**Response to Anonymous Referee #1's comments:**

This study evaluates the impact of spatial resolution on hourly flood simulation accuracy in large watersheds. It identifies key flood characteristics that influence multi-grid simulation performance and provides practical guidance for selecting spatial resolutions. The study is interesting, and the methodology offers valuable insights. However, there are still some points that should be addressed to improve the quality of the paper.

We sincerely appreciate your positive, insightful and constructive comments. We fully agree with all the concerns you raised and have made the revisions accordingly.

**Major comments:**

1. Section 2.3.2 selects flood characteristics potentially sensitive to grid resolution, but the criteria for selecting these characteristics are unclear. Additionally, the manuscript lacks detailed explanation of how these indicators are calculated. It is recommended to provide the selection criteria and explain the calculation process in detail, which would help improve the clarity and credibility of the paper.

We thank the reviewer for the constructive comments. In the revised manuscript, by citing existing studies on flood characteristics, we have clearly stated the sources and selection criteria for these indicators in Section 2.3.2. The detailed calculation processes for all indicators have been provided in Appendix A. We have also checked and refined the calculation methods, and clarified the source of each indicator with appropriate citations.

**Line 170:** *"These indicators, derived from existing studies (Wang et al. 2021; Liu et al. 2022; Zhang et al. 2025), effectively represent rainfall, underlying surface conditions, and watershed characteristics. Descriptions of these indicators are provided in Table 3, with detailed calculation methods given in Appendix A."*

2. The manuscript uses partial dependence plots (PDPs) for feature-effect interpretation. Since SHAP values are more commonly used for nonlinear analysis, the authors should explain why they chose PDPs instead of SHAP.

Thank you for your valuable comment. (1) In the revised manuscript, we have clarified the reasons for selecting PDPs instead of SHAP to demonstrate the nonlinear impacts of key flood features. As a global interpretation method, PDPs provide a clear and intuitive visualization of the marginal effects of each flood characteristic on model predictions. Since our study focuses on revealing the global average effects of flood characteristics, PDPs are a more direct choice. Moreover, PDPs have been widely used in previous studies to investigate the nonlinear impacts of key flood features. We have included the relevant references in the manuscript (Wang et al., 2026; Yao et al., 2026). (2) In Section 4.3, we have added a discussion on previous studies that employed the SHAP method to identify the controlling factors of spatial resolution.

Notably, the key controlling factors identified by the PDPs and SHAP methods show a high degree of consistency.

**Lines 205-210:** *"Since our study focuses on revealing the global average effects of flood characteristics, partial dependence plots (PDPs) are adopted as a global interpretation method. Compared with the SHAP method, PDPs provide a clearer and more intuitive visualization of the marginal effect of each flood characteristic on model predictions. Moreover, PDPs have been widely used in previous studies to investigate the nonlinear impacts of key flood features (Wang et al., 2026; Yao et al., 2026). Therefore, PDPs were chosen to visualize how each flood characteristic affects the IMP."*

**Lines 330-335:** *"Moreover, Luo et al. (2025) identified rainfall, elevation, and slope as key drivers of flash floods using the SHAP method, and these factors show strong consistency with the key controlling factors of spatial resolution identified in this study."*

3. In section 2.3.4, three different criteria are used to determine the optimal spatial resolution. What is the purpose and rationale behind setting different scenarios? This should be explained in the manuscript.

Thank you for the comment. In the revised manuscript, we have clarified the rationale for setting three different scenarios to determine the optimal spatial resolution. The purpose of setting three different schemes is to select the optimal spatial resolution from both accuracy and computational efficiency perspectives, as the optimal resolution is not unique. In the S1 scheme, the optimal spatial resolution is determined solely based on flood simulation accuracy. In contrast, the S2 and S3 schemes, while satisfying predefined accuracy thresholds (e.g., NSE > 0.8 for S2 and relative error ≤ 5% for S3), prioritize selecting coarser grids to improve computational efficiency. Observing the differences in the optimal spatial resolutions obtained from these schemes provides valuable insight for selecting appropriate spatial resolutions in large-watershed modeling.

