Supplement of Geosci. Model Dev., 19, 3709–3724, 2026
https://doi.org/10.5194/gmd-19-3709-2026-supplement

Geoscientific Model Development
Open Access
EGU

Supplement of

Flood volume allocation method for flood hazard mapping using river model with levee scheme

Muhammad Hasnain Aslam et al.

Correspondence to: Muhammad Hasnain Aslam (hasnain0386@gmail.com)

S1. Frequency Analysis

We performed a frequency analysis on the annual maximum flood storage values simulated by the CaMa-Flood model to estimate flood storage values corresponding to various return periods. Daily floodplain and river channel storage values were aggregated annually (1979–2019), and the annual maximum storage was extracted for each grid cell to construct a time series for extreme value analysis. Fig. S1 presents extreme value analysis showing fitted frequency curves using different distributions for a single simulation grid (unit catchment).

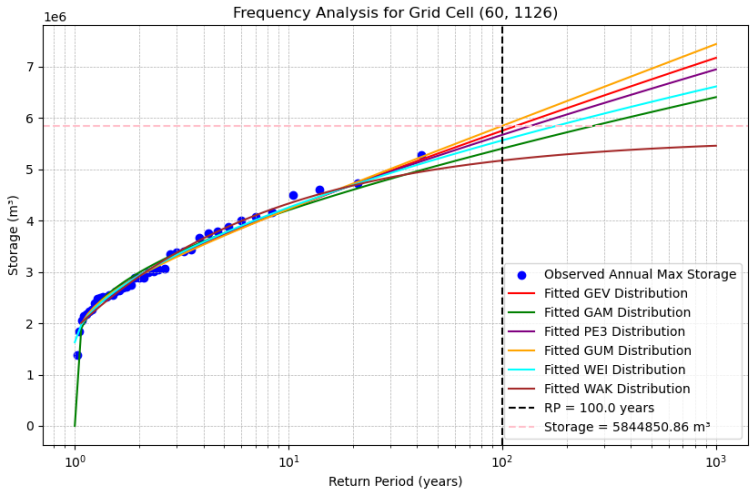
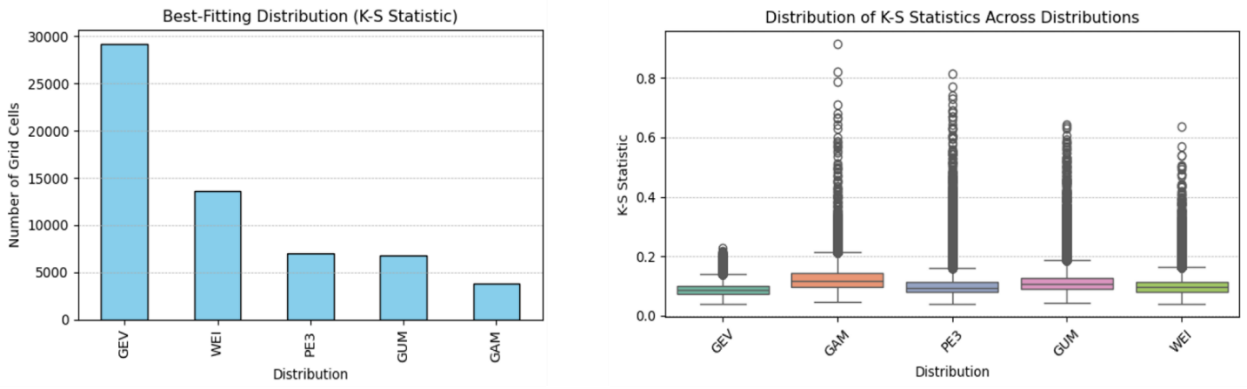


Figure S1: Fitted frequency curves on simulated annual maximum storages for a unit catchment

We tested multiple probability distributions, including Generalized Extreme Value (GEV), Gamma (GAM), Pearson Type III (PE3), Gumbel (GUM) and Weibull (WEI). Each distribution was fitted to the annual maximum storage using the L-moment method (Hosking, 2015).

To ensure robust estimation of return-period flood storage across the domain, we compared the performance of all the probability distributions highlighted above.To evaluate the goodness-of-fit of each distribution, we employed the Kolmogorov–Smirnov (KS) test (Massey, 1951), that measures the maximum absolute difference between the empirical and theoretical cumulative distribution functions. The distribution yielding the lowest KS statistic, was selected as the best-fitting model to determine the storage for various return periods

Fig. S2a Summarizes the distribution of best fits across all evaluated grids. The GEV distribution emerged as the most dominant, being selected in over 29,000 grid cells, substantially more than any other candidate. This strong performance is further reinforced by Fig. S2b, which presents the distribution of KS statistics for each distribution using a boxplot comparison.



a. Distribution of best fits across all evaluated grids

b. KS statistics for each distribution using a boxplot comparison

Figure S2: Performance of different Frequency Distributions in Japan river basins

The GEV distribution consistently yielded the lowest median KS statistic across the domain and exhibited relatively narrow interquartile ranges less outliers, indicating both accuracy and stability. These results collectively support the use of the GEV distribution as the most reliable model for frequency-based flood storage estimation

The final return-period storage maps were generated using the best-fitting distribution, and the results were saved in binary format for downscaling and inundation mapping.

