

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 thanks to you and to both expert reviewers for your time and for providing such thorough and constructive feedback. We found the comments to be incredibly valuable and agree that addressing them has significantly improved the quality, clarity, and impact of our work.

We appreciate your summary of the required revisions and are confident that we have diligently addressed all points raised. Specifically, in this revised manuscript, we have:

- Clarified the rationale for our study's focus on tidal flooding within the broader context of coastal hazards.
- Substantially revised the presentation of the model architecture, providing a simplified schematic in the main text and moving detailed explanations to the supplementary material for improved clarity.
- Added a new, dedicated section with a rigorous computational efficiency analysis, quantitatively demonstrating the significant time savings of our model compared to the hydrodynamic simulator.
- Incorporated a comprehensive uncertainty quantification using a deep ensemble method, including new figures and discussion on the practical implications for risk assessment.
- Expanded our discussion on how our framework's design, particularly the loss function and attention mechanism, specifically addresses the critical challenge of data imbalance.

Following this letter, you will find our detailed point-by-point response to each comment from the reviewers. To facilitate an efficient review process, all changes in the revised manuscript and the supplementary material have been color-coded: revisions made in response to Reviewer 1 are highlighted in blue, and those for Reviewer 2 are highlighted in red.

We are confident that the manuscript is now much stronger, clearer, and more robust. We hope that the extensive revisions are to your and the reviewers' satisfaction 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 # 1

The manuscript proposes a novel deep learning framework for predicting coastal flooding (only tidal). The framework has been applied to the urban areas of Abu Dhabi and San Francisco. The proposed model performs better in comparison to the existing DL methods. The manuscript is free from grammatical errors and is interesting to read. However, the following serious concerns need to be addressed,

Response:

We thank the reviewer for their positive and encouraging assessment of our manuscript. We are grateful for the feedback on the paper's novelty and clarity, and we sincerely appreciate the time and effort taken to identify areas for improvement.

We have carefully considered all of the serious concerns raised and have thoroughly revised the manuscript to address each point. In the following sections, we provide a detailed, point-by-point response describing the specific changes we have made and where they can be found in the revised text. We believe these revisions have significantly strengthened the manuscript.

The developed model is only tuned for tidal flooding, and it does not account for the storm surge, which is a major source of coastal flooding. Especially, San Francisco is vulnerable to flooding by hurricanes. The paper doesn't mention the idea behind the omission of flooding due to storm surge, and why modelling only tidal flooding is crucial in these example cases. The motivation for setting up this kind of prediction system is very unclear.

Response:

We thank the reviewer for the comment. We acknowledge that the San Francisco area does indeed experience storm surges especially along the Pacific coastline and to a lesser extent within San Francisco Bay. In the Supplementary materials section S1, line 16 of the original manuscript, we justified not using a storm surge model for San Francisco Bay due to significant wave heights typically of about 0.07-0.2 m, an order of magnitude smaller than the 2.0-3.0m found on the California Coast outside San Francisco Bay. We have updated Section 2.1.2 in our revised manuscript to clarify our focus area which is San Francisco Bay and its shoreline urban communities, as opposed to the coastal communities facing the Pacific Ocean. We have focused here because this is an area with a larger urban infrastructure such as highways that could suffer flood induced interruptions that affect the entire Bay Area transport network. Additionally, independent of storms, the addition of coastal protections (in the form of sea walls or levees) to one part of the Bay may affect other parts of the Bay; Hollemann and Stacey, 2014 showed that protecting the southern portion of the Bay may increase maximum water levels by 0.2 in the north of the Bay; Wang et al., 2018 showed that protection of the southern portions of the West Bay adversely affects the water levels of the East Bay, and vice-versa.

Meanwhile, the Abu Dhabi coastline experiences storm surges that are caused by sustained northwesterly winds that generally occur in the winter (called the Shamal winds). They are incorporated into the model by the coupling of Delft3d with SWAN, a spectral wave model, and storm runup calculations were performed along the Abu Dhabi coastline in the same manner described in Chow and Sun, 2022. Details of the Abu Dhabi hydrodynamic models were

provided in the Supplementary material, but the detailed descriptions of the Abu Dhabi Delft3d and SWAN coupled models have been added into Section 2.2 of the main body of the manuscript.

Also, what is the level of damage that can occur due to tidal flooding (as the storm surge is not included here) needs to be discussed so that the readers can understand the significance or the need for this framework.

Response:

Thank you for the question. We have included in Section 2.1.2 in our revised manuscript that the motivations of studying the changes to tidal flooding within San Francisco Bay due to different protection scenarios are based on protection observations from previous studies that adding coastal protections to one part of the Bay may affect other parts of the Bay (Holleman and Stacey, 2014; Wang et al., 2018).

Meanwhile storm surges and tidal flooding have both been modeled for the Abu Dhabi coastline since there are effects both from storm surges with the presence of sustained onshore winds and tidal interactions between the multiple mangrove islands in the Abu Dhabi area (as shown in Figure 2).

The manuscript needs to contain the details of how the hydrodynamic model, whose results have been used for training the DL model, is validated. What kind of events were used? How closely they matched the observed flooding also needs to be included in the paper.

Response:

Thank you for the comment. Previous work has performed validation of the hydrodynamic models used in this paper with the current tidal gauges so that the observed water levels (without sea level rise) are well reproduced by the simulation – see Wang et al., 2017 and Chow and Sun, 2022. The simulations were performed with a 0.5m SLR rise which reflects a possible future scenario for San Francisco Bay in the year (somewhere between 2050-2100 depending on the climate change scenario pathway (between SSP2-4.5 and SSP5-8.5) from IPCC AR6 report (referenced in Section 3.1.1 of the Supplementary Materials). As such the simulated flooding without SLR serves as a predicted flooding extent that is not currently observed. We have added some details addressing this in Section 2.2 of the revised manuscript.

The paper argues that the developed model is computationally efficient, but it fails to elucidate the details on how fast it is in comparison to the hydrodynamic model. This is very crucial to establish the significance of this work.

Response:

We thank the reviewer for pointing out this critical omission. We agree that elucidating the specific details of the computational speed-up is crucial to establishing the significance of our work.

To address this, we have added a new subsection to our Results titled "Computational Efficiency Analysis (Section 5.1.4)," which is dedicated to this topic. This section provides a direct and detailed comparison of CASPIAN-v2's speed against the hydrodynamic model. We now

explicitly state the inference time per scenario for our model (0.22s) versus the simulation time for Delft3D (14-19 hours).

Furthermore, to provide a powerful practical example, we calculate that evaluating our entire 72-scenario test set would take over 49 days with the hydrodynamic simulator, whereas CASPIAN-v2 accomplishes this in just 16 seconds. This represents a significant computational time reduction. We have also updated Table 5 to include the raw data (training time, inference time, model parameters) that supports this analysis.