**Lines 215-220:** *"The purpose of setting three different schemes is to select the optimal spatial resolution from both accuracy and computational efficiency perspectives, as the optimal resolution is not unique. In the S1 scheme, the optimal spatial resolution is determined solely based on flood simulation accuracy. In contrast, the S2 and S3 schemes, while satisfying predefined accuracy thresholds (e.g., NSE > 0.8 for S2 and relative error ≤ 5% for S3), prioritize selecting coarser grids to improve computational efficiency. Observing the differences in the optimal spatial resolutions obtained from these schemes provides valuable insight for selecting appropriate spatial resolutions in large-watershed modeling."*

4. Results and Discussion section has limited discussion or reference to other studies. For example, Section 4.3 lacks a discussion of the results from previous studies on grid controlling factors. Section 4.6 should also include a discussion of how previous studies have analyzed the

impact of rainfall gauge density on the accuracy of different grid resolutions.

This is a very useful comment. In the revised manuscript, we have extensively discussed the results by citing previous studies:

(1) In Sections 4.1, we have added discussions on flood simulation accuracy at different spatial resolutions.

**Lines 255-260:** *"Aerts et al. (2022) also observed this phenomenon using the daily CAMELS dataset: finer spatial resolution does not bring significant improvement at the watershed outlet. In contrast, for stations with smaller drainage areas (BZ, MB, HF, YC, ZJC, and JG), refining the spatial resolution substantially enhances the representation of runoff generation and flow routing, thereby significantly improving simulation accuracy (Modi et al., 2025)."*

**Line 275:** *"Small watersheds are highly sensitive to localized rainfall, steep gradients, and rapid flow routing, leading to strong nonlinear responses (Wilkinson and Bathurst, 2018). Consistent with these findings, Jiang et al. (2025) also reported that higher-resolution CaMa-Flood models improve daily peak-flow simulation, with the most pronounced improvements occurring in smaller tributaries."*

(2) In Section 4.2, we have added a discussion on computational time.

**Lines 290:** *"Aerts et al. (2022) also found that computational time does not scale linearly with the number of grid cells when running models at different spatial resolutions on high-performance computing equipment."*

(3) Section 4.3 on Grid Controlling Factors: Previous studies on spatial resolution in hydrological modeling have highlighted key flood characteristics, such as watershed area, topographic features, and rainfall distribution, that significantly influence model performance. In the revised manuscript, we have incorporated a detailed discussion of these studies, which strengthens the credibility and applicability of our research findings.

**Line 320:** *"For all four IMP metrics, the trained XGBoost model achieved average recall values of 78.3%–97.8% on the training sets and 64.9%–73.0% on the test sets, with corresponding AUC of 0.78–0.98 (training) and 0.65–0.76 (test). These results indicate satisfactory fitting capability and acceptable generalization performance, comparable to the AUC range (0.75–0.83) reported in similar XGBoost-based classification studies (Kumar et al., 2026)."*

**Line 325:** *"Previous studies have also confirmed the importance of drainage area for the selection of spatial resolution by evaluating the improvement in accuracy achieved through spatial resolution refinement across different drainage areas (Aerts et al., 2022; Jiang et al., 2025)."*

**Lines 330-335:** *"Moreover, Luo et al. (2025) identified rainfall, elevation, and slope as key drivers of flash floods using the SHAP method, and these factors show strong consistency with the key controlling factors of spatial resolution identified in this study."*

(4) Section 4.6 on the effect of rainfall gauge density: There is substantial research on the impact of rainfall input spatial resolution, such as Michelon et al. (2021), Pan et al. (2024), and Huang et al. (2019), all of which recognize the significant impact of spatial resolution on flood simulation accuracy. In the revised manuscript, we have discussed how previous studies have analyzed the impact of rainfall gauge density on model performance at different grid resolutions.

