Dear Reviewers,

Thank you for your comments on our manuscript, *"Impact of dataset design on LSTM-based urban pluvial flood prediction: Length, feature dimensions, and rainfall stratification."* Your suggestions have helped improve the rigor and clarity of our study.

We have addressed all comments in the revised manuscript. Below are our detailed, point-by-point responses:

Part I: Response to Specific Comments

1. Introduction and Generalization: The motivation is clear and well-structured. However, line 87 states that insights from the LSTM model are expected to be transferred to other ML models. The authors should justify this claim, explaining why these findings (especially threshold effects) would apply to other models.

Response: We agree that claiming absolute generalizability to other machine learning models without empirical evidence lacks rigor. In the revised manuscript, we have redefined the LSTM as a "representative baseline sequence model" and revised the discussion and conclusion to clarify: "Although the patterns derived under the LSTM baseline framework offer reference values for other sequence learning models, broader generalization (e.g., extension to Graph Neural Networks) necessitates further empirical validation".

2. Section 2.2: "An NSE is reported for the hydraulic model, but more detail is needed. What variable does this NSE represent... At which control points and under what rainfall events was this validated?"

Response: We have supplemented Section 2.2 with detailed information on the hydrodynamic model validation.

1）**Variables and Control Points:** The NSE represents the pipe flow rate (m^3/s) at key monitoring nodes within the drainage pipe network.

2）**Rainfall Events for Validation:** In addition to the initial calibration, we added a supplementary validation phase for three typical observed rainfall events (with total rainfall volumes of 80 mm, 120 mm, and 50 mm, respectively) (Table 2), yielding NSE values of 0.82, 0.75, and 0.88. Furthermore, we added Figure 13 to directly compare the inundation area hydrographs simulated by InfoWorks ICM and the

LSTM model under real rainfall events, demonstrating that the NSE values for the simulated events exceed 0.8.

3. Line 134: "Is there a specific reference for the IDF formula used?"

Response: We have added the source and specific parameters of the storm intensity formula (IDF). The Chicago design storm instantaneous rainfall intensity formula (Eq. 5) was fitted from local historical rainfall data; we have detailed the geographical and statistical significance of each empirical parameter (C, b, n).$q = \frac{1602(1+1.037\lg p)}{(t+11.593)^{0.681}}$. We have also detailed the geographical and statistical significance of each empirical parameter (e.g., A_1, C, b, n).

4. Line 147: The equation for a distributed model is an oversimplification as it neglects lateral flows between cells and surface storage. This needs a more rigorous explanation.

Response: $Q(t) = P(t) - I(t) - D(t)$ We have clarified that Equation 6 is strictly used to calculate the net surface runoff rate (excess infiltration rate) prior to concentration. The actual physical processes—2D spatial concentration, lateral flow between grid cells, and surface depression storage—are handled by the InfoWorks ICM 2D module based on shallow water equations and a DEM. This hydraulic conversion is formally expressed via Equation 8, which incorporates surface slope and depression storage.$Y(t) = \varphi(V(t), DEM, S, S_d)$

5. Line 148: "There is a typo in "III"; it should likely be "I" for Infiltration."

Response: We have corrected this typographical error and now uniformly use "I" to represent Soil Infiltration throughout the manuscript.

6. Section 2.3 (Line 161): "For consistency, 'Configuration 4' should be labeled: 'Rainfall (P) + Soil infiltration (I) + Pipe drainage (D)'."

Response: Accepted. In Table 4 and the relevant text, Combination 4 has been standardized to: "Precipitation (P) + Soil Infiltration (I) + Pipe Drainage (D) → Inundation Area (Y)".

7. Line 172: "Specify the unit for sequence length. Are these time steps or hours?"

Response: The temporal sampling frequency is once per minute; the LSTM sequence length of "30" represents 30 time steps (30 minutes). This has been clarified in the manuscript.

8. Section 3.3 (Line 332): Data Leakage Prevention? The manuscript mentions the use of overlapping sliding windows. It is crucial to clarify whether the data split (train/val/test) was performed before or after generating these windows. If done after, there is a high risk of information leakage, which would invalidate the reported generalization performance.