S2. Model Evaluation

S2.1. Discharges

To evaluate the hydrological performance of the CaMa-Flood simulations over Japan, we conducted a comprehensive comparison against daily discharge observations from over 400 stream gauges. First, we allocated the stream gauges in the models' 1-minute resolution map to obtain their XY locations by comparing the upstream map area with the observed areas. Once the stream gauges are allocated on CaMa-Map, the time series of simulated daily discharges for those virtual locations are extracted and compared with the observations.

Model performance was assessed using runoff from the bias-corrected VIC (GRADES) dataset. Scatter plots comparing mean and peak daily discharges (Fig. S3) The model shows strong agreement, achieving correlations of $R = 0.94$ and $R^2 = 0.89$ for mean daily discharges and $R = 0.80$ and $R^2 = 0.64$ for peak daily discharge. RMSE values, represented by color in the plots, indicate generally moderate errors with some peak flows were slightly underestimated.

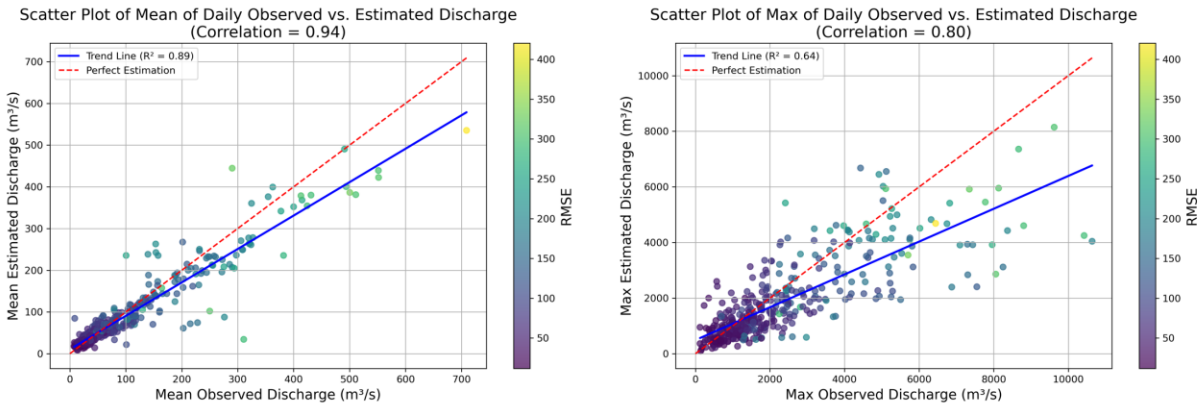


Figure S3: scatter plots comparing Magnitudes of mean and max annual of daily Obs and Simulated discharges

The performance of the simulated discharge across all gauging stations was also assessed using four statistical metrics: Kling–Gupta Efficiency (KGE), Nash–Sutcliffe Efficiency (NSE), Root Mean Square Error (RMSE) and Pearson's R. As shown in Fig. S4. Most stations exhibited KGE values between 0.50 and 0.75, indicating good overall agreement between simulations and observations. NSE values were primarily distributed within the 0.25–0.50 and 0.50–0.75 ranges, suggesting that the model captures flow dynamics reasonably well at most locations. A limited number of stations showed negative NSE, indicating poorer performance for a few stations. RMSE values were concentrated in the 30–70 m³/s range, with higher values observed at a few stations. The majority of stations achieved Pearson's correlation values between 0.45 and 0.75, with R values clustering around R ≈ 0.6. These distributions confirm that the model achieves satisfactory performance in simulating river discharge across Japan, with consistent accuracy in most basins.