**Lines 445-450:** *"Huang et al. (2019) also found that, in watersheds with sparse gauge networks, refining spatial resolution yields model performance similar to that of the lumped model. Furthermore, Ziaee and Abedini (2023) also emphasized that rain gauge density is a key factor affecting peak flow, and insufficient monitoring station density often leads to underestimation of the peak flow."*

**Lines 465-470:** *"Pan et al. (2024) demonstrated that coarser spatial resolutions of rainfall data degrade simulation accuracy. Similarly, Michelon et al. (2021) showed in a small Alpine catchment that high-density rainfall observations are essential for capturing key hydrological response characteristics, including runoff volume and peak flow timing."*

**Minor comments:**

1. Figure 1: Please include the control areas for all hydrological stations.

Thank you for the suggestion. We have updated Figure 1 to include the control areas (i.e., watershed boundaries) for all ten hydrological stations.

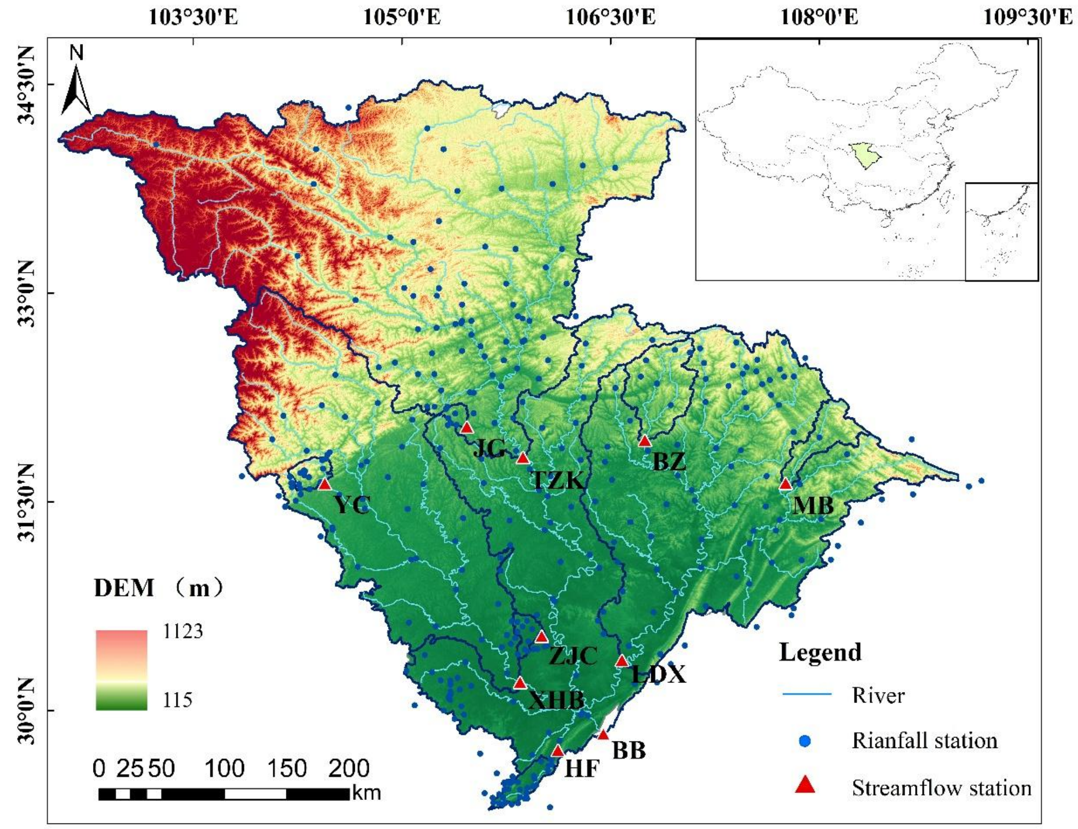

Figure 1. Location of the study area and the distribution of rainfall and streamflow stations.

2. Lines 100-105: Add the sources of the DEM and soil data. Additionally, please provide information on the number of reservoirs and their storage capacities within the study area.

