

Response to Anonymous Referee #2

We highly appreciate the reviewer for the constructive and detailed comments. We have carefully considered all comments and have outlined our responses and the corresponding revisions to the manuscript below.

Major Comments:

1. The authors selected data from May to September for analysis (or was this criterion solely used for station screening? Please clarify). While this period encompasses the high-incidence season for flash floods, this strategy inevitably excludes spring snowmelt floods and autumn/winter storm floods. Such seasonal truncation may lead to the oversight of specific hydrological processes. It is suggested that the authors elaborate in the Discussion section on the potential impact of this data screening strategy on the generalizability of the conclusions.

Response 1: We appreciate these comments. In the revised manuscript, we have addressed this concern through two modifications. In Section 3, we have explicitly clarified that the May-September time window was used as a criterion for screening station data integrity, not as a temporal filter applied to the analysis itself. Given the difficulty of obtaining complete hourly streamflow records for the entire year, data completeness was assessed over the May-to-September period, when flash flood activity peaks across much of CONUS. Once a station satisfied this criterion, all available data across its full period of record were used for trend analysis, so cool-season flood events at retained stations are included. Additionally, we have expanded Section 5 to include a discussion of the limitations associated with this screening approach.

2. The study utilizes NLDAS-2 precipitation data with a resolution of approximately 12 km. For the many small catchments mentioned in the text (e.g., L213, with an area of 78.8 km²), this resolution implies that the catchment is covered by very few grid points, which makes it difficult to capture the spatial variability of localized convective precipitation events that cause flash floods. I recommend that the authors discuss the influence of this limitation on simulation accuracy.

Response 2: We appreciate these comments. The 0.125° resolution of NLDAS-2 does provide limited grid coverage for the smallest catchments, which may smooth localized convective peaks and introduce uncertainty in basin-averaged precipitation estimates. However, the DRIVE model experiments isolating temperature and land cover effects employ identical precipitation forcing across scenarios, so the relative attribution conclusions are unaffected by precipitation resolution. We have discussed this limitation in Section 5.

3. Figure 5 indicates that some catchments are dominated by “Temperature,” and the text attributes the flood attenuation caused by warming to “rising temperature and evapotranspiration.” The current explanation is somewhat generalized. Please further clarify the specific physical mechanisms involved in this process.

Response 3: We appreciate these comments. We have expanded the mechanistic explanation in Section 4.4. Rising air temperature elevates the vapor pressure deficit, enhancing evapotranspiration and progressively depleting antecedent soil moisture. The resulting increase in available soil storage at storm onset allows more rainfall to infiltrate rather than generate direct runoff, attenuating flood peaks. This pathway is corroborated by the widespread ASM decline documented in Section 4.2. DRIVE sensitivity experiments confirm that 67% of catchments experience a 0 to 6% flow reduction under observed warming (Section 4.5, Fig. 7). This attenuation is most effective for moderate events. Additionally, for extremes above the 90th percentile, rainfall intensities far exceed infiltration capacity regardless of antecedent conditions, limiting the storage effect on peak flows (Section 4.5, Fig. 6).

4. Figure 5 displays the spatial distribution of the primary drivers. It would be beneficial to supplement this with a chart (such as a stacked bar chart or table) to quantitatively present the proportion of catchments dominated by distinct single drivers versus combined drivers (e.g., P&T, P&Land Cover), along with a brief discussion on the spatial distribution characteristics of these combined driving patterns.

Response 4: We appreciate these comments. Table S1 has been added to the Supplementary Material, listing the count and percentage of catchments for each driver category. Of the 294 catchments, 69.0% are governed by a single driver and 29.3% by combined drivers (Table S1). P and P&T together account for 90.1%, all other combinations individually represent less than 2%. We have also expanded Section 4.4 to describe the spatial contrast between these two dominant categories: among western CONUS catchments, P&T co-dominance is more common than in the East (36.4% versus 26.1%), whereas P-dominated catchments are proportionally more concentrated in the humid East (49.5% versus 39.0%; Fig. 5).

5. The abstract mentions that LUCC has remained stable over the past 40 years, which may be attributed to the fact that the 294 selected stations are mostly located in upstream areas with less human interference. Please add a restrictive description of the study station attributes in the abstract. Furthermore, the statement that “our results show 10% increase in impervious surfaces could lead to 20% peak flow increase” appears to be based on specific cases like Peachtree Creek rather than the general analysis. The wording in the abstract needs adjustment to explicitly state that this conclusion applies to specific contexts (e.g., rapidly urbanizing areas).

Response 5: We appreciate these comments. We have made the corresponding revisions to the abstract in the revised manuscript.

6. The current description of the Random Forest regression is too brief. Key details, such as hyperparameter settings and the division of training/testing sets, should be added to the Methodology section.

