination a in subsets s of all plausible futures N that meets the given flood control target F_0 . PRIM is applied to identify clusters of successful cases by searching across the full set of futures N for each option or combination. Specifically, for each option or combination a , we select the subset of future s that leads to the most successful outcome by balancing coverage and density with given flood control target.

RC2-12: *"What performance metric is ultimately adopted for calculating $f(a_i, w_j)$ in Eq. 1 and 2?"*

Response: In the original text, we did not explicitly state which performance metric $f(a, w_j)$ represents – an important oversight. We have now clarified this both in the Methodology section where we introduce the equations and in the Results context:

In **Section 2.2 (Robustness Analysis)**, when we first introduce $f(a, w_j)$ in equations (1) and (2), we now immediately explain what f is in our case study. For example, we added: "...where $f(a, w_j)$ is the performance of option a under future w_j . In this study, we quantify performance in terms of flood risk reduction, specifically using the Average Risk Reduction Rate (ARRR) as the metric (defined in Section 2.2)."

Later in the Methodology equation 2, "Where P_i is the average performance of alternative options, $f(a_i, w_j)$ represents the performance value of scenario j , and $f(a_0, w_j)$ is the performance of the baseline option of scenario w_j ", we revised to: Where P_i is the performance value of average risk reduction rate (ARRR) of alternative options or combinations in all plausible futures N , $f(a, w_j)$ represents the performance value of pluvial flood risk of option a in future w_j , and $f(a_0, w_j)$ is the performance value of pluvial flood risk of the baseline option a_0 in future w_j .

RC2-13: *"Eq. 4 (L297) is unclear as many of the variables that appear in it are not defined anywhere. Eq. 4 is also formally incorrect, as distinct conflicting objectives considered in multi-objective problem formulations (for which trade-off solutions should be determined) should be expressed as separate objective functions to maximize or minimize."*

Response: Equation (4) in our manuscript was intended to represent the multi-objective optimization (balancing robustness and adaptiveness), but we admit it was not well-presented. Equation (4) has been reformulated to clearly represent a multi-objective optimization problem, with each objective explicitly defined and evaluated. The revised equation expresses the decision-making goal as identifying the adaptation pathway l that simultaneously maximizes four distinct criteria: flexibility, effectiveness period, economic efficiency, and risk reduction performance. This is now formulated as:

$$\max_{l \in L} F(l) = [y_1(l), y_2(l), y_3(l), y_4(l)]$$

Where $l \in L$ is a candidate adaptation pathway from the set of feasible pathways L ; $y_1(l)$: Flexibility — number of successful alternatives reachable from pathway l ; $y_2(l)$: Effectiveness period — duration before performance drops below threshold; $y_3(l)$: Benefit-Cost Ratio (BCR) — economic efficiency of pathway l ; $y_4(l)$: Average Risk Reduction Rate (ARRR) — robustness of flood risk performance.

Equation (4) has been revised to more accurately represent a multi-objective decision problem, with each objective now clearly separated and all variables explicitly defined. In this study, a manual trade-off analysis was conducted based on an **equally weighted aggregation** of multiple criteria, rather than solving the problem through a formal algorithm such as a genetic algorithm. The revised manuscript clarifies that the decision-making process relied on Pareto-front comparison and visual inspection of trade-offs between robustness, flexibility, effectiveness period, and economic efficiency. While a weighted sum approach was adopted using equal weights for simplicity and transparency, the structure of Equation (4) has been retained in a general form to remain compatible with more advanced multi-objective optimization techniques. This makes it possible for future research to apply algorithmic methods—such as evolutionary algorithms—using the same formulation. These revisions address the reviewer’s concerns by correcting earlier formal issues, improving clarity, and making the method adaptable to a broader set of applications.