We believe this new, detailed analysis firmly establishes the significance of our work by clearly quantifying the dramatic efficiency gains. The new subsection and updated table have been highlighted in the manuscript.

There is no spatial validation of the over-/under-/correctly predicted flooded areas. For example, see Nithila Devi et al. (2024), where a spatial fitness index has been used to assess the spatial accuracy.

Response:

We thank the reviewer for the suggestion and for referencing Nithila Devi et al. (2024). We agree that spatial validation adds important insight into model performance. In Section 2.2 the revised manuscript, we have referenced past tidal gauge validation performed for the San Francisco model (Wang et al., 2017) and the Abu Dhabi model (Chow and Sun, 2022) for vertical water level validation with observations. While no flooding extent has been observed in the case of zero sea level rise, our model attempts to predict the extent of flooding in a future scenario of 0.5m sea level rise.

In San Francisco, there is also riverine and pluvial flooding present. How only modelling the tidal flooding sufficient in this case for assessing the flood protection capabilities?

Response:

Thank you for the comment. Although our San Francisco model does include riverine input from the Sacramento and San Joaquin Rivers, the inflow rates into the Bay were baseline values rather than for extreme fluvial flood events. While we acknowledge that incorporating more refined hydrodynamic forcing conditions to include pluvial and riverine floods, as well as extreme storm events, can refine the hydrodynamic model to reflect more extreme flooding, our overall scope in this paper is in the use of machine learning to be able to act as a surrogate for a hydrodynamic model running at a more average condition but under a Sea Level Rise of 0.5m and in the presence of a coastal protection over many different stretches of the shoreline. Our focus on tidal flooding within San Francisco Bay was to highlight the unique tidal behavior within the Bay where the construction of sea walls along certain portions of the shoreline Bay may in fact exacerbate the sea level within the Bay to increase by up to 1 m (Holleman and Stacey, 2014). We have added some details of this discussion in Section 2.2 of the revised manuscript.

Flood protection measures on the coast are very vaguely discussed. It is very difficult to understand what they are, how they function, and how they are incorporated into the

hydrodynamic and DL models. It will be nice to discuss with figures showing their placement, functionality and model representation explicitly.

Response:

Thank you for the comment. In the revised manuscript, we have included more information originally contained in Supplemental Materials into the main body of the paper, namely: The revised Section 2.1.2 (old Sections 3.1.2) for San Francisco, which references previous work that provides more details of the coastal protection assumptions; and revised Section 2.1.1 (old Section 3.1.1) that describes the flood protection measures taken for Abu Dhabi, which also references previous work that provide more details.

Regarding the writing style, I find that the paper is more from a computer science background, rather than explaining clearly the application and relevance to the hydraulic modelling application. For example, the manuscript lacks details on training data used, accuracy of the training data, spatial prediction accuracy, specifics of protection measures, etc., as mentioned in the earlier comments.

Response:

We appreciate the observation of the reviewer regarding the writing style and the need to better align the manuscript with the expectations of the hydraulic modeling community. In the revised version, we have moved some of the material presented in the Supplementary Materials section into the main body of the manuscript, and to better highlight the application relevance and hydrodynamic context of our work (with some of the model development and validation material taken from Wang et al., 2018 and Chow and Sun, 2022). Finally, we have made editorial changes to our explanation of the training data (including its source and accuracy), provided further clarification on spatial prediction accuracy, and included additional details on the definitions of the Operational Landscape Units (OLUs) and the protection scenarios of OLU that were used. This will ensure that the manuscript clearly communicates both the modeling and application aspects of the study.

The paper is very wordy with several irrelevant information and lacks crucial information necessary for the understanding of the readers. For example, information on kind of damages due to tidal flooding, validation etc. Please discuss the important information briefly rather than just citing other works.

Response:

Thank you for the feedback. In the revised manuscript, we have reduced wordiness and removed less relevant content. We have also briefly included key information, such as tidal flood impacts and data validation, to improve clarity and minimize reliance on external citations.

The following are the specific comments listed for each section,

Introduction and Related Works

The introduction is very wordy and unorganised. A separate section of “related works” is unwarranted and has a lot of redundant information. Therefore, sections 1 and 2 (introduction

and related works) need to be shortened and merged into one, ensuring that there is a proper flow of content in them.

Response:

We thank the reviewer for this constructive feedback regarding the structure of the manuscript. We agree that the separate Related Works section contained redundant information and that a single, integrated introduction would provide a more cohesive narrative for the reader.

To address this, we have performed a careful merge, eliminating the Related Works section entirely and strategically integrating the most essential literature context into the main Introduction. The revised introduction now seamlessly incorporates a more detailed overview of the state-of-the-art, summarizing the progression from traditional ML models to hybrid systems and distinguishing between 1D and 2D DL approaches. This enhanced context allows us to more clearly situate our work and sharpen the problem statement by highlighting that many existing models overlook the combined, long-term impacts of SLR and dynamic shoreline adaptations.

We are confident that this new structure is more organized, concise, and effective. The revised Introduction has been highlighted in the manuscript.

Lines 17-20 can be shortened.

Response:

We thank the reviewer for this suggestion. We agree that the original sentence was unnecessarily long due to the extensive list of examples and the detailed parenthetical definition.

As recommended, we have revised this sentence to be much more concise. We have removed the long definition and trimmed the list to three key, representative examples to illustrate the point without disrupting the flow of the introduction. We believe the revised sentence is now more direct and readable, as intended. The change has been highlighted on lines 17 to 19 in the manuscript.

Lines 26 – 34 need to be shortened, retaining information relevant to the kind of study the paper deals with.

Response:

We thank the reviewer for this suggestion. We agree that the original paragraph was overly detailed with its examples of coastal fortifications.

As part of our broader effort to streamline the Introduction section (in line with the reviewer's previous comments), we have significantly shortened this specific passage. The revised text now establishes the key point that cities are actively armoring their shorelines, without the extensive list of specific locations and project details. For instance, we removed the detailed breakdown of examples from San Francisco and New Orleans and streamlined the New York City case study to its core message.

We are confident the revised paragraph (lines 29 to 33) is now more concise and focused on the information directly relevant to our study.

Please use terms like computationally expensive/intensive instead of the term “computationally prohibitive” if possible.

Response:

We thank the reviewer for this important point regarding terminology. We agree that "computationally expensive" is a more appropriate and precise term in this context. As recommended, we have replaced "computationally prohibitive" with "computationally expensive" throughout the text, specifically in the abstract and introduction. The revisions have been highlighted in the manuscript for convenience.

Mention literature related to computationally efficient modelling approaches (subgrid, parallelisation, simplified models, etc.), such as De Almeida et al. (2013), Neal et al. (2012), Li and Hodges (2019), Sanders and Schubert (2019), Nithila Devi et al. (2024), etc. Discuss such methodologies and bring out the importance of DL/ML/AI in flood forecasting and modelling.

