Reviewer Comments to Authors

Manuscript: Flood estimation in natural and urban catchments using hydrological simulation and Bayes theorem

Recommendation: Major Revision

General Comments

The manuscript introduces a potentially valuable design-flood framework that integrates the continuous DDD rainfall-runoff model, the stochastic DDDEvent model, and Bayes' theorem. The primary contribution is the advancement beyond the conventional assumption that precipitation and runoff must share the same return period, instead representing variability in antecedent subsurface storage and rainfall sequencing. Comparisons with FFA-GEV and SCHADEX, as well as applications to natural, urban, and ungauged catchments, enhance the study's practical relevance.

The approach demonstrates scientific promise, especially for rainfall-dominated catchments. However, the probability formulation, uncertainty treatment, model validation, and scope of applicability require substantial clarification prior to publication. Additionally, the conclusions should clearly distinguish results substantiated by the present cases from applications that remain preliminary.

1. The application of Bayes' theorem requires clearer theoretical justification.

The manuscript should specify whether the terms in Equation 6 represent probability densities, conditional exceedance probabilities, or event probabilities. While Equations 4 and 5 condition on exact values of P and Q, Equation 6 uses exceedance probabilities and return periods. Furthermore, the conditional distribution of P given Q is derived from uniformly sampled precipitation values rather than a realistic precipitation marginal distribution. It is necessary to demonstrate that Equation 6 is derived from a coherent joint distribution and to assess the impact of employing a fitted extreme-value distribution for precipitation. If this cannot be established, the formulation and interpretation of return periods should be revised.

2. Uncertainty and model validation should be addressed more systematically.

The analysis addresses uncertainty in extreme precipitation but largely maintains other components as fixed. Key sources of uncertainty include DDD calibration, regionalization, critical-duration estimation, the S-given-P relationship, the gamma approximation, hyetograph generation, and model structure. Table 1 presents calibration KGE values, yet lacks clear independent validation, and KGE does not specifically assess extreme peaks. Provide validation or high-flow diagnostics where feasible, particularly for urban catchments with short records, and explicitly state that current estimates are conditional on fixed model structures and parameters.

3. Applicability to snowmelt-affected catchments is limited.

Model performance declines as snowmelt becomes a more significant factor, since the current framework does not fully represent combined rainfall-snowmelt generation. The abstract and methods sections should state at the outset that the framework is primarily applicable to rainfall-dominated catchments. Conclusions should avoid suggesting broader applicability until the framework incorporates snowmelt processes.

4. Validation of the ungauged-catchment application is preliminary.

The manuscript presents only four ungauged tests. Consequently, conclusions regarding transferability should be moderated or supported with additional independent tests. Furthermore, the calibrated applications utilize approximately 32 years of precipitation data, while the PUB applications use 52 years. This discrepancy confounds the comparison; therefore, the authors should evaluate calibrated and regionalized models using identical precipitation inputs, with record-length sensitivity assessed separately.

5. The urban-catchment application needs clearer scope.

The manuscript should specify which engineered urban processes DDD/DDDEvent represents and which it omits, such as storm sewers, connected impervious areas, detention, and inlet limitations. This clarification is particularly important for catchments as small as 0.06–1.9 km^2, which exhibit rapid response times and relatively short calibration records.

6. Antecedent storage and runoff attribution require further support.

The S-given-P relationship is derived from simulated DDD states rather than independent observations, and the gamma approximation should be substantiated with goodness-of-fit or sensitivity analyses. The separation of runoff from antecedent storage and event precipitation should be justified, as subtraction may not yield a unique source attribution in a nonlinear model. The manuscript should clearly distinguish among modeled subsurface storage, soil moisture, and groundwater, and should avoid generalizing the greater-than-50% contribution beyond the two illustrated catchments.

7. The precipitation-sequence generator needs additional validation.

Given that rainfall sequencing alters simulated peaks by approximately 20%, the manuscript should demonstrate that the beta-generated hyetographs resemble observed extreme storms. Additionally, the sampled parameter ranges, event-selection procedure, kernel-density settings, and a convergence check confirming that 2,500 simulations per precipitation value are sufficient should be reported.

8. The comparison metrics require clearer interpretation.

SimScore should be explained in intuitive terms, including its range and the reasons it may diverge from MAPE. The manuscript should clarify whether FFA-GEV is considered a benchmark or an approximation of the true flood-frequency relationship. If MAPE values are averaged across quantiles, the equation should explicitly include the averaging operation.

9. The methodology should be reorganized for clarity.

A concise step-by-step workflow linking DDD simulation, estimation of S given P, DDDEvent simulation, construction of the conditional distributions, application of Bayes' theorem, and estimation of design floods would improve reproducibility and readability.

10. The conclusions should be aligned more closely with the evidence.

The strongest evidence supports rainfall-dominated catchments with reliable precipitation estimates and an adequately calibrated DDD model. Claims concerning urban, ungauged, mixed, and snowmelt-dominated catchments should be described as preliminary and in need of further testing.

11. Very long return-period estimates require additional justification.

Estimating a 1,000-year flood from about 32 years of precipitation data requires substantial upper-tail extrapolation. The manuscript should provide uncertainty intervals, assess sensitivity to the choice of distribution and record length, and explicitly state the stationarity assumption. Conclusions based on the 1,000-year estimates should be appropriately qualified.

Minor Comments

Correct the inconsistent urban MAPE range in the Results and Table 2. Revise the claim that the simulations cover "all possible" states and sequences. Runoff volume should be distinguished from peak discharge. Please correct "Fig. 87," "s26.20," the inconsistent spelling of Røykenes, and the Gringorten reference year. Revise the code-and-data statement to specify the exact datasets, configurations, parameter files, and scripts required for reproduction.