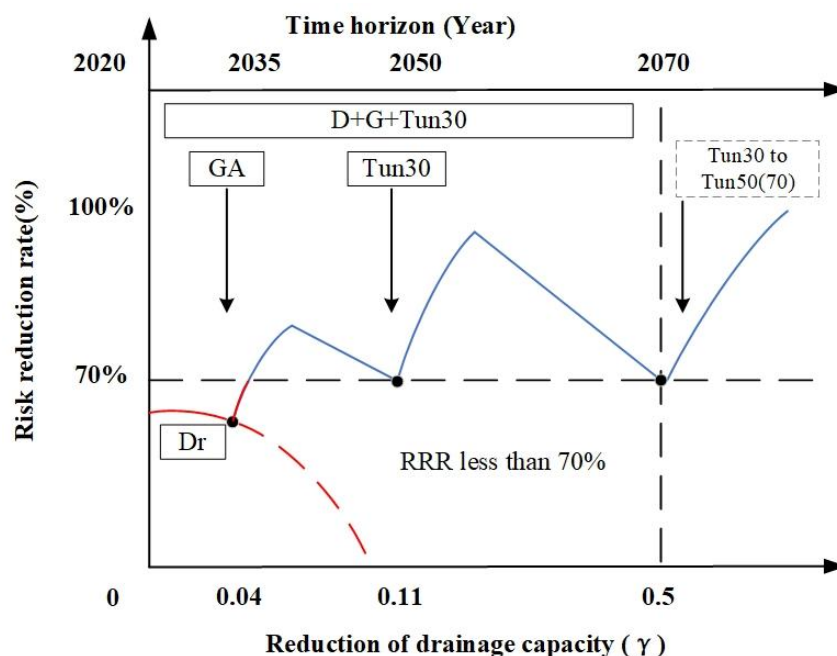
RC2-14: *"Fig. 5 and 6: what is the maximum time horizon considered for normalizing gamma in the x-axis (see Supplementary materials)? In other words, how many years correspond to gamma=1? For both figures, I suggest showing the time scale both in terms of gamma and the number of years."*

Response: Thanks for the suggestion. We have clarified the meaning of the non-dimensional time variable γ , and to avoid the duplication of time horizon information, we updated Figures

5 accordingly.

In our case study, γ was a normalized time, and the maximum value equal to 0.5. Based on the Supplement or our simulation setup, we identified the actual number of years that correspond to $\gamma = 0.5$. For example, suppose $\gamma = 0.5$ means 50 years (just as an illustration, the exact value is in our data 50 years). We confirmed this from our Supplement or model assumptions and then stated it explicitly in the text.

We modified Figures 5 to include a secondary x-axis or additional tick labels indicating the real time in years. For instance, below the γ axis, we add markers like 2035, 2050, 2070. In the figure captions, we now explain: “ γ is a dimensionless time, where $\gamma = 1$ corresponds to **Year 2070** (the end of our planning horizon).



In the **Results section text (Section 4.3)**, we also mention the time horizon when first discussing γ . E.g., "We normalized time as $\gamma = t/T$ (with $T=50$ years, the simulation period), so γ corresponds to the year 2070."

RC2-15: "Fig. 5, L501. Why, from the time of GA and Tun30 installation (black dots, if I am not mistaken), is there an initial gradual performance increase (i.e., risk reduction) over time? Given that gradually worsening conditions (in terms of extreme precipitation trends, land subsidence, and sea level rise) are considered, should not performance (i.e., risk reduction) steadily decrease over time, with only “instantaneous” jumps associated with the installation? Even if we consider a finite period for the installation time, shouldn't the performance increase only after the installation is complete? Please clarify."

Response: Thanks very much for the observation of an issue in Figure 5: the first black dot representing the implementation of the GA (Green Areas) measure is misplaced. It currently appears after the risk reduction rate has already started to rise, which misrepresents the actual decision logic. The black dot should be repositioned at the **point in time where the risk reduction rate begins to increase**—that is, where the drainage (Dr) strategy becomes insufficient and the GA measure is triggered—even if that performance level is still below the 70% threshold. This correction reflects the correct interpretation of the pathway logic and aligns with the underlying simulation output. The figure will be revised accordingly, and the updated caption and accompanying text will clarify that the placement of decision points is based on system performance trends, not fixed thresholds.

Added text Section 4.4: “It is noted that the slight rise in performance after GA and Tun30 installation reflects a short ramp-up period in our model, during which newly implemented measures gradually reach full effectiveness, then performance begins to decline as expected under continued climate-induced stressors.”

RC2-16: *"Fig. 5, L501. In the decreasing limbs, why does performance decrease non-linearly for Dr alone but then later it decreases linearly for Dr+GA as well as for Dr+GA+Tun30? I would have expected more non-linearity in performance decrease when multiple flood control measures operate together, since the curve would be reflecting the cumulative effects of supposedly different performance reduction dynamics over time. Please clarify."*

Response): We appreciate the reviewer’s comment and agree that the differences in performance decline across strategies require clearer explanation.