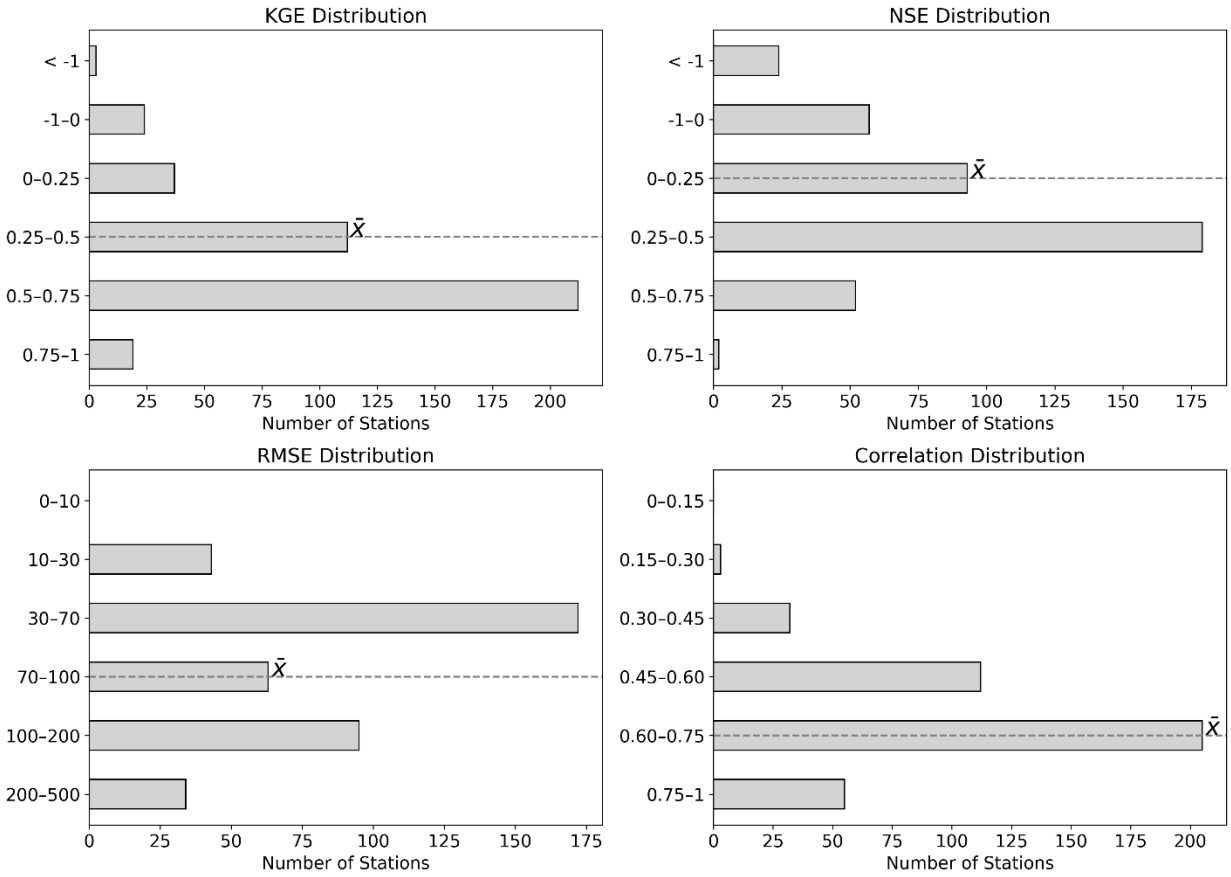


Figure S4: Distribution of Pearson R for simulated daily discharge against observations

Additional spatial distribution maps and sub-scatter plots (Supplementary Fig. S5-S8) show that stations with smaller upstream catchment areas (highland stations) exhibit lower model performance across all metrics. These smaller basins are more sensitive to localized hydrological variability and land-use effects which are often not fully captured by large-scale input data.or the simplified routing scheme. As a result, discrepancies at these stations likely reflect characteristic scale mismatches. While the model performs consistently well in larger basins, results in smaller catchments should be interpreted with greater attention to the local context, especially when applying outputs for flood risk or water management applications.

Daily time series comparisons were also conducted for all stream gauge stations included in this study to ensure comprehensive validation of model performance further. This thorough validation indicates that the model captures seasonal variability and baseflow trends but underestimates peak discharges at certain locations. Despite these peak underestimations at some stations, the model consistently reproduces the temporal flow trend across regions, supporting its utility for large-scale flood hazard assessments.