We have added the sources of the DEM and soil data in Section 2.1 “Study Area and Data” of the revised manuscript. In addition, we have provided information on the number of reservoirs and their total storage capacities within the study area, along with an analysis of their impacts.

**Line 100:** *“A 30 m Digital Elevation Model (DEM) was obtained from the United States Geological Survey (https://earthexplorer.usgs.gov/). The 1 km soil dataset was sourced from the Food and Agriculture Organization (Batjes, 1997)."*

**Line 105:** *“The watershed contains 3,806 small and medium-sized reservoirs, with a total storage capacity of 3.89 billion* $m^3$*, equivalent to a runoff depth of 24.7 mm over the entire basin,mainly distributed in the middle and lower reaches."*

3. Figure 2: Is there a reference for this figure? It would be helpful to add one.

We have added a reference for Figure 2 in the revised manuscript, specifying the source and citation of the model.

4. Section 2.3.1: The title “Equations” may not clearly convey the content of the section. It is recommended to revise the title to “Evaluation of GDHF Model Performance at Different

Spatial Resolutions” for greater clarity and alignment with the section's focus.

Thank you for this constructive suggestion. In the revised manuscript, we have changed the title of Section 2.3.1 to “Evaluation of GDHF Model Performance at Different Spatial Resolutions” as recommended.

5. Line 190: The statement "spatial refinement yields a significant improvement (IMPNSE > 0.10, IMPBIAS > 5%, IMPRPE > 5%, or IMPPTE > 1 h)" requires further clarification. The rationale behind selecting these specific thresholds should be provided.

Thank you for this important comment. In the revised manuscript, we have clarified the rationale behind these threshold selections. Minor improvements in IMP metrics may arise from noise or random fluctuations; simply labeling all positive IMP changes as 1 could lead to unstable model training and reduced generalization ability. Therefore, specific thresholds ($IMP_{NSE} > 0.10$, $IMP_{BIAS} > 5\%$, $IMP_{RPE} > 5\%$, or $IMP_{PTE} > 1$ h) are essential for identifying flood events where spatial grid refinement yields substantial improvements in model accuracy. These thresholds are based on the common ranges and practical significance of the metrics in hydrological modelling: an increase in NSE of more than 0.10 is considered a significant improvement; a reduction in BIAS and RPE of more than 5% represents a clear decrease in error; and a reduction in PTE of more than 1 h has practical significance for the timeliness of flood warnings.

**Lines 195-200:** *“Therefore, the IMP metrics were converted into binary labels: label = 1 if spatial refinement yields a significant improvement, otherwise, label = 0 (Ekmekcioğlu and Koc, 2022). Because minor improvements in IMP metrics may arise from noise or random fluctuations, labeling all positive changes as 1 could lead to unstable model training and poor generalization (Siam et al., 2022). This study adopts specific thresholds ($IMP_{NSE} > 0.10$, $IMP_{BIAS} > 5\%$, $IMP_{RPE} > 5\%$, or $IMP_{PTE} > 1$ h) to identify flood events where spatial grid refinement yields significant improvements in model accuracy and performance. These thresholds are based on the common ranges and practical significance of the metrics in hydrological modelling."*

6. Figure 4: Please ensure that it clearly specifies that the modeling process is based on a 10 km resolution.

Thank you for pointing this out. In the revised manuscript, we have specified that the modeling process in Figure 4 is based on a 10 km spatial resolution as an example.

7. Figure 12: The y-axis should be clearly labeled to indicate whether it represents the results of grid resolution selection or the precision of different grid resolutions.

Thank you for the suggestion. In the revised manuscript, we have modified Figure 12 to clearly display the results of optimal spatial resolution selection under different scenarios.

