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 require further refinement. In the revised manuscript, we will address 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 the helpful feedback regarding the clarity and justification of the CASPIAN-v2 architecture. In the revised manuscript, we plan to redraw Figure 4 to include a simplified schematic of the overall encoder–bottleneck–decoder structure, alongside the existing

detailed version. This will help improve readability and guide the reader through the hierarchical design.

Additionally, to better justify the inclusion of the MARX and SEE blocks, we will update the ablation study to show model performance when these components are entirely removed or replaced with alternative modules. While the current ablation study already includes results with different numbers of MARX and SEE blocks, we recognize the importance of evaluating their overall necessity. We have also experimented with other configurations in the bottleneck, including standard ResNet, ResNeXt blocks, and Squeeze-and-Excitation modules. These additional comparisons will be included to demonstrate why the selected components (ResNeXt blocks with CBAM) in the bottleneck are more effective in capturing spatial dependencies and multi-scale interactions critical for accurate flood prediction.

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 appreciate the reviewer's comment regarding the need for a computational efficiency analysis. In the revised manuscript, we will include a detailed comparison of training time, inference time, and GPU memory usage across all evaluated models. Specifically, our CASPIAN-v2 model requires approximately 23 hours to train on a local machine equipped with an NVIDIA RTX 4090 GPU. Its inference time is around 0.227 seconds per scenario, and the model is lightweight, comprising only 0.38 million parameters.

In contrast, simulating a single scenario using traditional hydrodynamic models (Delft3D for San Francisco and Delft3D coupled with SWAN for Abu Dhabi) takes approximately 14 hours and 19 hours, respectively, on high-performance computing infrastructure. This means that running all 72 scenarios in the test set (36 for Abu Dhabi and 36 for San Francisco) would take around 49 days using these simulators. In comparison, CASPIAN-v2 can process the same 72 scenarios in roughly 17 seconds.

We will include these comparisons in a new table and accompanying discussion to clearly demonstrate the significant time and resource savings achieved by our surrogate model, directly addressing the computational motivation highlighted in the manuscript.

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 appreciate the valuable feedback from the reviewer regarding uncertainty quantification. We fully agree that predictive uncertainty is essential for informed decision-making in coastal risk management. While we previously used Grad-CAM to provide qualitative insights into the spatial regions influencing the model's predictions, we acknowledge that this approach supports interpretability but does not quantify uncertainty.

To address this comment, in the revised manuscript we plan to incorporate a predictive uncertainty estimation method. Specifically, we propose to apply Monte Carlo Dropout by enabling dropout at inference time to generate multiple outputs for the same input. Alternatively, we may implement deep ensembles, where multiple independently trained models are used to derive a distribution over predictions. In addition, we will explore the use of random cutout–based test-time augmentation, where different variants of the same input are passed through the model, and the variability in outputs is used to estimate uncertainty. This method leverages our existing data augmentation strategy and provides a simple, architecture-agnostic way to probe model robustness.

All three approaches will help generate confidence maps alongside the predicted inundation, offering a more robust basis for risk-sensitive planning. We will include the resulting uncertainty maps and a corresponding discussion in the revised 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:
Thank you for highlighting this important point. We acknowledge the significant class imbalance in the dataset, as shown in Figure 9, with non-inundated points dominating the spatial distribution. While we did not apply explicit sampling or augmentation strategies to overcome this, our current formulation incorporates a hybrid loss function and custom architectural components (e.g., MARX and SEE blocks) designed to learn and prioritize spatially meaningful features, particularly around protection boundaries and inundation-prone regions.

As further evidenced in Figure 10, the model tends to focus more on areas near unprotected OLUs, which are more likely to flood. This behavior suggests that despite the numerical dominance of non-inundated points, the network implicitly assigns greater representational focus to the spatial characteristics associated with flood-prone zones. In the revised manuscript, we will clarify this point and also briefly discuss possible future enhancements, such as spatial weighting schemes or scenario-focused data balancing, to further address the imbalance in flood prediction tasks..

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:
Thank you for this insightful comment. In the revised manuscript, we will include a workflow explanation describing how CASPIAN-v2 can be applied in real-world coastal planning contexts. Specifically, we will outline how the output of the model, such as scenario-based flood maps and uncertainty estimates, can support planners, engineers, and policymakers at various decision

points. We will also describe how the model complements traditional hydrodynamic simulations by enabling rapid scenario analysis, which is particularly useful for emergency preparedness and evaluating the impacts of different shoreline protection strategies. This addition will clarify the practical value of the work and strengthen its positioning as a decision-support tool.

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:
Thank you for pointing out the issue with notation clarity. In the revised manuscript, we will standardize the mathematical notation, ensuring consistent use of subscripts, tensor dimensions, and variable references. Additionally, we will include a notation table to clearly define all symbols and improve readability.

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:
Thank you for highlighting this important point. In the revised manuscript, we will elaborate on the rationale behind our choice of evaluation metrics. While we included multiple metrics to comprehensively assess model performance from different perspectives, we now recognize the need to clarify what each metric represents and how it relates to practical flood prediction and decision-making.

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:
Thank you for this observation. Due to space constraints in the main manuscript, we provided a detailed description of the data preprocessing steps in Section S2 of the supplementary material. This section outlines each stage of the transformation from raw simulation outputs to the standardized 1024×1024 input grids used by the model. We will ensure that this is clearly referenced in Section 3.3 of the revised 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:
Thank you for this helpful suggestion. In the revised manuscript, we will include a summary table of the ablation study results in the main text, highlighting the performance impact of each novel component. This will provide a clearer understanding of their individual contributions and key insights from the experiments currently detailed in the supplementary material.

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:
Thank you for pointing this out. In the revised manuscript, we will improve the clarity of key figures, particularly Figure 7, by adopting more perceptually distinct color scales and adding explanatory elements to better highlight variations between model predictions. We will also enhance the figure captions and in-text descriptions to guide interpretation and ensure that the visual comparisons are more accessible and informative.

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 response, we plan to undertake substantial revisions to enhance the clarity of the model architecture presentation, provide a comprehensive computational efficiency analysis, incorporate or discuss uncertainty quantification, clarify our approach to data imbalance handling, and better contextualize 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.