Dear Dr. Lelys Bravo de Guenni,

We are pleased to submit the revised version of our manuscript, "*Climate Adaptation-Aware Flood Prediction for Coastal Cities Using Deep Learning*" (ID: EGUSPHERE-2025-838), for your reconsideration for publication in *Hydrology and Earth System Sciences*.

We would like to extend our sincere gratitude to you, the original reviewers, and the new reviewer for providing additional valuable feedback. We have carefully addressed all the remaining concerns and are confident that these revisions have further strengthened the manuscript's clarity, rigor, and practical significance.

We gave special consideration to the new reviewer's insightful comment regarding the significance of our work in the context of long-term planning. To address this, we have revised the Introduction and Discussion section to more clearly articulate the critical role of our model. Specifically, we highlight that the primary use-case of the proposed model is not simply to simulate one future scenario under a given SLR value, but to allow coastal engineers to evaluate the costs and benefits of thousands of potential protection configurations across multiple plausible SLR scenarios to identify an optimal investment strategy or to perform a sensitivity analysis (tasks that would not only be computationally daunting but also resource-costly to carry out with traditional hydrodynamic models, even for few thousand scenarios).

In addition to this crucial clarification, and in response to the other excellent points raised, we have also:

- Provided a new quantitative analysis of our model performance on data imbalance by adding confusion matrices and a dedicated results subsection.
- Added practical, actionable guidance for our uncertainty quantification, establishing an empirical threshold for planners to identify high-risk areas.
- Further improved the clarity of our architecture figure by enhancing its caption to serve as a descriptive legend, better connecting it to the detailed supplementary figure.

Following this letter, you will find our detailed, point-by-point responses to all comments from this review round. To facilitate an efficient review, all new changes made in this second round of revisions have been highlighted in blue throughout the manuscript and supplementary material.

We are confident that the manuscript is now much stronger and more robustly justified. We hope that these extensive revisions are satisfactory and that the manuscript is now suitable for publication.

Thank you once again for your time and for reconsidering our work.

Sincerely,

Dr. Bilal Hassan

(on behalf of all authors)

Anonymous Referee # 3

The authors developed a machine learning model designed to rapidly predict coastal flooding under varying sea-level rise (SLR) projections and shoreline adaptation scenarios, positioning it as a potential surrogate for traditional hydrodynamic models. Overall, the manuscript is well-structured and sufficiently clear for readers. The authors have also addressed the comments from two reviewers during the first round of review. However, my primary concern pertains to the significance of this study.

While the machine learning model offers the advantage of high computational efficiency compared to hydrodynamic models, this efficiency is typically most beneficial for short-term predictions, such as those related to storm surges induced by tropical cyclones. The focus of this study, however, is primarily on assessing the impacts of future SLR scenarios, which are generally characterized by slow processes that do not necessitate extremely high computational efficiency. Consequently, I believe that the significance and necessity of this research are considerably diminished. The value of this research would be greatly enhanced if it could be demonstrated that the machine learning model is also applicable for short-term scenario predictions. Conversely, for future SLR scenarios, the accuracy of the predictions becomes paramount. However, it is important to note that the accuracy of the machine learning model is inherently limited by the accuracy of the hydrodynamic model from which it was trained. Therefore, the accuracy of the hydrodynamic model directly influences the performance of the machine learning model. As such, the significance of this study should be clearly justified and thoroughly discussed before its publication.

Response:

We thank the reviewer for this thoughtful question, which illuminated a shortcoming of our presentation, and the helpful suggestion which pointed us to additional use-cases of our model. For clarity, our response below is organised into two parts.