Response:

Thank you for this helpful recommendation regarding various methods of employing spatially nested models. In the revised manuscript, we have revised the introduction and Section 2.2 to include a brief discussion of computationally efficient hydrodynamic modeling approaches, as suggested by the reviewer. We wish to note that nesting has not been necessary for the DELFT3D portion of the model, as the model grid has been downscaled to about 30 m horizontal resolution for areas of interest such as the city of Abu Dhabi; and that the model is capable of modeling dry and wet grid cells. But we did employ a nesting method to incorporate the SWAN to model storm surges and waves for Abu Dhabi. We have highlighted the limitations of traditional approaches and clearly motivate the role and advantages of AI-based methods in improving computational efficiency and scalability for flood forecasting and modeling.

Highlight the need for high-resolution modelling, especially in the complex urban terrain and cite relevant literature.

Response:

We thank the reviewer for this crucial suggestion to more explicitly motivate our work. We agree that the need for high-resolution modeling in urban areas was not sufficiently justified in the original text.

To address this, we have added a new sentence to the third paragraph of the Introduction (lines 39 to 42). This text now explicitly highlights that flood dynamics in complex urban terrain are critically dependent on fine-scale features like individual buildings and road elevations. We state that high-resolution modeling is therefore a necessity, not just a preference, for accurately predicting inundation and its impact on critical infrastructure. As requested, we have supported this statement with relevant citations from the literature.

We believe this addition better contextualizes our study and clarifies the motivation behind our choice to develop a high-resolution deep learning framework.

Line 56, what do you mean by high dimensionality of outputs?

Response:

We thank the reviewer for requesting this clarification, and we agree that the term required a clear definition. We have revised the text in the Introduction to provide a concise, parenthetical explanation of this concept.

The text on line 57 now clarifies that "high dimensionality of outputs" refers to "the challenge of predicting an inundation value for every pixel in a large spatial grid." We have placed this definition where the term is introduced as a specific challenge for deep learning models. This logically follows the earlier discussion where we establish why high-resolution modeling is a scientific necessity for urban flood prediction.

We believe this approach both clarifies the term and strengthens the overall logical flow of the introduction. The revision has been highlighted in the manuscript.

Since this paper expands on a deep-vision based framework, explain briefly about this in the introduction so that the readers from diverse backgrounds can appreciate the work. (Line 57)

Response:

We thank the reviewer for this excellent suggestion to improve the accessibility of our manuscript. We agree that a brief explanation of the foundational "deep vision-based framework" is essential for a broader audience.

To address this, we have expanded the relevant paragraph in the Introduction to clarify this concept. The revised text on lines 58 to 65 now explains that the core of this approach involves transforming the discrete shoreline protection inputs into 2D spatial maps. This allows to treat the problem as an image-to-image translation task, which has two key benefits: (1) it enables powerful Convolutional Neural Networks to learn the complex geometric relationships between coastal defenses and flood patterns, and (2) it unlocks the use of image-based data augmentation techniques to overcome data scarcity.

We believe this added context makes the motivation and novelty of our work much clearer. The revised paragraph has been highlighted in the manuscript.

The paper lacks a dedicated and concrete objective description; rather, it mentions the conclusions from the study in the introduction. Please move the lines 61 – 81 to conclusions or discussions, if redundant, considering removing them.

Response:

We thank the reviewer for this insightful comment on the structure of the manuscript. We understand how the initial phrasing of our contributions list could be interpreted as a premature summary of the study's findings. While listing the main contributions in the introduction is a common convention in our field to help readers quickly grasp the scope and novelty of the work, we agree that the language could be significantly improved to avoid this ambiguity.

To address this, we have carefully rephrased the "contributions" part (line 68 to 78 now). We have revised the wording to frame each point as a forward-looking objective of the paper rather than a backward-looking summary of its results. Specifically, we have shifted the language to focus on the actions our paper undertakes (e.g., "propose," "present," "conduct," and "employ") and removed phrases that stated the outcome of these actions (e.g., "demonstrating superior performance").

We believe this revision clarifies the purpose of this section as a statement of the paper's primary contributions and objectives, not as a list of conclusions, thereby resolving the structural concern the reviewer identified.

Lines 83 – 90, redundant information. Please remove them while merging sections 1 and 2.

Response:

We thank the reviewer for pointing out this redundancy. We agree that the opening paragraph of the former Related Works section (which contained lines 83-90) repeated information already present in the introduction.

As part of the major revision where we merged the Introduction and Related Works sections, we have removed this redundant paragraph entirely. This ensures the new, unified Introduction is more concise and avoids the repetition the reviewer correctly identified, making for a stronger and more focused opening to the manuscript.

Lines 91 – 98, irrelevant information.

Response:

We thank the reviewer for pointing this out. We agree that the discussion on traditional flood management techniques, which was the subject of lines 91-98 (in original manuscript), was not directly relevant to the core focus of our paper on computational surrogate modeling.

As part of the comprehensive revision where we merged the Introduction and Related Works sections, this entire paragraph has been removed. This action helps ensure that the new, unified Introduction is streamlined and focuses squarely on the background and literature pertinent to our proposed methodology, as the reviewer recommended.

Mention what kind of flood protection strategies exist here.

Response:

We thank the reviewer for this insightful comment on clarifying the scope of our work. To address this, we have revised the second paragraph of the Introduction to include a more pertinent discussion of coastal adaptation strategies. The revised text now introduces the widely recognized strategic framework of 'Protect, Accommodate, and Retreat', providing important context for the reader. We then clarify that our study focuses specifically on 'Protect' strategies and further explain that our analysis is concerned with the optimal spatial configuration of these defenses, not their specific engineering designs. We believe this revision effectively clarifies the scope and context of our work, and the updated paragraph has been highlighted in the manuscript.

Lines 123 – 125, not clear.

Response:

We thank the reviewer for highlighting this lack of clarity. The original lines were part of the former Related Works section. While this section has been removed as part of our larger structural revision, we have ensured that the core idea, identifying the limitations of prior work, was retained and improved in the new Introduction.

Specifically, we have rephrased the sentence that describes the research gap to be much clearer. The revised text (lines 53 to 55 now) explicitly states that many existing studies "do not jointly consider the complex, long-term impacts of both SLR and dynamic shoreline adaptation strategies," directly addressing the ambiguity the reviewer pointed out.

How is the training done for different SLRs?

Response:

We thank the reviewer for this important question, which made us realize our training strategy was not sufficiently detailed. To address this, we have substantially revised the Experimental Setup section to clearly explain our methodology for handling different SLR scenarios.

