Response to reviewers' comments

"The dynamics of peak head responses at Dutch canal dikes and the impact of climate change" https://doi.org/10.5194/egusphere-2024-1495 submitted to Natural Hazards and Earth System Sciences

We thank the reviewer for her/his thorough, insightful and valuable feedback, both on a general and more detailed level.

Below, we reply to the reviewer's comments and explain how we will address them. The reviewer's comments are shown in *Italicized text in gray*, our responses are shown in blue. We provide detailed responses to the major comments, along with specific **actions** to improve the manuscript. For minor comments, we offer brief responses, as we will incorporate these suggestions to enhance clarity and refine terminology throughout the text.

Anonymous Referee #1

The paper reports on a comprehensive analysis of dike strength and failure for a large number of polder dikes across the Netherlands. Using a model for dike 'ground'water head calculation and time series of precipitation and evapotranspiration, the model produces time series of water heads in dikes. Results are statistically analyzed, including grouping of dike types across the area.

The subject fits well to the scope of NHESS, and addresses a field that has received increasing attention, both in water management and (applied) research into dike stability and flood risk.

Thank you for your positive feedback; we appreciate the recognition of our study's relevance to dike stability and flood risk research.

Still the manuscript requires revision before final publication. My main points are:

- *I miss a spatial component in the discussions of clusters and 'correlation': how would a map indicating cluster member of dike sections look - is there any spatial correlation, or not?*

  We acknowledge that the spatial component of clustering and correlation is not explicitly discussed in the paper. However, the coincidence matrix in Figure 9 provides an initial indication, showing that each cluster includes dikes from multiple water authority regions (different IDs). Additionally, Figure 8 suggests no clear spatial correlation in the impulse response characteristics, as noted in Lines 367–368: "*Initially, the variation of head responses across regions suggests no specific pattern, likely due to the heterogeneous subsoil conditions within the canal dike system, as shown by the random color distribution in Fig. 8.*" Nevertheless, we acknowledge the lack of a spatial discussion of the clusters and a map visualizing cluster members, which would add relevant insights.

  Action: In the revised manuscript, we will include a spatial analysis and maps illustrating:
  - The spatial distribution of clusters, and
  - The sensitivity of dikes based on head decimate heights.

  These findings will be integrated into the relevant sections—Section 4.2.1 (clusters) and Section 4.3.1 (decimate height sensitivity). To maintain readability, the maps will be placed in the appendix.

- *The meaning of the clusters determined from the statistical analyses remains a unclear - what do they tell us, and how do these contribute to the overall objective, i.e. predicting dike stability over large areas, under conditions of extreme precipitation. I got the impression that a different type of clustering would have added more insight.*

  The clusters in this study group dikes based on similarities in head peak responses to the same weather events. This was determined by analyzing the coincidence of simulated head peaks over 30 years of rainfall and evaporation (Lines 388–389). These clusters represent variations in head response across the dike system, which is relevant for regional risk assessments. Since extreme loads may result from different weather events in different clusters, this variation affects system reliability and has direct implications for the flood risks in polders, as discussed in the introduction (Lines 56–57) and Section 5.2 (Lines 532–544). Additionally, these clusters with different head responses can help identify potential instabilities based on predicted rainfall. Therefore, it would be even more useful to understand how dikes with specific characteristics respond to heavy rainfall or which types of dikes exhibit certain head response patterns. Both aspects are examined in Section 4.2.2, where we analyzed potential relationships between physical dike characteristics, dike clusters, and impulse response function parameters. The lack of clear relationships suggests that general attributes such as soil type or geometry do not directly determine head response behaviour. This is an important finding, as it indicates that commonly used physical characteristics cannot be linked to the head response and it is likely influenced by more local dike properties, for example local subsurface . This is also clearly stated in the conclusions.

  Regarding the reviewer's comment about different type of clustering, we are unsure what specific approach is being suggested. However, we believe this may relate to the detailed comment on Lines 586–587, which we address separately further below.

  Action: We will revise Section 4.2 to clarify the meaning of the clusters and how they contribute to the overall objective of predicting dike stability under extreme precipitation. Additionally, we will refine related discussions in other sections based on the reviewer's detailed comments.

- *The discussion section should me restructured and complemented with thoughts about climate change impacts, and implications for the dikes, and their management.*

  This study quantifies the impact of intensified rainfall and drier summer periods—one of the key impacts of climate change—on peak head levels, one of the key elements for dike stability. However, it assumes that the head response remains unchanged over time. The potential impact of non-stationary responses, such as variations in hydraulic conductivity due to repeated dry-wet cycles, is not explicitly addressed, which may be what the reviewer is referring to. Other climate change impacts on dike stability and management, like deterioration processes and other failure mechanisms e.g. horizontal translation, are also relevant but are not currently discussed in detail. While Section 5.2 already addresses some management implications, we recognize that the reviewer is requesting a more comprehensive discussion.