Response 6: We appreciate these comments. We have expanded Section 2.4 to provide the implementation details necessary for reproducibility. Specifically, we now specify that data were split into 80% training and 20% validation sets for each catchment.

Hyperparameters were optimized per catchment using the TPE algorithm over 100 evaluations, with the validation set R^2 as the objective function. The tuned hyperparameters include `n_estimators` (50-500), `max_depth` (None or 10-40), `min_samples_split` (2-20), and `min_samples_leaf` (1-10).

In addition, we revised the variable importance screening procedure. The original fixed 15% threshold lacked an objective basis. In the revision, variables are ranked by mean decrease in impurity (MDI) and a cumulative importance curve is constructed for each catchment. The inflection point, defined as the rank of maximum absolute second-order difference, marks the transition from dominant to marginal contributors. As reported in Section 4.4 of the revised manuscript, the median inflection point across 294 catchments falls at approximately 22% cumulative importance. The 15% threshold is more stringent than all observed inflection points and thus retains only the most prominent drivers in every catchment.

7. The TS-DUH model operates at a 90 m resolution, while the DRIVE model runs at 0.125°. How was this significant resolution discrepancy resolved during the coupling process? In particular, the processing of land cover data within DRIVE when extracting vegetation parameters requires a more detailed explanation.

Response 7: We appreciate these comments. The coupling between the two models is one-directional. DRIVE first generates gridded runoff at 0.125° resolution, and these runoff fields are then uniformly disaggregated to the 90 m TS-DUH grid. The TS-DUH model subsequently routes these runoff contributions along the high-resolution flow network, using detailed topographic and land cover information to estimate flow velocities and travel times.

Regarding the derivation of vegetation parameters within DRIVE, we have expanded the description in Section 2.3 of the revised manuscript. Specifically, the 30 m land cover data were first reclassified to match the DRIVE land cover categories via a cross-walking table, then spatially aggregated to compute fractional coverage within each 0.125° grid cell. These fractions were used to derive area-weighted vegetation parameters, such as leaf area index, albedo, and roughness length from the VIC parameter library.

8. The study analyzes the long-term trend of soil moisture (ASM) and finds a downward trend. However, in the rainfall analysis, the Random Forest utilizes the annual mean ASM. For events like flash floods, the soil moisture prior to the event is a more direct factor determining runoff response. It is recommended that the authors explore the role of “antecedent conditions” more deeply in the Discussion/Results section.

Response 8: We appreciate these comments. We acknowledge that our attribution framework operates at the annual scales. We recognize that ASM cannot fully capture the role of antecedent soil moisture in runoff generation. In the revised Discussion, we now discuss this limitation. We further note that event-scale attribution linking

antecedent moisture windows to individual flood responses warrants further investigation.

9. A paragraph regarding model uncertainty could be added to the Discussion section. For example, discuss the uncertainty of DRIVE model parameters and the potential impact of selecting different atmospheric forcing data (e.g., NLDAS-2 vs. MERRA-2) on the results.

Response 9: We appreciate these comments. We have revised the relevant sentence in the Section 5 to explicitly acknowledge both the inherent limitations of the hydrological models and the uncertainties associated with different data sources.

Minor Comments

Please unify the format of separators between keywords.

Response 10: We appreciate these comments. In the revised manuscript, we have standardized the format of separators between keywords.

In the abstract, please specify the exact temporal precision of the “sub-daily data.” Additionally, indicate the statistical test method and significance level used for the reported 17.3% increase in flash flood frequency.

Response 11: We appreciate these comments. We have replaced “sub-daily” with “hourly” and added the significance level ($p < 0.05$) in the abstract. Detailed descriptions of the Mann-Kendall test and Poisson regression remain in Section 2.2.

When citing Ivancic and Shaw (2015), I suggest adding recent literature regarding “Flash Drought leading to Flash Flood” to provide a more comprehensive context.

Response 12: We appreciate these comments. We have added a sentence and recent literature regarding the rapid transition from flash droughts to flash floods in the revised manuscript.

There is a discrepancy between Line 334 (90.8%) and Line 361 (90.3%) regarding the percentage of observation points with discharge changes within $\pm 3\%$. Please verify.

Response 13: We appreciate these comments. We apologize for the typographical error in the original manuscript. We have verified our statistical results and confirmed that the correct proportion of catchments with discharge changes within $\pm 3\%$ is indeed 90.8%.

Provide the full name or a brief description when “GLC_FC30” first appears.

Response 14: We appreciate these comments. We have expanded the first mention of “GLC_FC30” in Section 3.

The manuscript totals 25 pages, with the reference section accounting for 6 pages. authors are recommended to streamline content, focusing on optimizing references to balance core research and citations.

Response 15: We appreciate these comments. We have systematically streamlined the references in the revised manuscript.