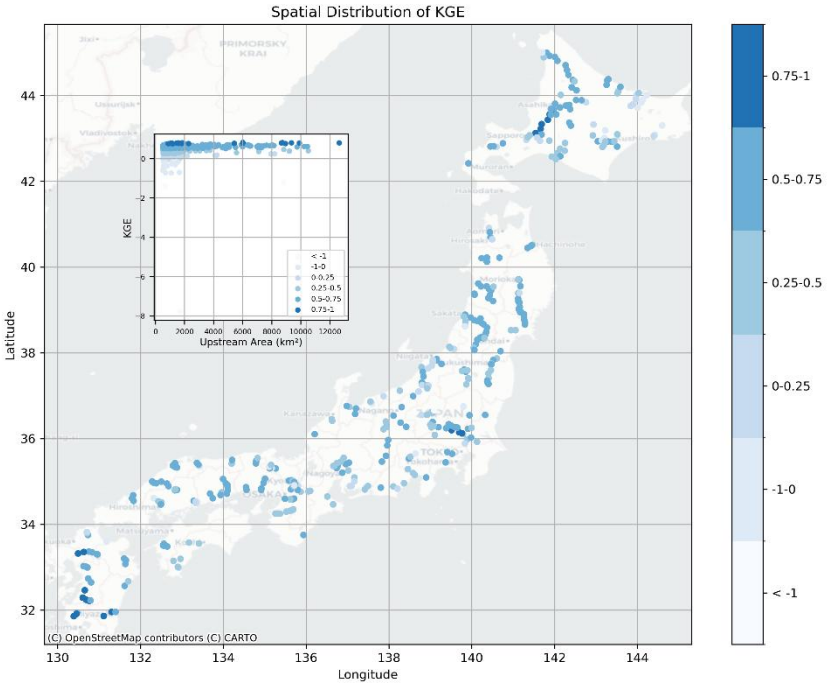


Figure S5: Spatial distribution of Kling-Gupta Efficiency (KGE) values across Japan for daily simulated discharges compared to MLIT observations. Circle colour intensity represents KGE performance, with higher values indicating better agreement. The inset scatter plot shows the relationship between KGE and the upstream catchment area. © OpenStreetMap contributors (https://www.openstreetmap.org/copyright) © CARTO

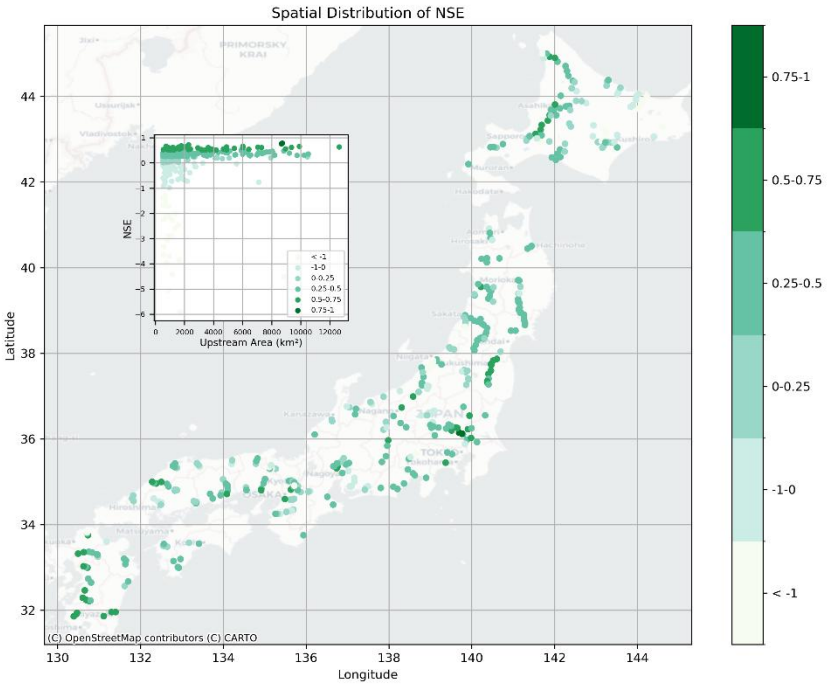


Figure S6: Spatial distribution of Nash–Sutcliffe Efficiency (NSE) for model-simulated daily discharges. The color scale reflects model accuracy, with negative values indicating poor simulation skill. The inset illustrates the trend of NSE values concerning upstream catchment size, showing generally higher NSE in larger basins. © OpenStreetMap contributors (https://www.openstreetmap.org/copyright) © CARTO

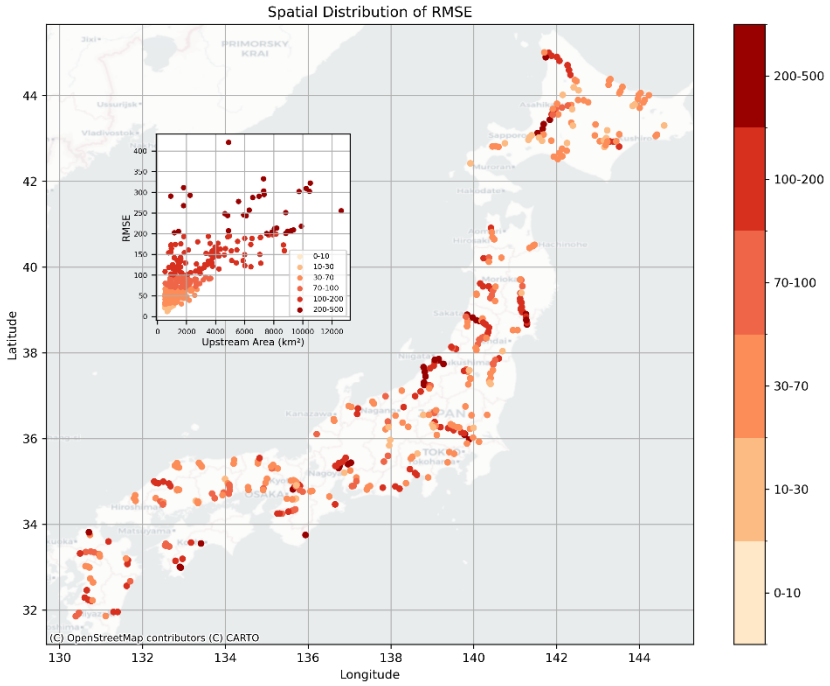


Figure S7: Spatial distribution of RMSE for model-simulated daily discharges. © OpenStreetMap contributors (https://www.openstreetmap.org/copyright) © CARTO