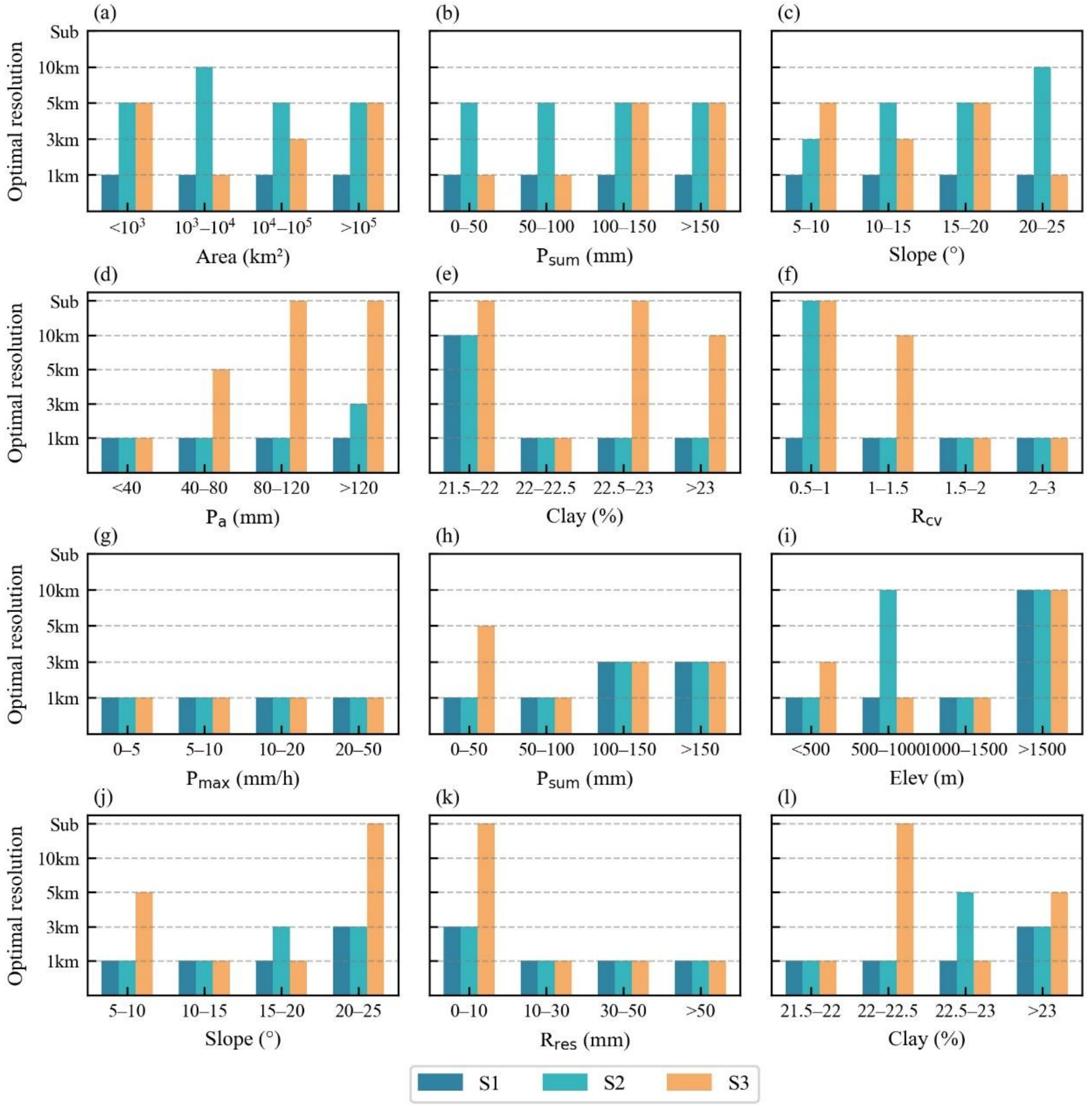

Figure 12. Spatial resolution recommendations for different flood characteristic categories based on the three schemes.

8. Appendix A: Why are references not included in Appendix A? Any references used to describe details and equations in the appendix should be cited.

Thank you for this comment. We have added appropriate citations in Appendix A to reference the sources of the flood characteristic indicators and the equations used to describe their calculation principles.