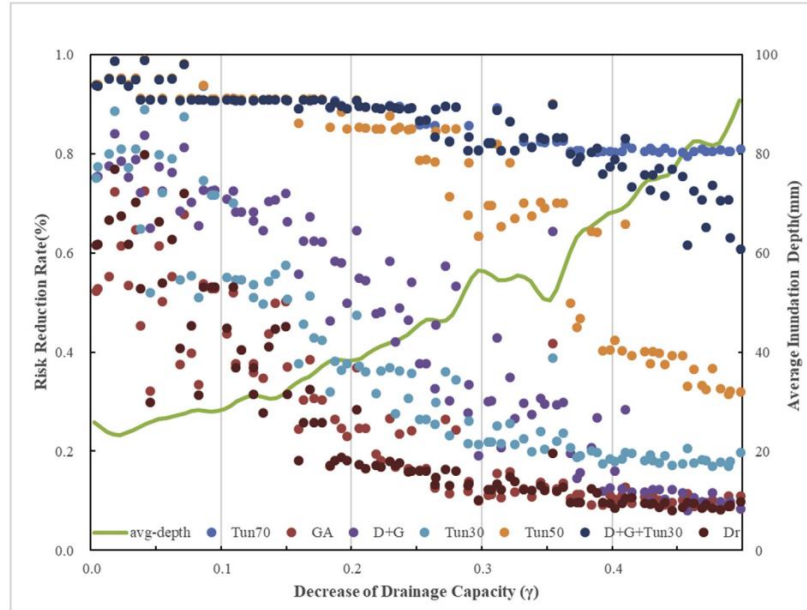


Fig. 6. Risk reduction rate of the seven selected strategies and the average inundation depth across the combinations of solution and rainfall scenarios at the given level of γ . Tun70: deep tunnel with 70% runoff absorbed under the baseline; GA: green area expansion; D + G: drainage enhancement + GA; Tun30: deep tunnel with 30% runoff absorbed under the baseline; D + G + Tun30: drainage enhancement + green area + Tun30; Tun50: deep tunnel with 50% runoff absorbed under the baseline; Dr: drainage enhancement. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Based on the Figure (Figure 6 from Hu et al., 2019), the Risk Reduction Rate (RRR) trajectories for different strategies reveal distinct behaviors that reflect both the composition and resilience of the interventions.

The Dr-only strategy (drainage enhancement) exhibits a non-linear, concave decline in RRR as γ increases. This pattern arises because drainage alone quickly loses effectiveness under worsening conditions, leading to a steep initial drop in risk reduction. As drainage capacity further degrades, the rate of decline slows, forming a concave curve. This behavior is also supported by Figure (Figure 6 from Hu et al., 2019), which shows a wide interquartile range for Dr, indicating high variability and sensitivity to uncertainty.

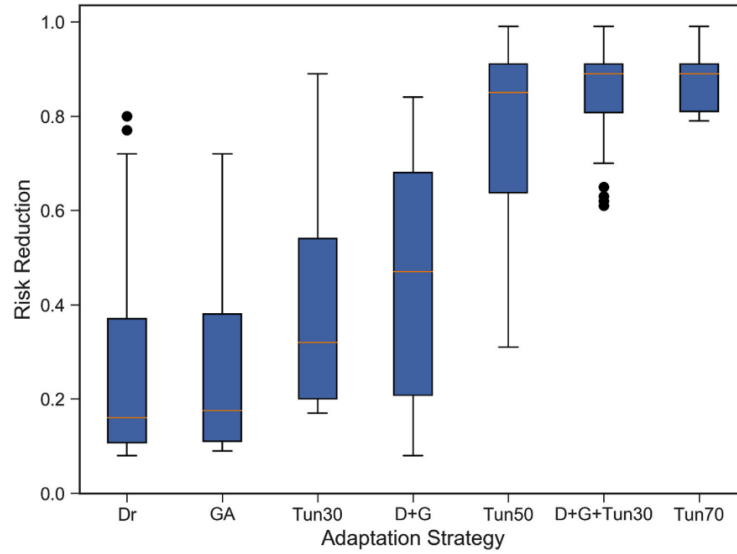


Fig. 7. Box plots of potential risk reduction rates. Dr: drainage capacity enhancement; GA: green area increase; Tun30: deep tunnel with 30% runoff absorbed; D + G: Dr + GA; Tun50: deep tunnel with 50% runoff absorbed; D + G + Tun30: Dr + GA + Tun30; Tun70: deep tunnel with 70% runoff absorbed. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