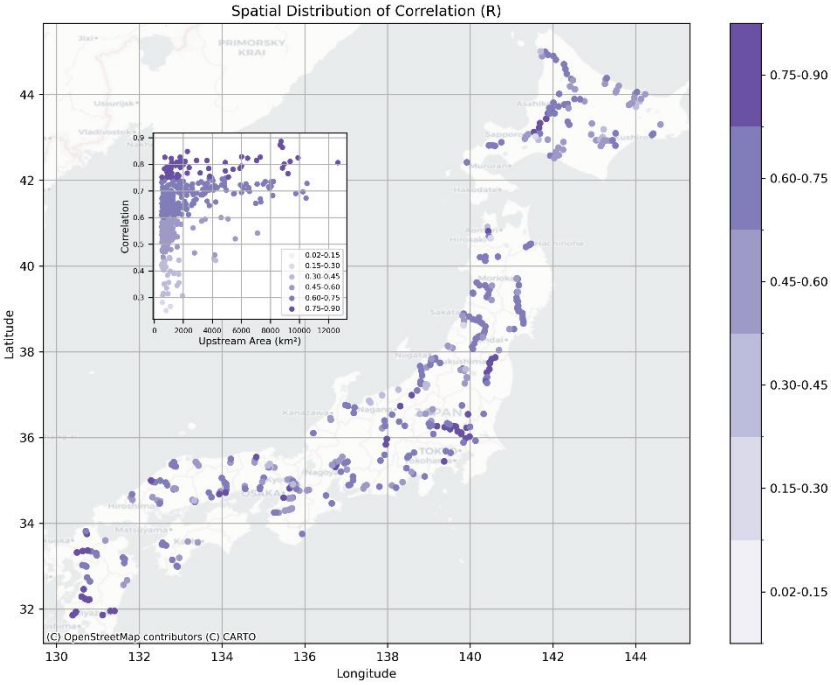


Figure S8: Spatial distribution of Pearson Correlation Coefficient (R) between observed and simulated daily discharges . Higher R values indicate stronger agreement in flow timing. The inset plot shows correlation versus upstream area, reinforcing the challenge of simulating discharges in smaller basins. © OpenStreetMap contributors (https://www.openstreetmap.org/copyright) © CARTO

S2.2. Water Levels

While evaluating the model's performance against observed water levels, a direct mass comparison between simulated and observed heights was not possible. This limitation arises primarily from the lack of riverbed surveys in the topographic data used by CaMa-Flood. Additionally, nationwide station datum heights reported in the observed data are not standardised and vary significantly. These inconsistencies in vertical reference frames introduce significant bias when comparing absolute water level values. Given the above, our water level validation approach evaluated the temporal consistency between model outputs and observations. We assessed how well the model captured seasonal variability and interannual trends using the Pearson correlation coefficient (R). The analysis shows that most stations achieved moderate to strong agreement, with most R values falling between 0.45 and 0.75. This indicates that the model reliably reproduces the timing and relative magnitude of water level fluctuations, which is critical for flood risk applications.