**Line 505:** *"The total rainfall $P_{sum}$ and the average rainfall intensity $P_{avg}$ represent the total rainfall and the average rainfall over all time steps, respectively (Zhou et al., 2021)"*

**Line 510:** *"The rainfall spatial non-uniformity coefficient $\eta$ and the temporal variation coefficient $R_{cv}$ (Wang et al., 1999) are used to describe the spatiotemporal distribution characteristics of rainfall:"*

**Line 520:** *"The antecedent precipitation $P_a$ quantifies the soil moisture condition prior to a rainfall event and can be used to analyze the watershed's nonlinear response and the runoff generation process (Schoener and Stone, 2019):"*

**Line 525:** *"These indicators characterize the topographic features and storage conditions of the watershed (Ma et al., 2021)."*

**References:**

Aerts, J.P.M. et al., 2022. Large-sample assessment of varying spatial resolution on the streamflow estimates of the wflow_sbm hydrological model. Hydrology and Earth System Sciences, 26(16): 4407-4430.

Batjes, N.H., 1997. A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling. Soil Use and Management, 13(1): 9-16.

Huang Y, Bárdossy A, Zhang K.: Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, Hydrology and Earth System Sciences, 23(6), 2647-63.

Jiang, R., Lu, H., Yang, K., Cho, H. and Yamazaki, D., 2025. Analysis and comparison of the flood simulations with the routing model CaMa-Flood at different spatial resolutions in the CONUS. Environmental Modelling & Software, 185: 106305.

Wang, W.Z., Jiao, J.Y. and Hao, X.P., 1999. Nonuniformity of spatial distribution of rainfall and relationship between point rainfall and areal rainfall of different patterns of rainstorm on the Loess Plateau. Adv Water Sci (in Chinese), 10(2): 165-169.

Kumar, A., Pandey, G. and Kale, R.V., 2026. Ensemble machine learning and deep learning framework for flood susceptibility mapping in the transboundary Rapti River Basin. Environmental Earth Sciences, 85(7): 168.

Liu, Y., Li, Z., Liu, Z. and Luo, Y., 2022. Impact of rainfall spatiotemporal variability and model structures on flood simulation in semi-arid regions. Stochastic Environmental Research and Risk Assessment, 36(3): 785-809.

Luo, L. et al., 2025. Exploration of the spatiotemporal characteristics and triggering factors of flash flood in China. Ecological Indicators, 176: 113698.

Ma W, Ishitsuka Y, Takeshima A, et al. 2021. Applicability of a nationwide flood forecasting system for Typhoon Hagibis 2019. Scientific reports, 11(1), 10213.

Michelon, A., Benoit, L., Beria, H., Ceperley, N. and Schaefli, B., 2021. Benefits from high-density rain gauge observations for hydrological response analysis in a small alpine catchment. Hydrology and Earth System Sciences, 25(4): 2301-2325.

Pan, X. et al., 2024. Study on the Influence of Temporal and Spatial Resolution of Rainfall Data on Watershed Flood Simulation Performance. Water Resources Management, 38(8): 2647-2668.

Schoener, G. and Stone, M.C., 2019. Impact of antecedent soil moisture on runoff from a semiarid catchment. Journal of Hydrology, 569: 627-636.

Siam, Z. S., Hasan, R. T., and Rahman, R. M. (2022). Effects of Label Noise on Regression Performances and Model Complexities for Hybridized Machine Learning Based Spatial Flood Susceptibility Modelling. Cybernetics and Systems, 53(4), 362–379.

Wang, Q., Xu, Y., Cai, X., Tang, J. and Yang, L., 2021. Role of underlying surface, rainstorm and antecedent wetness condition on flood responses in small and medium sized watersheds in the Yangtze River Delta region, China. CATENA, 206: 105489.

Wang, T., Fu, Z. and Luo, M., 2026. Deep learning for decoding climate–urbanization synergies: flood susceptibility forecasting under SSP-RCP scenarios in Beijing, China. Journal of Hydrology, 672: 135364.