In the Dataset Splits subsection (Section 4.1), we have updated Table 2 to include the datasets used for fine-tuning (SF 0.5m and 1.5m SLR), clarifying the purpose and composition of all data used in the study. Furthermore, in the Model Optimization and Training Protocol subsection (Section 4.2), we now explicitly describe our two-phase approach. We detail the primary training phase, which used the larger AD (0.5m) and SF (1.0m) datasets, followed by the fine-tuning for generalization phase, where the model was adapted to the new SLR scenarios using a curriculum-based strategy to ensure stable learning.

We believe these revisions provide a clear and transparent account of our experimental protocol, resolving the ambiguity the reviewer rightly pointed out. The relevant sections and the table have been highlighted in the manuscript.

Figure 1 can be moved to Methodology and please describe the overall framework and the steps involved there.

Response:

We thank the reviewer for this excellent suggestion to improve the structure of the Methodology section. We agree that providing a high-level overview of the entire framework before detailing the model architecture would create a more logical flow for the reader.

As recommended, we have moved Figure 1 to the beginning of the Method section and created a new introductory subsection titled "Proposed Framework (Section 3.1)." This new subsection provides a clear, step-by-step description of the end-to-end workflow illustrated in the figure, covering data generation, preprocessing, the primary training path, the inference path, and the fine-tuning loop.

This new structure ensures that the reader first understands the overall process before delving into the specific architectural details of the CASPIAN-v2 model, providing a much clearer and more effective narrative. The new subsection and relocated figure have been highlighted in the revised manuscript.

Study Area and Data Description

Line 136, what do you mean by environmental effects?

Response:

Thank you for the suggestion. In the revised manuscript, we have clarified by stating that the environmental impacts are mostly focused on the effects on transportation links, and specifically whether important arterials such as shoreline highways or freeways will be flooded due to SLR.

Briefly describe OLU discretization in a few sentences for the benefit of readers.

Response:

Thank you for the question. In Section 2.1.2 the revised manuscript, we introduce the idea of Operational Landscape Units (OLUs) and clarify that for San Francisco Bay, the discretization of the was taken from SFEI, 2019. In Section 2.1.1, for Abu Dhabi, the urban coastline was also discretized into OLUs based on the Abu Dhabi Plan 2030 (Abu Dhabi Urban Planning Council, 2007). Maps of the OLU boundaries used for both San Francisco Bay and Abu Dhabi have been moved from the Supplementary information section to the main body of the revised text (Figures 1 and 2).

Mention past flood damages caused by tides.

Response:

Thank you for the question. Our focus on tidal flooding within San Francisco Bay was to highlight the unique tidal behavior within the Bay where the construction of sea walls along certain portions of the shoreline Bay may in fact exacerbate the sea level in other areas of the Bay. For example, Hollemann and Stacey, 2014 showed that protecting the southern portion of the Bay may increase maximum water levels by 0.2 in the north of the Bay; Wang et al., 2018 showed that protection of the southern portions of the West Bay adversely affects the water levels of the East Bay, and vice-versa. This discussion has been added to Section 2.1.2 in the revised manuscript.

Briefly mention here about the training dataset.

Response:

We thank the reviewer for this suggestion. To provide earlier context on the data, we have revised the Data Sources and Hydrodynamic Simulations section (Section 2.2). We have expanded the accompanying text to briefly summarize the composition and purpose of the datasets used for primary training, blind testing, and generalization. This revision provides the reader with a high-level understanding of our data strategy upfront, with the more granular details on splits and protocols reserved for the Experimental Setup section (Section 4.1). The updated text has been highlighted.

How was the Delft 3D model validated? How accurate is it?

Response:

For San Francisco Bay, the Delft3D model was adapted from the CoSMoS model originally developed by Barnard (2014) and adapted to San Francisco Bay by Wang et al (2017), and validated in the past using tidal gages at 9 tidal gage locations in and around San Francisco Bay. Pearson correlation coefficients ranged from 0.9862 to 0.9996, while the RMS ratios (the ratio of modelled versus measured RMS amplitudes) ranged from 0.973 to 1.027 (please refer to Wang et al., 2017). This detail has been added to Section 2.2 of the revised manuscript.

For Abu Dhabi, the Delft3D model was validated using water level data from 196 tidal gage locations throughout the Persian Gulf (as the hydrodynamic model encompassed the entire Gulf in addition to the western portions of the Gulf of Oman). The water levels at these locations were compared with one month's worth of hydrodynamic simulation, and the resulting absolute RMSE values ranged from 0.0013 to 0.0043 m in the vicinity of Abu Dhabi. More validation details for Abu Dhabi can be found in Chow and Sun, 2022. This detail has been added to Section 2.2 of the revised manuscript.

Rather than points, use line or raster to areas susceptible to flooding. Figure 2a.

Response:

Thank you for the suggestion. As recommended, we have revised Figure 2a to replace the point-based visualization that more clearly delineates areas susceptible to flooding. This will improve readability and better convey spatial flood risk patterns across the region.

Table 1 – What does main set, hold out set, etc., mean? Please explain when you have a table or figure in the manuscript for a general reader.

Response:

We thank the reviewer for pointing out the need for clearer definitions of our dataset terminology. To address this, we have revised the text accompanying Table 1 in the Data Sources and Hydrodynamic Simulations section (Section 2.2) to explicitly define the purpose of each dataset partition. The paragraph now clarifies that the Main Set is for primary training/testing, the Holdout Set is for blind testing on challenging scenarios, and the Generalizability Set is for evaluating model adaptability. Furthermore, we have updated the table caption to include these concise definitions, ensuring the table is more accessible to a general reader. The relevant text and caption have been highlighted.

What are Shamal winds? Describe in a sentence or two with relevant literature.

Response:

Thank you for the question. While we have referenced Langodan et al., 2023 in Section 3.1.1 relating to the Shamal winds, we have added an explanation of these winds Section 2.2 of the revised text. While the Persian Gulf does not typically experience tropical cyclones, it is known for its northwesterly winds generally occurring with winds at about 20 m/s with sudden onset and sustained over a period of up to 3-5 days. These are called the Shamal winds (meaning "North" in Arabic) and occur at least 10 times annually, mainly during the winter months (Senafi et al., 2015; Li et al., 2020). These references have also been added to the revised manuscript.

It is unclear what you mean by three months. Period of simulation or the computational time itself? What is the significance of choosing this?

Response:

Thank you for the question. A three-month period of simulation was chosen for both the San Francisco and the Abu Dhabi simulations. This was chosen to balance the need to model a larger number of tidal cycles that are modeled per simulation, with the computational time and storage space used for the simulations. Additionally for Abu Dhabi, the three-month simulation period also corresponds to a period of Shamal winds that lasted 3 months during the winter of 2017. We have added this explanation to Section 2.2 of the revised manuscript.

Method

Section 4.1, please shorten this to retain relevant general information. The rest can be moved to supplementary so that the readers don't lose interest.