Action: We will restructure the discussion section to improve clarity and explicitly include additional climate change impacts. Furthermore, we will expand on the implications for dike safety and management to provide a more detailed perspective.

- *The conclusions are too much a summary: rewrite these in 'conclusive' style, not a story of what you did.*

  We acknowledge the reviewer's feedback regarding the conclusion section. While we aimed to summarize the key findings, we understand the need for a more conclusive style that emphasizes the main insights rather than a summary of actions.

  Action: We will revise the conclusions to present the findings in a clearer and more conclusive style, focusing on their significance rather than repeating the steps taken in the study.

- *At several points the text is wordy, or unclear - see specific comments.*

  We appreciate the reviewer's feedback on clarity and conciseness.

  Action: We will critically review the manuscript, using the specific comments as guidance, to improve clarity and make the text more concise. Our revisions will focus on eliminating unnecessary wording while ensuring that the key messages remain clear.

In the attached pdf I have put my detailed comments.

We appreciate the detailed comments, of which many are focusing on terminology and clarity. The comments are incorporated in the revised manuscript to make the text more readable. We would like to highlight a few detailed comments and give a reaction:

- L586-587: *"so, this makes me doubting whether your clustering resulted in something that contributed to reaching your objective, instead of a statistical of sections by means of numerous characteristics as such."*
  We have no doubts about the validity of the clustering approach and its contribution to the study's objectives. Our goal is to assess the dynamics of peak hydraulic heads in canal dikes at a national scale, specifically in response to heavy rainfall events, by analyzing variations in head responses and head statistics. The clustering was designed to identify groups of dikes with similar head response behavior, making it a valuable tool for regional risk assessments. This approach helps to understand the variation in loading conditions and distinguish which dikes may experience similar loading conditions under extreme weather events (see also our response in the major revision). Therefore, we examined the relationship between dike characteristics and head response in Section 4.2.2 to determine whether physical attributes could explain (1) the clustering and (2) the impulse response function parameters. While no clear relationships were found, this is an interesting finding in itself, as it suggests that general dike characteristics alone may not be sufficient to predict head response behavior.

- L416-*423: "part if this is method, furthermore, I prefer reading about what causes behaviour, insead of listing all sorts of tests and p values - these distract here; instead describe which factors show (functions / causal) relationshops (underpin that with test results), that is interesting to the reader"*
  We appreciate this suggestion and understand the preference for focusing on causal relationships rather than listing statistical tests and p-values. While these tests are important

for validating our findings, we agree that emphasizing the key factors driving head response behavior will improve readability and clarity.

Action: We will revise this section to shift the focus toward explaining which factors show functional or causal relationships with head response behavior. The statistical results will be used to support these insights rather than being presented as a list. We will move the description of various tests to the methodology section, where we make a new subsection named “3.3. Statistical tests for relationships”.

- Line 449: *“it would be interesting to know how much height variation is still acceptable and 'safe' is that 10 cm, 50 cm, or 2 cm?”*
  This is indeed an interesting and important question. However, determining the exact height variation that remains acceptable and "safe" is complex and depends on many factor regarding the strength of dikes. While this falls outside the scope of the current study, it is a key focus of our follow-up research!

Response to reviewers' comments

"The dynamics of peak head responses at Dutch canal dikes and the impact of climate change" https://doi.org/10.5194/egusphere-2024-1495 submitted to Natural Hazards and Earth System Sciences

We thank the reviewer for her/his thorough, insightful and valuable feedback, both on a general and more detailed level.

Below, we reply to the reviewer's comments and explain how we will address them. The reviewer's comments are shown in *Italicized text in gray*, our responses are shown in blue. We provide detailed responses to the major comments, along with specific **actions** to improve the manuscript. For minor comments, we offer brief responses, as we will incorporate these suggestions to enhance clarity and refine terminology throughout the text.

Anonymous Referee #2

This manuscript outlines an application of time series models using impulse response functions to model the hydraulic heads observed with dike systems in the Netherlands. Different model structures are tested to simulate the heads, with a nonlinear-threshold model (TARSO) found to perform the best. This is an interesting result, that could teach us something about how the heads in dike systems respond to precipitation and potential evaporation. The study attempts to relate model characteristics to various physical characteristics of the dike systems, with moderate success. I generally found the manuscript well written and the figure quality appropriate. The topic fits the scope of the journal. I have a couple of major comments that should be addressed, and some minor technical comments at the bottom.

Thank you for your constructive feedback; we appreciate your insights and will carefully address your comments.