In contrast, the D+G strategy (drainage + green area) maintains a more linear decline in performance. The combination of structural and green infrastructure balances performance loss, moderating the pace of decline as γ increases. This indicates a more consistent risk reduction effect over time.

The D+G+Tun30 strategy shows a two-phase pattern: performance remains stable until around $\gamma = 0.25$, after which the RRR begins to decline linearly. This inflection point indicates that the addition of tunnel storage (Tun30) enhances early-stage robustness, delaying the onset of decline. Once γ exceeds the buffering capacity of the tunnel, the system transitions into a steady degradation phase.

We added a clarification in the text Section 4.4: “The differing curvature of the performance decline is due to the interaction of measures. For the Dr-only strategy, once implemented, its risk reduction gradually diminishes at an accelerating rate as climate stressors intensify – producing a concave-down curve (initially gentle slope, steepening later). In contrast, the strategies with multiple measures (D+G and D+G+Tun30) show a more linear decline. This is because when drainage alone begins to lose effectiveness, the next measure (GA, and later Tun30) either has just been implemented or is concurrently mitigating risk, effectively offsetting some of the non-linear drop. The combined result is a more steady (approximately linear) decrease in performance over time, as the measures’ effects complement each other.”

Reference

Hu, H.Z, Tian, Z., Sun, L.X, et al. (2019). Synthesized trade-off analysis of flood control solutions under future deep uncertainty: An application to the central business district of Shanghai. *Water research*, 166, 115067.

RC2-17: *"Notation used in equations 1, 2 and 3 (L221-247) should be consistent with the notation adopted later in the Results section (L406-459)."*

Response: We completely agree that consistency in notation is critical. We have reviewed all equations (1, 2, 3) in the Methodology and cross-checked them against the notation in the Results section (which corresponds to original L406-459 and surrounding text).

In the revision, we unified the terminology and symbols. For example, if the Methodology introduced certain symbols for options, scenarios, thresholds, etc., we ensured the Results use the *same symbols* when referencing those concepts. Originally, we suspect there might have been a lapse, e.g., using "*a*" in the equations but then talking about "option" or "strategy" in results without the subscript, or using different letters like *x*, *y* later on. All such inconsistencies have been fixed.

RC2-18: *"Eq. 4 in the Supplementary material, Fig 3 in the manuscript: In the case of combined flood control measures implemented gradually (e.g., $D+G$, $D+G+Tun30$), do the authors consider the effects of the discount rate on the investment costs for additional interventions deferred over time? Are these effects reflected in the benefit-cost ratios plotted in Fig. 6?"*

Response: Thanks very much for this important issue. We realized we had not explicitly mentioned discounting in the manuscript. We have now stated it clearly:

(1) In our analysis, we did not incorporate a discount rate for future costs in the benefit-cost ratio (BCR) calculations – all costs were treated in constant terms for simplicity. We have now explicitly acknowledged this in the paper. For instance, in Supplementary Text 4 we describe BCR, we added: "For simplicity, costs of measures were not discounted in our analysis, and we treated all monetary values in constant terms."

(2) We then addressed how this choice might affect results. Specifically, in Section 4.1 we added: It is noted if a strategy defers a major investment (like the Tun30) to later years, in reality its present value cost would be lower with discounting, potentially making the strategy more economically attractive than our simple BCR suggests.

RC2-19: *"L432-450, Table 3; L615-616: the authors should provide the definition of density and coverage in the context of PRIM algorithm. Also, it is not clear how Table 3 is obtained by applying the condition given in Eq. 3."*

Response: We thank the reviewer for the helpful observation. *Coverage* is defined as the

proportion of all scenarios that satisfy the identified condition (i.e., the scenario “box” coverage). *Density* is defined as the proportion of scenarios within that “box” that are successful (or meet the target criteria). We have added the missing explanations regarding the Patient Rule Induction Method (PRIM) results:

(1) In Section 2.1 6) adaptiveness analysis, we add a brief sentence like: “(In PRIM, ‘coverage’ refers to the fraction of the total scenario space included in the identified hyper-rectangle (subset), and ‘density’ refers to the fraction of cases in that subset that meet the specified success criteria.)”