Response:

We thank the reviewer for this constructive suggestion. As recommended, we have substantially revised Section 3.2 (previously 4.1) to be more concise and conceptually focused. The revised section now presents a high-level overview of the CASPIAN-v2 architecture, emphasizing the scientific purpose of the encoder, bottleneck, and decoder stages rather than their low-level implementation. All granular details, including specific network operations and equations, are presented in the Supplementary Material (Section S3). We are confident that this revision makes the core text more accessible and engaging for the journal's readership, while still providing the full technical details for readers who seek them. The revised Section 3.2 has been highlighted in blue in the manuscript.

Describe a spatial fitness metric.

Response:

We thank the reviewer for this valuable suggestion to incorporate a metric focused on spatial accuracy. As requested, we have now introduced a dedicated spatial fitness metric to our evaluation framework. Specifically, we have added the Dice Similarity Coefficient (DSC) to the Evaluation Metrics section (Section 4.4). We explain that this metric complements our existing point-wise error metrics by evaluating the spatial overlap and geometric alignment of the predicted flood extent against the ground truth. We also detail the methodology, which involves binarizing the continuous model output to create inundation masks and then computing the DSC. This allows us to explicitly quantify the model's performance in terms of correctly matched inundation areas, overprediction (False Positives), and underprediction (False Negatives). The new subsection describing the DSC has been added to the manuscript and highlighted for the reviewer's convenience.

Results

Tabulate and discuss the values of the spatial fitness index for different DL/ML models.

Response:

We thank the reviewer for this excellent suggestion to provide a comparative analysis of spatial accuracy. As requested, we have computed the Dice Similarity Coefficient (DSC) as a spatial fitness metric for our proposed CASPIAN-v2 model and all baseline models. The results have been incorporated in two key ways:

- Updated Table: The DSC scores for all models have been added to Table 6, providing a direct quantitative comparison.
- New Figure and Discussion: To visually illustrate these spatial differences, we have added a new Figure 8. This figure explicitly shows the areas of matched, over-, and under-prediction for CASPIAN-v2 compared to the best-performing baseline. A new paragraph discussing the findings from both the table and the figure has been added to the results section.

This analysis shows that CASPIAN-v2 not only has lower point-wise error but also achieves superior spatial accuracy, more reliably capturing the true flood extent. The relevant updates to the table, the new figure, and the discussion have all been highlighted in the manuscript.

Table 5, please describe the performance of the proposed method compared to the existing methods in the paper.

Response:

We thank the reviewer for this comment, which prompted us to improve the framing and context of our comparative analysis. To address this, we have added a new introductory paragraph to the Performance Benchmarking against State-of-the-Art Methods section. This paragraph now provides a high-level summary of the comparison and includes a direct cross-reference to the Baseline Models subsection (Section 4.3) in our Experimental Setup, where we detail the selection and implementation of all the state-of-the-art models used in our benchmark. This approach ensures the reader understands which models were used for comparison and can easily find the implementation details, while the Results section remains focused on discussing the outcomes. We believe this new structure better contextualizes the performance of CASPIAN-v2. The new paragraph has been highlighted in the manuscript.

Please use a zoomed-in figure to illustrate the effect of the representation of the protection measure in the manuscript.

Response:

We thank the reviewer for this excellent suggestion. We agree that zoomed-in figures are essential for illustrating the local-scale impact of shoreline protection on flood dynamics. To address this, we have updated both Figure 5 and Figure 6 to include new, zoomed-in insets that focus on specific Operational Landscape Units (OLUs). We have also revised the figure captions and the corresponding discussion in the Results section to use these insets to explain the physical phenomena. The text now clearly describes how inundation patterns are directly influenced by the protection status of the nearest OLU. For instance, showing how an area

remains dry when its coastal segment is protected, but floods when it is left unprotected, and we highlight our model's ability to accurately capture these critical, fine-grained effects.

We believe these revisions now effectively illustrate the physical impact of the protection measures and our model's ability to simulate them, as the reviewer recommended. The updated figures and text have been highlighted in Sections 5.2.1 and 5.2.2.

Section 6.2.2, Lines 475 – 478, lacks a clear explanation.

Response:

We thank the reviewer for highlighting this lack of clarity. We have rewritten the paragraph in Section 5.2.2 to provide a much more detailed and clear explanation of the model's performance on the holdout set.

The revised text now explicitly clarifies that the "abrupt changes" refer to the sharp inundation boundaries created by complex flood protection scenarios. We further explain that the holdout set was specifically designed with these challenging configurations (e.g., one side of the SF bay protected while the other is not) to test the prediction capabilities of the model. Finally, we elaborate on the significance of this result by explaining that the model's success on these unseen configurations, despite being trained on a small subset of all possible combinations (2^n , where n is the number of total OLU's), demonstrates that it is learning the fundamental spatial logic of flood protection. We are confident this revised section now provides the necessary context and detail. The updated paragraph has been highlighted in the manuscript.

Section 6.2.4, a repeated heading, please be more specific and clearer.

Response:

Response: We thank the reviewer for highlighting this important issue. We agree that the repeated headings were ambiguous. To address this, we have revised the subheadings in both the Quantitative (5.1 (previously 6.1)) and Qualitative (5.2 (previously 6.2)) Results sections to be more specific about their content, thereby ensuring they are unique and clear. For instance, the headings for generalizability are now "5.1.5 Numerical Assessment of Generalizability" and "5.2.4 Visual Assessment of Generalizability". Similar adjustments have been made to all parallel subsections to clearly distinguish between metric-based and visual evidence. The updated headings are highlighted in the revised manuscript.

Line 494, how is the ground truth information collected? What is the associated error?

Response:

We thank the reviewer for this critical question regarding the source and quality of our ground truth data. We have substantially revised the Data Sources and Hydrodynamic Simulations section (Section 2.2) to explicitly address this.

The revised section now clearly states that our "ground truth" data is derived from extensively validated, physics-based hydrodynamic simulations (Delft3D). To answer the question about the associated error, we have integrated the validation details for both the San Francisco and Abu Dhabi models directly into this section. We now report the key validation metrics, such as

Pearson correlation coefficients and RMSE values from comparisons with real-world tidal gage data, and provide citations to the original validation studies.

This revision ensures that the source and reliability of our ground truth data are clearly established at the beginning of the methods section, providing the necessary context and transparency for the reader. The updated section has been highlighted in the manuscript.

Line 506, summarize in a few sentences about data augmentation.

Response:

We thank the reviewer for pointing this out. As requested, we have expanded the sentence to include a brief summary of our data augmentation technique, specifying that it involved applying random spatial cutouts and scaling factors. To provide full context without adding length, we have also added a direct cross-reference to Section 4.1, where the augmentation strategy is detailed. The revised sentence is highlighted at line 551 in the revised manuscript.

Figure 9, please move it to the supplementary.

Response:

We agree with the reviewer that this figure is better suited for the Supplementary Material. As suggested, we have moved the data distribution figure (formerly Figure 9) to the supplementary section (now labeled as Figure S6). This helps to streamline the main text and keep the focus on the primary results. The reference in the manuscript has been updated accordingly.