- *One thing I was missing in the manuscript is an explanation and interpretation of why the threshold nonlinear model structure (TARSO) works best for the hydraulic heads in Dikes in the Netherlands. This is a surprising outcome to me, that deserves more thought and might be informative for future attempts to model the heads in dikes. This model was designed for a different type of system (groundwater levels in polders, influenced by ditches falling dry and being activated). Perhaps there is topping-off of the heads in dikes. These types of models are commonly used to gain understanding of how groundwater systems function, and why. A discussion of this type is currently missing from the manuscript but would be a welcome addition.*

  The strong performance of the TARSO model in modeling hydraulic heads in Dutch canal dikes can be explained by the non-linear characteristics of the head response in these dikes. The manuscript gives some suggestions in line 330 – 332: "*This non-linear behaviour can be the result of various soil layers in the dike body, each with distinct hydraulic properties, and changes in infiltration rates or nonconstant storage capacities of the unsaturated zone during the dry season*" The suggestion that there may be a "topping-off" effect in dike heads is an interesting perspective that aligns with the need to account for non-linear responses.

Action: We will expand on this discussion in the revised manuscript, providing a more extensive interpretation of why the TARSO model performs well for dikes. However, as these explanations remain hypotheses at this stage, we will also emphasize that further research is needed to verify these mechanisms.

- *Looking at Figure 6, I was very surprised by the simulated behavior of the FlexModel given that all models share the same input and not be that far off from each other. I made a quick script to model the data using the Pastas default values to better understand the result but got an average of R2=0.68 for the FlexModel, much higher than the reported value of 0.32. I suspect some suboptimal choices were made for that model. Perhaps the Authors can revisit the scripts and double-check this, or explain this result in more detail.*

  Thank you for this remark. It is unclear how the reviewer obtained an average R^2 of 0.68 for the FlexModel. When we apply the FlexModel using default values, we obtain a significantly lower average R^2. We have experimented with different initial parameter settings but were unable to achieve a substantial improvement in model performance.

  If the reviewer is willing to share their script, this would be very helpful for comparison and to better understand the differences in our findings. Additionally, this relates to the subsequent comment regarding the reproducibility of results, which we address separately.

- *I assessed the manuscript for its reproducibility. I appreciate the authors providing the original head, precipitation, and evaporation data is provided. This data provides a unique dataset on head measurement in dike systems, which might be worth highlighting in the manuscript. I note here that none of the data underlying the results shown in the figures and tables are shared, nor are the scripts that lead to the results. This makes it difficult to verify the results and/or build upon this work. I would recommend the authors to share the scripts and output data on a FAIR repository to improve the reproducibility of this study.*

  Thank you for assessing the reproducibility of our study and for highlighting the uniqueness of the dataset. We appreciate the importance of making research more transparent and reproducible.

  Action: We will emphasize the uniqueness of the dataset in the manuscript. Additionally, we will share the scripts and output data on 4TU.ResearchData, ensuring alignment with FAIR principles to improve the reproducibility and accessibility of our work.

Minor technical suggestions:

L84: Potential evaporation Correct; will be revised

L195: Figure 4 and 5 appear to be the same. I am not sure if another figure is meant to be here. Otherwise Figure 5 can be removed. Correct; we will remove Fig. 5.

L192: How well does GeoTop actually work for dikes? I can imagine that these are not related at all, given that dikes are built by humans with specific materials. This may influence the results later on relating outcomes to the soil types, i.e., would the relationship with the soil type improve if the GeoTop data is left out? Some consideration about this would be good here.

This is a good suggestion; We will check this and write down the findings in the revised paper.

L238: I think Sm is substituted by R, not the other way around. Correct; will be revised

L276: simulate "the heads". Good suggestion, will be revised

L288: How was this threshold of 0.7 determined? How sensitive are the result to changing this threshold. We will elaborate on the sensitivity of this threshold.

L302: Analyses Will be revised

L303: every "…"? We will rewrite this sentence: "of every" will be deleted.

L327: r2 was previously referred to as R2, check throughout Will be checked and corrected in the manuscript

L329: The FlexModel is a nonlinear recharge model, the other three models are not. I think the TARSO model is meant here, which still computes recharge using a linear equation.

In literature, tha TARSO-model is often referred to as a nonlinear model, since it accounts to some extent for the nonconstant relationship between precipitation excess and water table depth caused, in contrast to linear model (Knotters en Gooijer, 1999)

L342: replace by "in the summer of 2019" Will be revised

L343: disturbances Will be revised

L345: scatter plot in Figure XX. Sentence will be rewritten.

L366: I don't understand what "peak block" is, please clarify. This is explained in lines 347-349.

L496: Apparently, the head time series were filtered using some reliability criteria in this study. This should be mentioned in the section describing the data. What reliability criteria were used?

This is discussed in section 3.1.3 – Model calibration and selection. Lines 287-294 describes the reliability criteria used

L501: Remove "explicitly". Uncertainty was not considered, as I understood from the manuscript.

We believe you mean line 511: we will remove "explicitly".

Knotters, M., & De Gooijer, J. G. (1999). TARSO modeling of water table depths. Water Resources Research, 35(3), 695-705.