Dear Editor,

Thank you for the opportunity to revise our manuscript and for the constructive comments provided by both reviewers. Detailed point-by-point responses to the reviewers were already provided during the interactive discussion. In the present revision, we have implemented the corresponding changes throughout the manuscript and carefully checked the consistency of the revised text, figures, tables, methodology, evaluation protocol, and conclusions. For transparency, we are submitting both **a clean revised manuscript** and **a PDF of the marked-up manuscript highlighting all changes relative to the original submission**.

The main revisions and their correspondence to the reviewer comments are summarized below.

No.	Main revision implemented in the manuscript	Main location(s)	Reviewer comment(s) addressed
1	**Reframed the overall scope of the framework.** The manuscript now consistently presents the method as a rapid, solver-trained SWE surrogate for reproducing hydrodynamic-solver-generated flood-depth and inundation fields, rather than as a standalone forecasting model or replacement for high-fidelity hydrodynamic solvers. Claims concerning physical consistency, mass dynamics, and hydraulic fidelity were also moderated.	Abstract; Introduction; Methodology; Results/Discussion; Conclusion	**Reviewer 1:** Concerns 1–2. **Reviewer 2:** Comments 2, 5, 6
2	**Clarified the forcing and forecasting interpretation.** We now explicitly state that the FNO is conditioned on an externally prescribed forcing trajectory and does not independently forecast future inflows. The bathtub baseline and Magnifier are described as terrain-aware and volume-aware depth-reconstruction components rather than full SWE models of velocity, discharge, or momentum.	Sections 2.2, 2.3.2–2.3.3, 2.4	**Reviewer 1:** Concern 1. **Reviewer 2:** Comments 2, 5
3	**Clarified the validation framework.** The reported evaluation is now explicitly described as surrogate-to-ANUGA assessment. FIM-bench/observations are used to support evaluation of the underlying hydrodynamic setup, while the surrogate performance metrics quantify agreement with ANUGA-generated reference solutions.	Section 3 and related Results discussion	**Reviewer 1:** Concerns 1–2. **Reviewer 2:** Comments 2, 6
4	**Added computational-efficiency analysis.** A new subsection reports ANUGA runtimes for the Neuse, Fall River, and Chowilla cases and compares these with the approximately 1-s surrogate inference time. The discussion clarifies that the reported acceleration represents reduced marginal cost after training for repeated ANUGA-like scenario evaluations.	New subsection: *Computational efficiency and marginal inference cost*	**Reviewer 1:** Concern 2

Continued on next page

No.	Main revision implemented in the manuscript	Main location(s)	Reviewer comment(s) addressed
5	**Clarified and standardized the training/test split.** The manuscript now consistently states that each case study contains 150 realizations, with 100 used for training and 50 completely held out for final testing. Ambiguous terms such as "training and validation" and "validation/testing" were removed.	Figure 3; Table 1; Section 3.3; case-study descriptions; Results; table captions	**Reviewer 2:** Comment 1
6	**Clarified the meaning of generalization.** The held-out hydrographs are now explicitly described as unseen realizations generated within the same controlled perturbation envelope. We therefore avoid presenting the current experiment as strict out-of-distribution extrapolation and identify independent historical-event and withheld-forcing-range tests as future work.	Appendix C; limitations discussion	**Reviewer 2:** Comment 1
7	**Added event-level distributional evaluation.** Two violin-plot figures were added to show event-wise relRMSE/NSE and CSI distributions over the 50 independent test scenarios. These plots complement the aggregate performance tables and show event-to-event variability.	Results, following Tables 2 and 3	**Reviewer 2:** Comment 7
8	**Added inundation-threshold sensitivity analysis.** CSI is now evaluated at wet–dry thresholds of 0.01, 0.05, 0.10, and 0.20 m for both the coarse FNO and fine Magnifier across all three domains.	Section 3.5.3; Section 4.2; new CSI-sensitivity figure	**Reviewer 2:** Comment 7
9	**Clarified peak-magnitude and peak-timing metrics.** The manuscript now explains that negative signed values indicate underestimation or earlier predicted timing, whereas positive values indicate overestimation or delayed timing. The corresponding figure label was also corrected from "top 10%" to "top 5%" to match the stated peak-region definition.	Section 4.4; peak-metric figure	**Reviewer 2:** Comment 4
10	**Added a plain-language methodology overview.** A short intuitive explanation was inserted before the technical neural-operator formulation to explain the coarse-to-fine workflow: the FNO captures broad flood evolution, while the Magnifier refines local predictions using high-resolution terrain and learned residual corrections.	Beginning of Methodology section	**Reviewer 1:** Comment 5
11	**Clarified ANUGA-to-raster preprocessing.** Section 3.3 now describes only the preprocessing actually used in the study: ANUGA water-depth outputs on the unstructured mesh are interpolated onto structured Cartesian raster grids to form consistent tensors for FNO/Magnifier training and evaluation.	Section 3.3	**Reviewer 2:** Comment 3

Continued on next page

No.	Main revision implemented in the manuscript	Main location(s)	Reviewer comment(s) addressed
12	**Expanded implementation and reproducibility details.** The revised manuscript now reports additional information including batch size, weight decay, random seed, supervised depth-loss objective, checkpoint-selection procedure, and training/evaluation workflow.	Section 3.4, *Implementation and Training Configuration*	**Reviewer 2:** Comment 8
13	**Revised the interpretation of physical and uncertainty results.** Volume preservation, flood-depth agreement, inundation accuracy, and uncertainty performance are now interpreted as evidence of successful emulation of solver-generated depth and inundation fields, rather than independent proof of complete SWE consistency. The manuscript explicitly acknowledges that velocity, discharge, momentum, and local continuity residuals are not evaluated.	Results and Discussion	**Reviewer 2:** Comments 5, 6
14	**Revised the Conclusion.** The final section now consistently describes the model as a fast, physically anchored, solver-trained surrogate for repeated scenario simulation and uncertainty-aware analysis. It also explicitly states that high-fidelity hydrodynamic solvers remain essential for generating training data, calibration, validation, and detailed hydraulic analysis.	Conclusion	**Reviewer 1:** Concerns 1–2. **Reviewer 2:** Comments 2, 5, 6

We believe these revisions have substantially improved the manuscript by more clearly defining the scope and intended use of the proposed framework, strengthening the evaluation and reproducibility of the study, and addressing the methodological and interpretive concerns raised by both reviewers.

We thank the Editor and the reviewers again for their constructive feedback and for the opportunity to improve the manuscript. We hope that the revised manuscript is now suitable for publication in *Hydrology and Earth System Sciences*.

Sincerely,

Chaopeng Shen