References

1. *Bates, P. D., Horritt, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. Journal of Hydrology, 387(1–2), 33–45. <https://doi.org/10.1016/j.jhydrol.2010.03.027>*
2. *De Almeida, G. A. M., & Bates, P. (2013). Applicability of the local inertial approximation of the shallow water equations to flood modeling. Water Resources Research, 49(8), 4833–4844. <https://doi.org/10.1002/wrcr.20366>*
3. *Li, Z., & Hodges, B. R. (2019). Modeling subgrid-scale topographic effects on shallow marsh hydrodynamics and salinity transport. Advances in Water Resources, 129, 1–15. <https://doi.org/10.1016/j.advwatres.2019.05.004>*
4. *Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resources Research, 48(11), 1–16. <https://doi.org/10.1029/2012WR012514>*
5. *Nithila Devi, N., & Kuiry, S. N. (2024). A novel local-inertial formulation representing subgrid scale topographic effects for urban flood simulation. Water Resources Research, 60(5), e2023WR035334.*

6. Sanders, B. F., & Schubert, J. E. (2019). PRIMo: Parallel raster inundation model. *Advances in Water Resources*, 126, 79–95. <https://doi.org/10.1016/J.ADVWATRES.2019.02.007>

Response:

Thank you for sharing these references. We have incorporated the relevant literature, as suggested, into the revised manuscript to strengthen the discussion on computationally efficient flood modeling approaches and better contextualize the role of deep learning in this domain.

Response References:

1. Abu Dhabi Urban Planning Council. Plan Abu Dhabi 2030. Urban Structure Framework Plan. 2007. Available online: <https://www.ecouncil.ae/PublicationsEn/plan-abu-dhabi-full-version-EN.pdf>
2. Al Senafi, F.; Anis, A. Shamals and climate variability in the Northern Arabian/Persian Gulf from 1973 to 2012. *Int. J. Clim.* 2015,35, 4509–4528.
3. Chow, A.C. and Sun, J., 2022. Combining Sea level rise inundation impacts, tidal flooding and extreme wind events along the Abu Dhabi coastline. *Hydrology*, 9(8), p.143.
4. Holleman, R.C. and Stacey, M.T., 2014. Coupling of sea level rise, tidal amplification, and inundation. *Journal of Physical Oceanography*, 44(5), pp.1439-1455.
5. IPCC. Summary for Policymakers. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press: Cambridge, UK, 2021.
6. Li, D.; Anis, A.; Al Senafi, F. Physical response of the Northern Arabian Gulf to winter Shamals. *J. Mar. Syst.* 2020, 203, 103280.
7. San Francisco Estuary Institute (SFEI), 2019 (Beagle, J., Lowe, J., McKnight, K., Safran, S., Tam, L. and Szambelan, S.J., authors). San Francisco Bay shoreline adaptation atlas: Working with nature to plan for sea level rise using operational landscape units (No. SFEI publication# 915).
8. Wang, R.Q., Herdman, L.M., Erikson, L., Barnard, P., Hummel, M. and Stacey, M.T., 2017. Interactions of estuarine shoreline infrastructure with multiscale sea level variability. *Journal of Geophysical Research: Oceans*, 122(12), pp.9962-9979.
9. Wang, R.Q., Stacey, M.T., Herdman, L.M.M., Barnard, P.L. and Erikson, L., 2018. The influence of sea level rise on the regional interdependence of coastal infrastructure. *Earth's Future*, 6(5), pp.677-688.

Anonymous Referee # 2

Summary

This manuscript presents CASPIAN-v2, a novel deep learning framework for predicting coastal flooding under varying sea level rise (SLR) scenarios and shoreline protection strategies. The authors test their approach on two distinct geographical regions (Abu Dhabi and San Francisco) and demonstrate superior performance compared to state-of-the-art methods. The paper makes several contributions, including a new CNN-based architecture, comprehensive datasets from vulnerable coastal areas, and validation of generalizability across different SLR scenarios.

While the work represents a significant advancement in data-driven coastal flood prediction, several aspects require substantial revision before the manuscript is suitable for publication in HESS.

Response:

We thank the reviewer for their thoughtful summary and for recognizing the key contributions of our work. At the same time, we fully understand that several areas required further refinement. In the revised manuscript, we have addressed all raised concerns in detail to ensure the work meets the standards of HESS. We believe these revisions will significantly improve clarity, rigor, and impact of our work.

General Comments

Model Architecture Presentation and Justification

The CASPIAN-v2 architecture is sophisticated but presented in an overly complex manner that hampers understanding. Figure 4 contains too much information without sufficient explanation of design choices. The authors introduce multiple novel components (MARX blocks, SEE blocks) without adequate justification for these specific innovations over simpler alternatives.

Example: In Section 4.1.2, the rationale for integrating ResNeXt blocks with CBAM is not clearly connected to the specific challenges of flood prediction. The authors should explain why this combination addresses spatial dependencies in coastal flooding better than other attention mechanisms.

Recommendation: Provide a simplified schematic of the architecture alongside the detailed one and clearly justify each novel component in relation to the specific requirements of flood prediction tasks.

Response:

We thank the reviewer for this critical feedback. We agree that the manuscript required a stronger, empirically-backed justification for our novel architectural components and a clearer

presentation. To address this, we have made the following significant revisions, which also align with feedback from Reviewer 1 regarding the manuscript's structure:

- **Simplified Architecture Figure and Text:** In response to Reviewer 1's suggestion to shorten the Section 3.2 (previously 4.1), we have moved the detailed explanation of the CASPIAN-v2 architecture to the Supplementary Material. To address this reviewer's comment on clarity, we have now introduced a new, simplified schematic of the model architecture in the main text (Figure 4), which provides a clear, high-level overview. The original, more detailed figure has been moved to the supplementary section S3 alongside the detailed technical description.
- **New Bottleneck Ablation Study:** To empirically justify our design choices, we have added a new subsection to the Supplementary Material (Section S4.1). This new study systematically evaluates the performance of our proposed MARX block against a series of alternatives (e.g., standard ResNet, other attention mechanisms) and demonstrates the superiority of our chosen configuration.
- **Enhanced Existing Ablation Studies:** We have strengthened the existing ablation studies for both the number of MARX blocks (Section S4.3) and SEE blocks (Section S4.4) by adding a "without component" baseline experiment to each. This clearly demonstrates that the inclusion of these blocks provides a fundamental performance improvement over a simpler architecture.

We are confident that these substantial additions provide the rigorous justification for our design choices that the reviewer rightly requested. The relevant sections in both the main text and the supplementary material have been updated and highlighted.

Computational Efficiency Analysis

A primary motivation for developing surrogate models is the computational burden of physics-based simulations. However, the paper lacks a rigorous comparison of computational efficiency between the proposed model and alternatives.