(2) In addition to the definitions of coverage and density, we now clarify how Table 3 is derived based on a combination of Equation (3) and the PRIM (Patient Rule Induction Method) algorithm.

Equation (3) defines the failure condition, where a strategy is considered to fail if its performance falls below the flood risk reduction target $F_0=0.7$ at any given level of the system stressor γ . PRIM is then applied to identify clusters of failure cases by searching across the full set of scenarios for each adaptation option. Specifically, for each option, we isolate the subset of scenarios that lead to failure according to Eq. (3).

Using the resulting failure boxes from PRIM, coverage is calculated as the proportion of all failure scenarios captured by the box, while density reflects the proportion of failure cases within the box relative to all scenarios it contains. Once these scenario clusters are identified, the Average Risk Reduction Rate (ARRR) of each option is calculated over its corresponding PRIM-derived subset.

The manuscript has been revised to clarify this connection more explicitly for the reader in Section 4.2 from L433-L440: “Scenario discovery confirms that declining drainage capacity is the most critical uncertainty influencing the risk reduction rate. Failure scenarios are identified when the flood control target $F_0=0.7$ is not met. Using the PRIM algorithm, we extract failure-prone subspaces for each option by optimizing the combined value of coverage and density. Table 3 summarizes these metrics, where coverage and density are derived from PRIM-identified failure boxes, and ARRR is calculated as the average performance within those clusters. The valid period is defined by the point (characterized by γ) when a single option or combination no longer meets the performance target.”

RC2-20: *"L352: the authors state that they performed a trade-off optimization, balancing robustness and adaptiveness. However, later in the paper the authors clearly state that they actually did not run any optimization, which is deferred to future work (L637-641). I would clarify from the beginning that the contribution of the manuscript only focuses on the*

formulation and not on the solution of a specific complex optimization problem defined using the proposed framework. It is worth highlighting here that the case study considered in this work had a very small number of candidate solutions, which made it possible to "manually" enumerate and compare all of them, to identify the optimal solution; however, more realistic applications, with possibly hundreds of candidate flood control scenarios, will require using an optimization algorithm to reduce the number of hydrologic simulation that would be otherwise required if all scenarios were simulated. Incidentally, L639-640 include an incorrect statement, as genetic algorithms are not machine learning methods."

Response: We appreciate this comment as it helps us clarify the scope and avoid confusion. We have made several changes to address these points:

(1) In Section 3.2, We went through instances where we might have implied we did an optimization. For example, at original L352 we change phrasing from “we performed a trade-off optimization” to “we conducted a trade-off *analysis*”.

(2) Similarly, Section 5.3, we ensure consistency. We emphasize that solving the optimization problem (especially via algorithm) was outside the scope and is left for future work. We also mention that if there were hundreds of possible interventions or sequences, then one would indeed need to integrate an optimization algorithm or heuristic to search through the possibilities efficiently.

(4) We now refer to genetic algorithms appropriately as a heuristic optimization technique or an evolutionary algorithm. We removed the phrase “machine learning” in that context. We replace it with: “an evolutionary optimization algorithm.”

RC2-21: *"L617-627: these limitations should be discussed earlier, possibly in a dedicated subsection of the revised Methodology section."*

Response: We have followed this advice. The limitations of our study (which were originally briefly listed around L617-627 in the conclusion/discussion) are now **discussed earlier** and more prominently in the manuscript.

Section 2.1, model simulation, we revised the paragraph to: “Evaluate each option (and combinations of options) under all scenarios using an appropriate flood simulation model. The framework can accommodate models of varying complexity: for instance, high-fidelity 1D/2D hydrodynamic models (e.g., SOBEK, MIKE 1D2D; Wang et al., 2018) could be used for detailed analysis at the cost of more computation, whereas simpler conceptual models allow faster simulation of many scenarios. In our case study, we employed a simplified hydrological model based on the SCS-CN method to simulate runoff and flooding, which kept computational demands manageable given the thousands of plausible futures simulations, although the

framework could integrate more complex models if needed.”