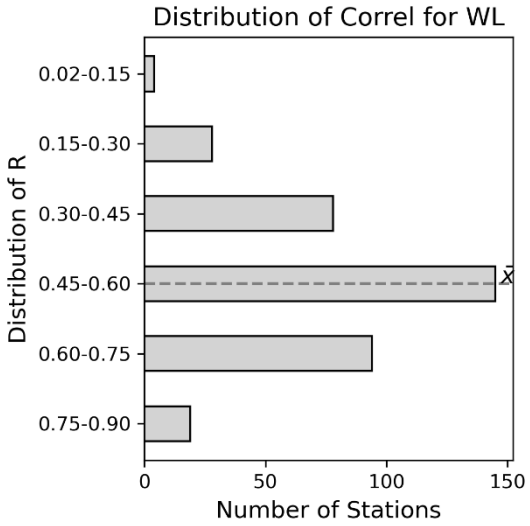


Figure S9: Distribution of Pearson R for Obs and Sim daily Water levels

S3. Additional Supporting Figures

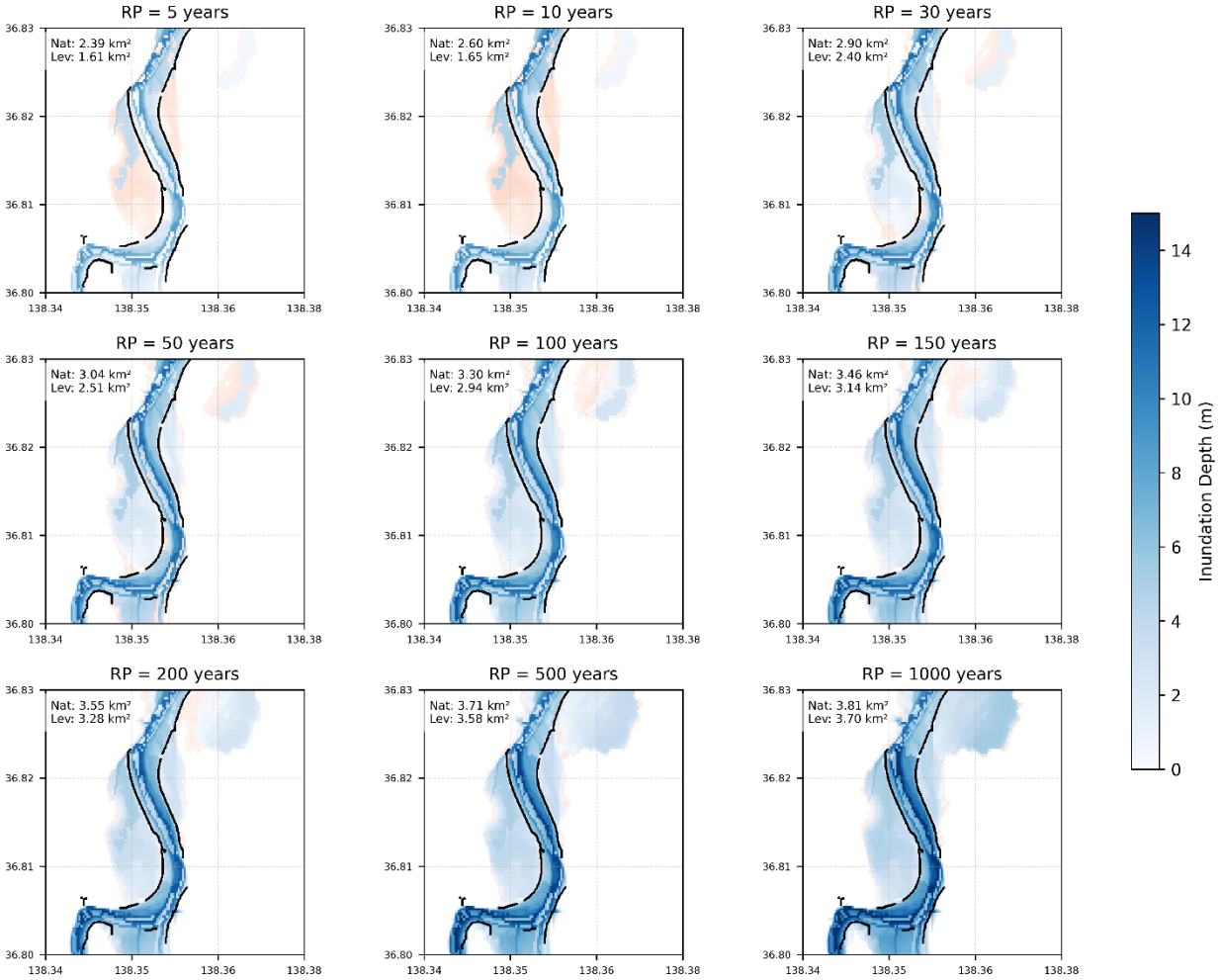


Figure S10: 22 km downstream of Chaikuma-Sai Confluence. Blue indicates inundations with Levee, Red backward layers are inundations in Natural. The black line represents levee data, and estimated inundation areas are mentioned in text boxes at top left