Response: This is an important concern regarding data leakage. Our data partitioning logic is as follows: First, we randomly sample independent storm events from the total sample space; we then perform sliding window interception within each selected independent event; finally, we employ 5-fold cross-validation. This ensures that training and validation sets remain independent. The 6-hour recession period between each event further guarantees physical independence.*within* each selected independent event; finally, we employ 5-fold cross-validation. This ensures that during validation, the training and validation sets not only remain independent during random shuffling but also feature a partitioning mechanism that strictly prevents future information leakage. Additionally, the 6-hour recession period between each event ensures absolute physical independence between the events.

9. Section 4.1 (Line 402): The claim that performance decays after Level 4 is difficult to discern in the current figure for low-intensity and mixed-intensity.

Response: In the revised manuscript, we have corrected the term "decays," describing it more accurately as a "saturation effect" or "plateau." As shown in Section 3.3.1 and Figure 17, when data volume reaches L4 (~14,400 samples), model performance achieves a qualitative leap. Beyond L4, increasing data to L5 and L6 yields only marginal NRMSE reduction and minimal R^2 improvement—the gains plateau rather than decay.

10. Line 423: Does the model account for manhole overflows? This is a critical factor in urban pluvial flooding.

Response: Yes, our coupled 1D–2D hydrodynamic model fully accounts for manhole overflows. When pipe drainage capacity reaches saturation, excess water overflows to the surface through manholes, and its inundation extent is calculated by the 2D module, as described in Sections 2.2 and 2.3.D_t) reaches saturation, the excess water overflows to the surface through manholes, and its inundation area evolution on the surface is calculated by the 2D module.

Part II: Response to General Comments

1. There is confusion between the use of MANOVA and ANOVA. The description suggests a factorial ANOVA for individual metrics, yet MANOVA is mentioned. Please clarify the exact statistical framework and how the covariance between multiple dependent variables was handled.

Response: We have corrected the statistical terminology in the revised manuscript: we conducted independent Multi-factor Analysis of Variance (Factorial ANOVA) for each dependent variable (training time, NRMSE, R^2) to quantify the main and interaction effects of dataset length, rainfall level, and feature combinations, rather than a joint MANOVA. Results are detailed in Figures 20–22.

2. A major omission is the time required to generate the dataset. Since InfoWorks ICM simulations are computationally intensive, the authors must provide details on the total simulation time, hardware used, and a comparison between the "data investment" time vs. the AI's real-time prediction advantage.

Response: This is a valid point. We have added the hardware specifications (Intel i9 + NVIDIA RTX 3090 GPU) and a comparison between the offline data generation cost and the AI's real-time prediction advantage. The LSTM training time is ~896 seconds (~15 minutes); once the training set is generated offline using the physical model, the AI model achieves second-level real-time predictions during deployment.

3. Figures and Legibility:

Regarding "Y0333333.1" in Figure 7: This was a data label artifact from the original plotting. We have redrawn the legends for Figures 5, 6, and 7 with clear descriptive labels.

Regarding the redundancy and legibility of Figures 8, 9, and 11: We have significantly streamlined and consolidated the cross-validation charts. We removed the old, redundant, and hard-to-read training time graphs, integrating the core 5-fold cross-validation mechanism into Figure 8 (matrix plot) and Figure 9 (flowchart). Simultaneously, the validation performance of each configuration was transformed into highly intuitive box plots (Figures 10, 11, and 12), clearly displaying data distributions and means.

Regarding the "Value" axis in Figure 17: We have explicitly relabeled the vertical axes of Figure 17. The vertical axes of the three subplots are now clearly labeled "NRMSE", "R^2", and "Training Time (s)", directly and clearly supporting the subsequent conclusion regarding performance saturation after reaching the L4 data volume threshold.

4. Discussion: The discussion lacks sufficient citations to back its claims, particularly in line 534. The results should be contextualized by comparing them with existing literature on dataset design in computational hydrology.

Response: We have enriched the discussion and literature review, comparing our findings with existing research on dataset design in computational hydrology—including recent work on extreme event prediction and rainfall distribution in deep learning, which further supports the need for stratified mixed sampling under data-scarce conditions.

Conclusion:

We appreciate your recognition of the value of our research in "shifting the focus from model architecture to dataset design." By supplementing the hydrodynamic model validation, clarifying the data split mechanism, correcting statistical terminology, and optimizing figure legibility, we have addressed concerns regarding computational feasibility, data handling, and hydrological physical basis. We hope the revised manuscript meets the journal's publication requirements.