Section 5.3. We revised the sentence to: “One advantage of our decision-support tool is that it can run comprehensive evaluations for thousands of future–option combinations within a few days, using only moderate amounts of input data. This computational efficiency is largely due to our use of a simplified model (SCS-CN) and a relatively small case-study area. However, this highlights a trade-off: using a more detailed 1D–2D model or expanding to a larger region would substantially increase computational time and data requirements. In other words, the ‘moderate’ resource demand we experienced may not hold in cases that require high-resolution modeling. This limitation suggests that careful model selection (or the use of techniques like emulators and parallel computing) is important when applying the framework to bigger or more complex systems.”

RC2-22: *"L676-677: The code used to perform the analysis would be an important asset for the reviewers to evaluate the proposed decision-making framework and the case-study implementation, especially in light of the unclear methodology section. In line with HESS journal policies, I encourage the authors to share polished open-source code and the complete dataset to enable the reproducibility of their analysis."*

Response: Thanks for the suggestion. We agree that sharing our code and data will greatly enhance transparency and reproducibility. We are committed to open science and will make our materials available in accordance with HESS policies. We will prepare a replicable repository of our analysis code and a curated version of our dataset.

Minor comments:

(We also address the reviewer’s minor comments below. Each minor issue raised has been corrected in the text to improve clarity and accuracy, as detailed.)

RC2-23: *"L98-101 The flow in this sentence is broken, like if there were some missing connectors or else two individual sentences got messed up into one."*

Response: We have rewritten the sentence spanning Lines 98–101 to fix the broken flow. The corrected text in the Introduction now reads: “However, an open question remains: to what extent can a traditional robustness evaluation (especially under risk-averse assumptions) suffice for rational decision-making, versus using a multi-objective trade-off analysis to gain a more comprehensive view?”

RC2-24: *"L182 “the calculation of indicators such as the average risk reduction rate (ARRR) and benefit-cost ratio (BCR) for each alternative option was performed for each scenario”. In*

this paragraph, some metrics are mentioned but their mathematical expression is not provided in the Methodology; additionally, the wording “indicators such as” is not appropriate in the Methodology section, as it is ambiguous. Does it mean that other metrics are calculated, besides ARRR and BCR? Or else do the authors intend that ARRR and BCR represent just examples of possible robustness metrics that the modeler may decide to use?”

Response: Thanks. ARRR and BCR were the main indicators of robustness in our analysis. We ensured ARRR and BCR are defined: e.g., ARRR is percentage reduction in expected damages compared to baseline, averaged over scenarios” and BCR is “ratio of total avoided damage to total cost”. These definitions are now in the methodology so the reader knows exactly how they’re computed (connecting also to RC2-12 about performance metric).

Additionally, we have clarified this part of the Methodology:

Assess each option’s performance across all scenarios using robustness criteria, which used to be depicted as $f(a, w_j)$ meaning the performance of option a under scenario w_j . In this study, we assume all scenarios are equally likely (an application of Laplace’s principle of insufficient reason) and compute performance indicators for each option under each scenario. Key indicators include the average risk reduction rate (ARRR), percentage reduction in expected damages compared to baseline, averaged over scenarios, and the benefit–cost ratio (BCR), ratio of total avoided damage to total cost. Using these, we evaluate how “robust” each option is: i.e., how well it performs on average and whether it consistently meets acceptable thresholds across scenarios.

RC2-25: *"L185 “Subsequently, the performance of each option and its combination was evaluated by quantitative comparison and ranking stability”. In this sentence, the notion of stability of the options is suddenly introduced, but it is never defined in the manuscript."*

Response: We realized that the term “ranking stability” was introduced without explanation. We have now removed it in Section 2.1 5) Robustness analysis.

RC2-26: *"L188 “Valid periods of the alternative options were determined based on the conditions of the successful scenarios under each (individual or combined) option, in conjunction with a specific flood control objective.” From this sentence, the concept of “valid period”, quite redundant throughout the paper, is not clear (also see my main comment #9). A clearer description of such concept is given in the Results section (lines 438-439); however, it would be beneficial to the reader if the concept of “valid period” were introduced earlier, from the first time that it is mentioned."*

Response: As addressed in our response to RC2-9, we have replaced the term “valid period” with “effectiveness period” and, importantly, we now introduce and explain this concept at its first occurrence (which is around L188 in the Methodology).