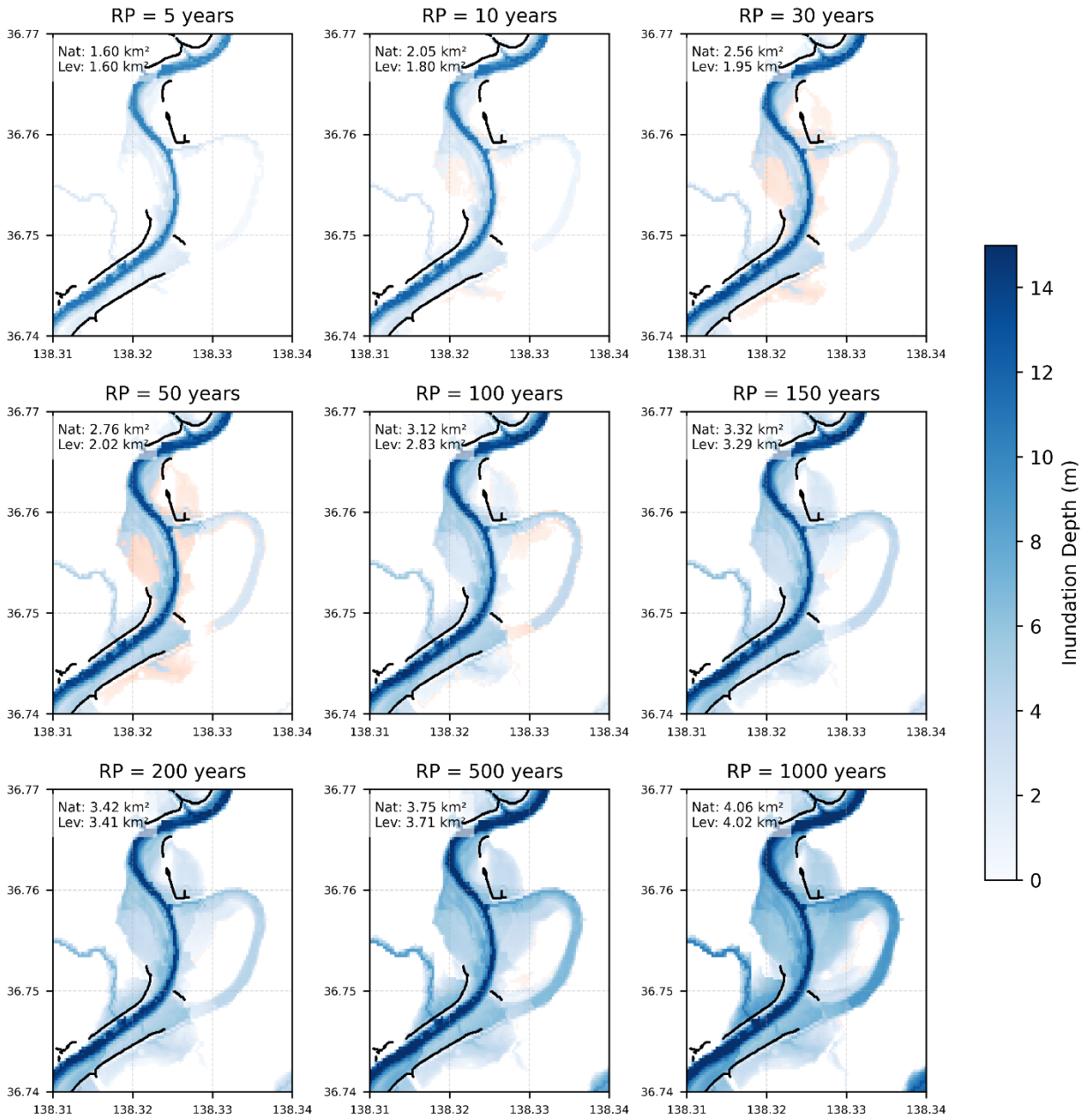


Figure S11: 14 km downstream of Chaikuma-Sai Confluence. Blue indicates inundations with Levee, Red backward layers are inundations in Natural. The black line represents levee data, and estimated inundation areas are mentioned in text boxes at top left.

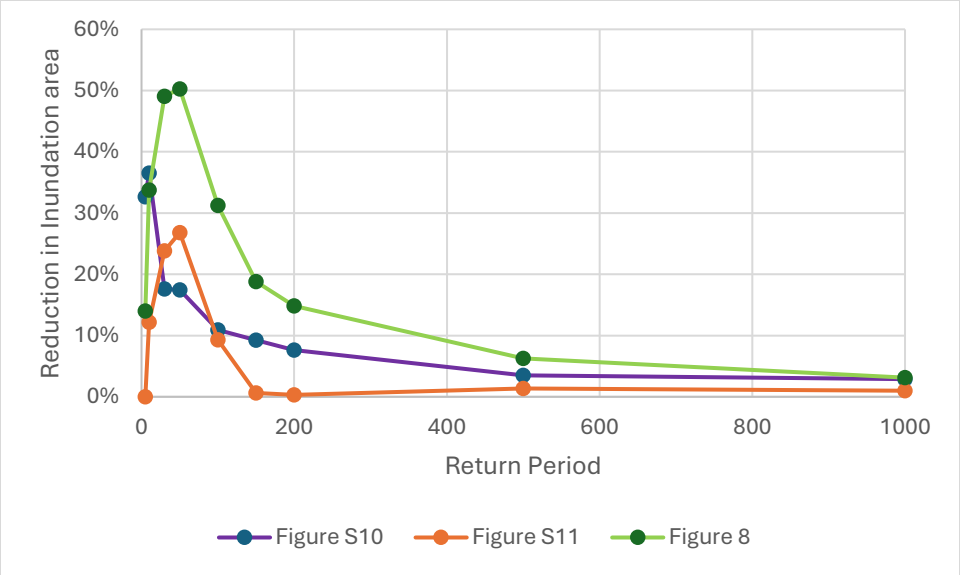


Figure S12: Percentage reduction in Inundation area referenced from Figures S10, and S11 and Main text Figure 8