Example: While Table 5 thoroughly compares prediction accuracy, it contains no information about training times, inference times, or memory requirements. This is particularly important given that lines 45-49 on page 2 emphasize computational burden as a key limitation of current approaches.

Recommendation: Include a comprehensive analysis of computational efficiency, comparing training and inference times across all evaluated models, and explicitly stating the practical time savings compared to hydrodynamic simulations.

Response:

We thank the reviewer for this crucial recommendation. We agree that a rigorous analysis of computational efficiency is essential to demonstrate the practical value of our surrogate model. To address this, we have made two major additions to the manuscript:

- We have updated Table 6 to include three new columns detailing the Computational Efficiency for all evaluated models: the number of trainable parameters, total training time (TT), and average inference time (IT) per scenario.
- We have added a new, dedicated subsection titled "Computational Efficiency Analysis (Section 5.1.4)" within the Results section. This new text provides a comprehensive analysis of the data from the updated table. It compares the efficiency of CASPIAN-v2 to both the ML/DL baselines and, most importantly, to the original physics-based Delft3D simulator, highlighting the trade-offs between speed and accuracy.

This new analysis now quantitatively demonstrates the dramatic practical time savings our model provides, reducing the time required to evaluate our full test set from an estimated 49 days with the hydrodynamic simulator to just 16 seconds with our model. We are confident this new section fully addresses the reviewer's concern and powerfully underscores the primary motivation for our work. The updated table and new subsection have been highlighted.

Uncertainty Quantification

The model provides deterministic predictions without addressing prediction uncertainties, which is crucial for risk assessment and decision support in coastal planning.

Example: The error maps in Figures 5-7 show where predictions differ from ground truth, but they don't indicate the model's confidence in its predictions, which is essential for reliable risk assessment.

Recommendation: Incorporate uncertainty quantification into the model (e.g., through ensemble methods, Bayesian techniques, or prediction intervals) or thoroughly discuss this limitation and its implications for practical use.

Response:

We thank the reviewer for this crucial recommendation. We fully agree that quantifying predictive uncertainty is essential for the practical application of our model in risk-sensitive decision-making. To address this, we have implemented a deep ensemble method, a well-established and robust technique for uncertainty quantification. We have added a new paragraph to the Discussion and Conclusion Section (Section 6) in the revised manuscript. This new section:

- Introduces and justifies the deep ensemble approach.
- Presents a new figure (Figure 11) with the resulting predictive uncertainty maps, which compellingly show that the model's uncertainty is highest at the flood boundaries and in other hydrodynamically complex areas.
- Discusses how this spatial uncertainty information is invaluable for coastal planners.

Furthermore, the full quantitative performance metrics for each of the five models in the ensemble, which demonstrate the stability of our approach, are now provided in a new table in the Supplementary Material (Section S8) and referenced from the main text.

We are confident that this substantial new analysis fully addresses the reviewer's concern and significantly enhances the practical utility of our work. The new section (lines 573 to 584), figure (Figure 11), and supplementary table (S9) have been highlighted in the manuscript.

Data Imbalance Handling

Figure 9 reveals severe class imbalance in the dataset, with non-inundated areas predominating. While the authors acknowledge this challenge, they don't adequately explain how their approach specifically addresses it.

Example: Section 7 mentions the imbalance issue but doesn't describe specific techniques beyond the hybrid loss function that were employed to mitigate its effects. It's unclear how the model achieves its reported high accuracy despite this challenge.

Recommendation: Elaborate on specific techniques used to address data imbalance, potentially including specialized sampling strategies, data augmentation approaches tailored to rare flood events, or custom components in the architecture designed for imbalanced spatial data.

Response:

We thank the reviewer for this important comment. We agree that a more explicit, evidence-backed explanation of how our model handles the severe data imbalance was needed.

To address this, we have substantially revised the paragraph in the Discussion and Conclusion section (Section 6, lines 555 to 567) that discusses this topic. The new text now clearly details our multi-faceted strategy for handling the data imbalance:

- We explain how our Hybrid Loss function is specifically designed for this challenge, with the Quantile loss component allowing us to up-weight the importance of the rare, positive flood values.
- We highlight the role of the attention mechanism in the MARX block, which forces the model to focus on the most critical and often under-represented inundated regions. Crucially, we now provide empirical evidence for this claim by referencing our Grad-CAM analysis (Figure 10), which visually demonstrates the model concentrating its attention on these vulnerable, unprotected areas.
- We reinforce that our choice of evaluation metrics (like DSC and Acc) was deliberately made to provide a balanced assessment of performance.

We are confident that this detailed, evidence-backed explanation now clearly describes the specific techniques that enable our model to achieve high accuracy despite the challenging nature of the data. The revised paragraph (lines 555 to 567) has been highlighted in the manuscript.

Real-world Application Context

The practical utility of the model for coastal planning is asserted but not demonstrated through concrete examples or integration pathways.

Example: The conclusion claims CASPIAN-v2 is "an essential tool for coastal resilience planning" (lines 539-540, page 28), but doesn't provide specific guidance on how planners might integrate this tool with existing decision-making frameworks.

Recommendation: Include a case study or conceptual workflow showing how the model could be integrated into actual coastal planning processes, identifying key stakeholders and decision points where the model adds value.

Response:

We thank the reviewer for this insightful recommendation to better connect our model to practical applications. We agree that demonstrating a clear pathway to real-world integration is essential.

To address this, we have added a new paragraph to the Discussion and Conclusion section (Section 6, lines 585 to 592). This new text explicitly outlines a conceptual workflow for how CASPIAN-v2 could be used by stakeholders such as coastal planners and engineers. We describe a two-tiered approach where our model is first used as a rapid scenario-assessment tool to explore a vast range of potential shoreline protection strategies, a task infeasible for traditional simulators. The most promising strategies identified by our model could then be selected for more rigorous analysis with high-fidelity hydrodynamic models.

We believe this addition provides the concrete guidance and integration pathway the reviewer requested, strengthening the paper's positioning as a valuable tool for coastal resilience planning. The new paragraph (lines 585 to 592) has been highlighted in the manuscript.

Specific Comments

Mathematical Notation Inconsistency

The paper uses inconsistent notation, particularly in Section 4.1, making the mathematical formulations difficult to follow.

Example: In Equations 1-5, subscripts sometimes denote indices and sometimes represent different variables entirely. The relationship between tensors across equations is not always clear.

Recommendation: Standardize notation throughout the paper and provide a notation table for reference.

Response:

We thank the reviewer for this critical feedback. We agree that the notation, particularly in the Bottleneck Stage equations (originally Eqs. 1-5) the reviewer highlighted, was inconsistent and required clarification.