In the revised Methodology text, we now say:

“Determine how long each option remains effective and how easily it can be adjusted. For each single or combined alternatives, we identify its effectiveness period – the duration or range of conditions over which it meets the flood risk target – by finding the point at which its performance falls below the acceptable threshold.”

RC2-27: *"L200-205: concepts such as “transient scenarios” and “signposts” are unclear in this section of the methodology. They become clear to the reader after reaching the Results section (Fig. 6), but the article should be organized in a way that all concepts unroll smoothly following sections order."*

Response: We have added explanations for “**transient scenarios**” and “**signposts**” right when they are first mentioned in the Methodology (L200-205).

In the revised text of Section 2.1, 8) robust adaptive pathway, we now write:

“Formulate and select robust adaptive pathways. Using the information on each option’s effectiveness period and flexibility, we identify sequences of actions that extend flood protection over time. We generate an adaptation roadmap by considering how the system could transition under transient scenarios. From the set of possible pathways, we then select a robust adaptive pathway that best satisfies the flood control objectives in the long term based on the multi-objective analysis from step 7. Along this pathway, we define key signposts – measurable indicators (e.g., a threshold of rainfall intensity or drainage failure rate) that signal when it’s time to shift to the next action. Monitoring these signposts will support future decision-making and adjustments to the plan.”

RC2-28: *"L208: “The choice of robustness option is the meta-problem of how to decide (Herman et al., 2005)”. This sentence is too vague for and not related to the methodology section. I suggest either removing it or else moving it to the Introduction and elaborate more."*

Response: Thanks for the suggestions, we revised the text to: “Deciding on a robustness criterion is essentially a meta-decision problem (Herman et al., 2015). In our context, robustness of a strategy refers to its satisfactory performance across a range of uncertain future states.”

RC2-29: *"L362-375: in this paragraph, parameters alpha, beta, and gamma are mentioned for the first time without prior introduction. These parameters are defined in the Supplementary material but should be also properly introduced in the manuscript."*

Response: We have introduced the parameters α , β , γ in the main text where they first appear.

- uncertain factor of the α , future rainfall assumed to increase from 7% to 18%
- Urban rain island effect β assumed to increase from 10% to 20% in central region (Xujiahui and Pudong rain gauges), decrease from -0.076% to -0.038% (other 9 rain gauges in Shanghai)
- decrease of drainage capacity γ , assumed to decrease from 0 to 50% due to the anthropogenic land subsidence and sea level rise

RC2-30: "L408-415: "Benefit-cost is the evaluation dimension for the robustness metrics" needs rewording; "benefit-cost, was defined as the average risk reduction rate (ARRR) per unit cost based on the robustness metrics of Laplace's Insufficient Reason" does not explain with sufficient detail how benefits and costs are calculated; also, the concept of Laplace's Insufficient Reason is mentioned a few times in the paper but never introduced."

Response: Thank you for the insightful comment. We have revised the relevant sentences in Section 4 to clarify that the benefit–cost ratio (BCR) is used as a complementary robustness indicator, defined as the **average risk reduction (ARRR)** divided by the **implementation cost** of each option.

We revised: "Specifically, the benefit is the reduction in expected flood losses compared to the no-action scenario (Equation 3), while the cost refers to the total implementation cost of each adaptation option (Equation S5 in [Supplementary materials Text 4](#)). We adopt Laplace's principle of insufficient reason, assuming all scenarios are equally likely when calculating average outcomes across scenarios."

We introduced **Laplace's Principle of Insufficient Reason** in the Methodology 2.2, we add: "In this study, we adopted For the risk aversion metric, the neutral risk aversion of Laplace's principle of insufficient reason as one robustness measure: in the absence of known scenario probabilities, we assign equal weight to all scenarios and identify solutions that perform best on average.

In conclusion, we have carefully addressed all comments from Reviewer #2 by making the text more precise, adding needed definitions, reorganizing sections for better flow, and correcting any inaccuracies. We trust that these revisions have significantly improved the clarity and quality of the paper, and we thank the reviewer for their thorough and constructive feedback.