(i) On the significance of this research:

We agree that a clear and thorough justification is essential, and we appreciate the opportunity to elaborate on this matter and improve our writing. The reviewer is correct in pointing out that this work's focus is on the climate adaptation-aware flood prediction (as also reflected in the paper's title), which is driven by slow-evolving processes and assumes a long-term planning horizon. However, for coastal planners, the critical task is not simply to simulate one future scenario under a given SLR value, but to evaluate the costs and benefits of thousands of potential protection configurations across multiple plausible SLR scenarios to identify an optimal investment strategy or to perform a sensitivity analysis. In this context, the usability of physics-based high fidelity simulators is highly limited, not only due to their extremely slow running time but also due to their high computing resource requirements (these tools need to be run on HPC clusters). To illustrate, consider the following example. For the San Francisco bay area, the number of candidate coastal segments for protection (OLUs) is 30, hence there are over a billion possible protection combinations. To test even a mere 10,000 of these protection scenarios for a given SLR value with a traditional hydrodynamic model would be computationally daunting, requiring over 5 years of continuous computation on an HPC cluster (approximately 0.85 to 1.41 million CPU-hours). This prohibitive barrier is what currently prevents comprehensive, data-driven resilience planning.

To address the reviewer’s concern, we have expanded the Introduction section with a clarification that while our focus is on long-term SLR scenarios, the core challenge we address is not the simulation of a single slow process, but rather the exploration of the vast combinatorial design space of potential, long-term adaptation strategies. Also, the “Computational Efficiency Analysis” section has been revised to include the actual CPU-hours of the hydrodynamic simulations as well as the hardware details of HPC nodes on which the simulations were run. Lastly, we have revised the Discussion section to clarify that this research enables a two-tiered decision-making workflow. Our high-efficiency surrogate model serves as a rapid, cost-effective scenario-assessment tool to explore this vast design space in hours, not decades, allowing planners to identify a small shortlist of the most promising and effective protection strategies. These top candidates can then be subjected to rigorous validation using the precise, but slow, physics-based simulators.

We also explicitly address the reviewer's valid point regarding the dependency of machine learning models on the hydrodynamic model's accuracy. In the revised Discussion section, we acknowledge that our surrogate's accuracy is indeed bounded by its training data. However, we clarify that its role is to serve as a high-fidelity proxy for broad-scale exploration, not to be a perfect replacement. Our extensive validation, which demonstrates high accuracy and strong spatial fidelity (DSC scores), confirms that CASPIAN-v2 is a reliable proxy for this purpose, capable of correctly ranking the performance of different strategies. This synergy, using the fast surrogate for exploration and the precise physics-based model for targeted validation, is what establishes the necessity of our work.

(ii) Applicability to short-term flood prediction:

We appreciate the reviewer’s suggestion and completely agree that the proposed model’s applicability to short-term flood prediction would significantly increase the contribution of this paper. As the paper demonstrates, with fine-tuning the proposed model was able to generalise to different SLR values (e.g., 0.5m or 1.5m). Likewise, given the availability of training data, we could demonstrate that CASPIAN-v2 can generalize to SLR of 0 meters, which would enable its application to prediction of short-term coastal flooding. However, the selected two study areas do not experience storm surges induced by tropical cyclones, and thus simulations at 0 meter of SLR would result-in no coastal inundation, that is prediction of all 0s in the inland locations regardless of a protection scenario. Therefore, for the current setting, illustrating the generalization of CASPIAN-v2 to 0 meter of SLR will not add any meaningful insights to the paper. However, to incorporate the reviewer’s suggestion, a somewhat more compact version of the above response has been appended to the Discussion section.

Anonymous Referee # 2

The authors have made substantial improvements to the manuscript and have adequately addressed the majority of concerns raised in the first round. The computational efficiency analysis and uncertainty quantification represent significant enhancements. The following minor issues should be addressed for final acceptance.

Response:
We sincerely thank the reviewer for their positive assessment of our revisions and for acknowledging the significant enhancements made to the manuscript. We appreciate their guidance and have diligently addressed the remaining minor issues as detailed below.

Specific Minor Revision Comments

1. Data Imbalance Performance Analysis
While the authors explain their theoretical approach to handling data imbalance, they should provide a brief quantitative breakdown showing model performance specifically on flooded vs non-flooded pixels. A simple confusion matrix or performance metrics stratified by flood/no-flood would suffice.

Response:
We thank the reviewer for this excellent suggestion. We agree that providing a quantitative breakdown of the model's performance on the imbalanced classes is crucial for validating our approach. To address this, we have added a new subsection to the Results section titled "Performance Analysis under Data Imbalance." This new section presents and interprets a new figure (Figure 5), which contains the normalized confusion matrices for the AD, SF, and combined test sets. The analysis explicitly highlights the model's high accuracy for both the majority (non-flooded) and, more critically, the minority (flooded) classes, including its extremely low false negative and false positive rates.