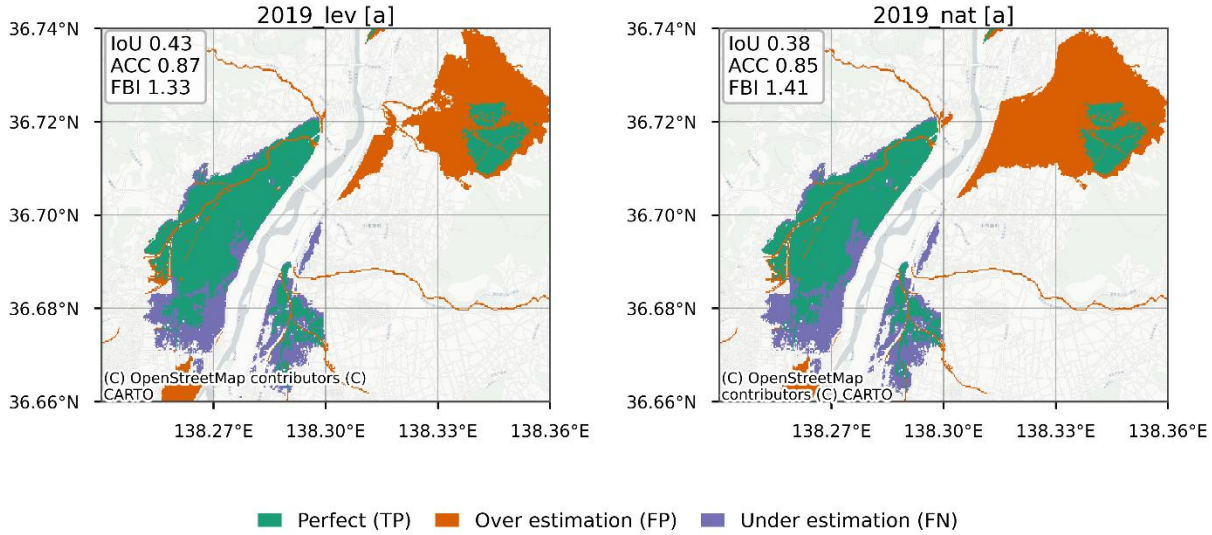


Figure S13: Estimated Inundation Extents in the Chikuma River reaches. © OpenStreetMap contributors (https://www.openstreetmap.org/copyright) © CARTO

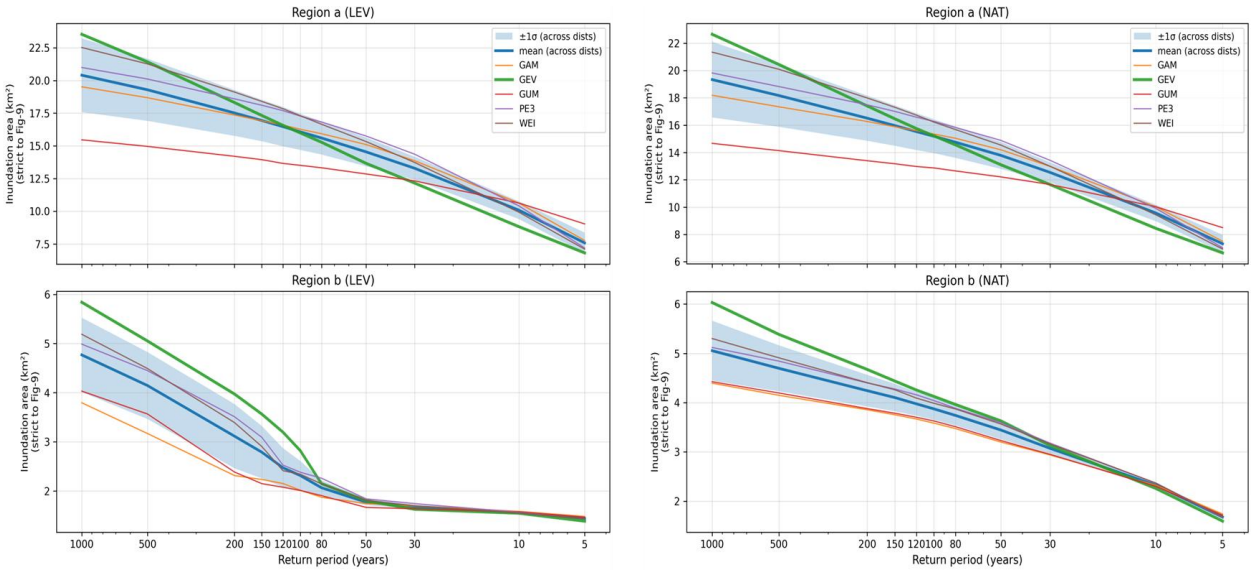


Figure S14: Sensitivity of inundation area to the choice of extreme-value distribution.
Mean inundation area (solid blue line) and associated uncertainty (±1 standard deviation, shaded) across fitted frequency distributions are shown as a function of return period for two representative sub-regions (a and b), under levee (LEV) and natural (NAT) conditions. Coloured lines indicate results obtained using individual distributions (GEV, GAM, GUM, PE3, and WEI). Variability in inundation area increases toward higher return periods, reflecting sensitivity to the upper tails of the fitted distributions.