To resolve this, we have performed a comprehensive standardization of our mathematical notation throughout the entire manuscript. The core change was to clearly distinguish descriptive subscripts (e.g., for "encoder" or "ResNeXt-1") from variable indices (e.g., for layer depth k or group g). We achieved this by formatting all descriptive subscripts in upright text. This principle has been applied to all equations in the Supplementary Material, including the Bottleneck and Decoder stages, as well as to the Evaluation Metrics in the main text.

Furthermore, as recommended, we have created a comprehensive Notation Table (S10) and added it to the Supplementary Material (Section S9). This table defines all mathematical symbols for easy reference. We are confident these changes directly address the issues the reviewer raised and significantly improve the clarity of our mathematical formulations. The revised text has been highlighted accordingly.

Evaluation Metrics Justification

While the paper employs multiple evaluation metrics, the rationale for these specific choices and their relevance to practical flood prediction applications isn't fully explained.

Example: The threshold exceedance metric ($\delta > \Delta$) is introduced in Section 5.4, but its practical significance for flood risk assessment isn't discussed.

Recommendation: Justify the choice of each evaluation metric in terms of its relevance to practical flood prediction applications and decision-making contexts.

Response:

We thank the reviewer for this insightful recommendation. We agree that a clear justification for our choice of metrics is essential for linking our evaluation to real-world applications. In response, we have substantially revised the Evaluation Metrics section (Section 4.4) to include a detailed rationale for each metric.

For every metric, we now introduce it with a brief explanation of what it measures and why it is relevant to practical flood prediction, risk assessment, and decision-making. For instance:

- We now clarify that ARMSE is used to highlight large, potentially catastrophic prediction errors, while AMAE provides a more general sense of the average error.
- For the Threshold Exceedance Metric ($\delta > \Delta$), which the reviewer specifically mentioned, we now explicitly connect it to operational decision-making, such as determining if a prediction error exceeds a critical infrastructure threshold (e.g., a building's floor height).

- Similarly, we justify the use of Acc by its importance in preventing false alarms in highly imbalanced flood data, and ARTAE by the need for a scale-invariant error measure.

We are confident that this expanded section now thoroughly justifies our evaluation framework and provides the practical context the reviewer requested. The entire revised section 4.4 has been highlighted in the manuscript for convenience.

Data Preprocessing Details

The data preprocessing section (3.3) lacks sufficient detail on critical aspects that could impact model performance.

Example: The method for mapping inundation coordinates onto a 1024×1024 grid (lines 182-184, page 9) is mentioned but not described in detail, despite this being a critical step that affects the spatial resolution of predictions.

Recommendation: Provide more detailed explanation of preprocessing steps, potentially with illustrative examples showing the transformation from raw data to model inputs.

Response:

We thank the reviewer for this important feedback. We agree that the original Data Preprocessing section was too succinct and did not adequately describe the critical step of transforming raw data into the spatial grid format required by our model.

To address this, we have expanded the section 2.3 (originally 3.3) in the main manuscript. The revised text now provides a more detailed summary of our preprocessing pipeline. Specifically, we describe how the spatial grid is defined, how inundation points are mapped to grid cells, and importantly, the conflict resolution strategy we employed to handle cases where multiple points fall within the same cell.

This enhanced description provides the necessary context for the reader to understand our methodology, while still directing them to the Supplementary Material (Section S2) for the full mathematical formulations and step-by-step breakdown. We believe this revision strikes the right balance between detail and conciseness, as the reviewer intended. The updated section 2.3 has been highlighted in the manuscript.

Ablation Study Presentation

The paper mentions ablation studies in the supplementary material but doesn't adequately summarize key findings in the main text.

Example: Line 341-342 on page 16 mentions "extensive ablation studies" but doesn't present the key insights derived from these experiments.

Recommendation: Include a summary table of ablation study results in the main text, highlighting the contribution of each novel component to overall performance.

Response:

We thank the reviewer for this excellent suggestion. We agree that summarizing the key findings from our ablation studies in the main text provides crucial insight into the contribution of our model's novel components.

To address this, we have made two significant changes. First, we have expanded our ablation studies in the Supplementary Material to include a new, detailed "Impact of Bottleneck Configurations" section (S4.1). This new study systematically evaluates our proposed MARX block against seven alternative configurations to empirically justify its design.

Second, to ensure these key findings are visible in the main text as requested, we have added a new summary table (Table 3) and an accompanying paragraph to the Model Optimization and Training Protocol section 4.2. This table now synthesizes the outcomes of all ablation studies, including the new bottleneck analysis, highlighting the contribution of each optimized component to the model's final performance.

This addition ensures that the main manuscript now clearly presents the key insights from our extensive ablation experiments, while still directing the reader to the Supplementary Material for the full detailed analysis. The new table and text have been highlighted in the manuscript.

Figure Clarity and Interpretation

Several figures are complex and difficult to interpret, with insufficient explanation in captions and text.

Example: Figure 7 compares model predictions across different approaches, but the subtle differences between models are difficult to discern with the chosen color scale and presentation format.

Recommendation: Improve figure clarity through better color scales, simplified presentations, or additional explanatory elements like difference maps to highlight where each model performs better or worse.

Response:

We thank the reviewer for this excellent suggestion. We agree that the original Figure 7 made it difficult to discern the specific improvements of our model over the baselines.

To address this, we have completely redesigned the figure to be more direct and informative, following the reviewer's recommendation to use difference maps. The new Figure 7 now presents a focused comparison for each region that includes: (a) the ground truth, (b) the absolute error map of our CASPIAN-v2 model, and crucially, (c-f) a series of new error

difference maps that compare CASPIAN-v2's performance directly against each of the key SOTA baselines.

These difference maps use a color scale to explicitly highlight regions where CASPIAN-v2 provides a clear improvement over each baseline. We have also rewritten the accompanying section and the figure caption to guide the reader through this new, more powerful visualization. We are confident that this revised figure now effectively and clearly communicates the qualitative superiority of our model, as the reviewer intended. The updated figure and text have been highlighted in the manuscript.

Summary Assessment

This manuscript presents valuable research on deep learning for coastal flood prediction, with promising results that could significantly advance the field. However, major revisions are needed to address issues related to model architecture presentation, computational efficiency analysis, uncertainty quantification, data imbalance handling, and real-world application context.

With these improvements, the paper has the potential to make a significant contribution to both the technical literature on deep learning for environmental modeling and practical coastal planning applications.

Response:

We sincerely thank the reviewer for their encouraging assessment and constructive feedback. We appreciate the recognition of the potential impact of our work and fully acknowledge the areas that require improvement. In the revised manuscript, we have made substantial revisions to enhance the clarity of the model architecture presentation, provided a comprehensive computational efficiency analysis, incorporated uncertainty quantification, clarified our approach to data imbalance handling, and better contextualized the results for real-world coastal planning applications. We are confident that these revisions will strengthen the manuscript and align it more closely with the expectations of both technical and applied research communities.