Yao, Y., Fu, G. and Webber, J., 2026. How can hydrological connectivity inform catchment scale stormwater flood management? Water Research, 297: 125767.

Zhang, J. et al., 2025. A study on the effect of spatially variation rainfall on urban flooding. Geomatics, natural hazards and risk, 16(1).

Zhou, Z. et al., 2021. The impact of the spatiotemporal structure of rainfall on flood frequency over a small urban watershed: an approach coupling stochastic storm transposition and hydrologic modeling. Hydrology and Earth System Sciences, 25(9): 4701-4717.

Ziaee, P. and Abedini, M.J., 2023. Investigating the Effect of Spatial and Temporal Variabilities of Rainfall on Catchment Response. Water Resources Management, 37(13): 5343-5366.

**Response to Anonymous Referee #2's comments:**

This study systematically evaluates the impact of spatial resolution on hourly flood simulation accuracy in large watersheds by integrating distributed hydrological modeling with machine learning–based analysis. The authors employ the GDHF model across multiple spatial resolutions and introduce an improvement index (IMP) framework to quantify performance gains relative to a sub-watershed scheme. By coupling model simulations with XGBoost and partial dependence analysis, the study identifies key controlling factors governing resolution sensitivity and reveals their nonlinear influence on simulation performance. The framework is further strengthened through experiments under varying rainfall station densities, providing insight into the limitations of spatial refinement under data-sparse conditions. The results demonstrate that finer resolutions significantly enhance simulation accuracy in smaller basins, while offering limited gains in large drainage areas, and highlight the dominant role of basin area and rainfall characteristics in determining optimal resolution. Overall, the manuscript presents the methodology that is well designed and aligned with the study objectives, with a clear framework and appropriate integration of hydrological modeling and machine learning. The results effectively address all objectives, providing a coherent evaluation of spatial resolution effects, controlling factors, and rainfall station density impacts. The conclusions are consistent with the results and accurately summarize the key findings without overinterpretation. The paper is well organized; however, some minor issues related to sentence structure and presentation consistency should be addressed to further improve readability and precision, as outlined below.

Thank you very much for taking the time to thoroughly review our manuscript and provide such detailed and constructive comments. We are pleased that you found the methodology well designed and the manuscript well-organized. We have carefully addressed all the minor issues you raised and make the revisions accordingly. Below, we provide our point-by-point responses.

**Minor Comments**

- In the Introduction section (lines 50–55), "key characteristic" should be revised to "key characteristics" to match the plural context.

  Thank you for pointing this out. We have revised the term to "key characteristics" accordingly in the revised manuscript.

- In the Introduction section (line 85-90), "pronounced topographic relief" is slightly repetitive with "hilly and mountainous terrain" mentioned earlier and could be revised for conciseness, and control areas mean "drainage areas" or "catchment areas"?

  We appreciate your careful reading. To avoid redundancy, we have removed the phrase "pronounced topographic relief" as suggested. In addition, we have replaced “control areas” with “drainage areas” for consistency and clarity.

- In the Data and Methods section (line 95-100), “km2·gauge-1” should be formatted using proper superscripts (e.g., km$^{2}$·gauge$^{-1}$) for consistency and clarity.

  We have carefully checked the formatting of all units throughout the manuscript and ensured that superscripts are used correctly (e.g., km$^{2}$·gauge$^{-1}$) to maintain consistency and clarity.

- In the Methods section (line 125-130), “Li et al. (2024” is missing a closing parenthesis and should be corrected to “Li et al. (2024)”.

  We have verified all citation formats throughout the manuscript and corrected any missing parentheses or other formatting issues, including the example noted by the reviewer.

- In Appendix A (line 470-475), there is a typo in “P(t) epresents”; it should be corrected to “P(t) represents”.

  We have corrected the typo to “P(t) represents” in Appendix A. We have also carefully reviewed the entire manuscript to avoid similar errors.

- In Appendix A (around line 515), the equation for Shape (A15) is repeated; one instance should be removed to avoid redundancy.

  Thank you for your detailed review. We have removed the duplicate equation as suggested.