This new quantitative evidence provides direct proof that our strategy for handling data imbalance is effective. We are confident this addition fully addresses the reviewer's concern. The new section and figure have been highlighted in the manuscript.

2. Uncertainty Quantification Practical Guidance
The deep ensemble approach is well-implemented, but the authors should add 1-2 sentences providing practical guidance on uncertainty threshold interpretation for coastal planners (e.g., "uncertainty values above X suggest areas requiring additional hydrodynamic validation").

Response:
We thank the reviewer for this excellent suggestion to make our uncertainty analysis more actionable for practitioners. To provide the concrete guidance requested, we have performed a new empirical analysis to derive a specific uncertainty threshold that is directly linked to our critical error metric ($\delta > 0.5$).

For all scenarios in our test set, we first identified the specific pixel locations where the prediction error exceeded 0.5 meters. We then calculated the average uncertainty value associated with these "critical failure" locations. This analysis allowed us to establish that a normalized uncertainty value of approximately 0.75 or greater serves as a strong indicator that a prediction may exceed the 0.5m error threshold. We have added this finding and its practical interpretation to the Discussion section, recommending that planners use this 0.75 threshold to flag high-risk zones that require further investigation or more conservative safety margins. This provides the direct, data-driven, and actionable guidance that the reviewer rightly suggested. The new text has been highlighted in the manuscript.

3. Figure 4 Clarity
The simplified architecture diagram may be too abstract. Consider adding labels for the novel components (MARX, SEE blocks) or a brief legend to help readers connect the simplified version to the detailed supplementary figure.

Response:

We thank the reviewer for this helpful feedback. We agree that the simplified schematic in Figure 4 would benefit from a more explicit explanation to better connect it to the detailed version in the supplement. While the figure already contained the literal labels (e.g., "MARX Blocks"), we understood the reviewer's feedback to be about adding conceptual labels, that is, explaining the function and purpose of these components. To address this, we have revised the caption for Figure 4 to serve as a descriptive legend. The revised caption now explains the role of each block (FE, MARX, SEE, and FR), clarifying the purpose of our novel components directly alongside the figure.

Furthermore, to create the explicit link the reviewer requested, the caption now concludes with a direct reference pointing the reader to the detailed, layer-by-layer schematic in the Supplementary Material (Figure S5).

We believe this revised caption now acts as the necessary bridge between the high-level schematic and the detailed architecture, making the figure significantly clearer and more informative, as the reviewer intended. The updated caption has been highlighted in the manuscript.

4. Model Limitations Discussion
Add a brief paragraph discussing model limitations, particularly regarding generalizability to coastal environments beyond the two study areas and data requirements for new regions.

Response:
We thank the reviewer for this important suggestion to clarify the limitations of the model regarding broader generalizability and data requirements. To address this, we have expanded the limitation paragraph within the Discussion and Conclusion section. The revised text now explicitly states

that applying CASPIAN-v2 to a new coastal environment would require generating a dedicated dataset of hydrodynamic simulations for that specific location.

Furthermore, as a pathway to address this data dependency, we have added a discussion on future work that could explore advanced techniques like few-shot learning. We suggest that these methods could potentially allow the model to be adapted to new regions with a drastically reduced number of simulations, significantly enhancing its practical utility.

We believe this addition provides a more transparent discussion of the model's current limitations and offers a clear, forward-looking perspective on how to address them. The revised paragraph has been highlighted in the manuscript.

5. Mathematical Notation Verification
Ensure the standardized notation described in the response letter is actually implemented consistently throughout the supplementary material.

Response:
We thank the reviewer for this follow-up. We confirm that we have performed a thorough check and have implemented the standardized notation consistently throughout the entire manuscript and all sections of the supplementary material. All equations now adhere to the clear and consistent format.

These targeted revisions will enhance clarity and practical applicability without requiring fundamental changes to the methodology or results.

Response:
We thank the reviewer for their summary. We agree that these final targeted revisions have enhanced the clarity and practical applicability of our work, and we appreciate the opportunity to make these final improvements.