8. Supplementary Information

S1. Overview of extreme events

Figure 6 provides an overview of the investigated flood events in the ten basins. Figure 7 provides an overview of the 2003 drought event, by showing the periods where the observed discharge drops below the defined threshold level for each of the ten basins.

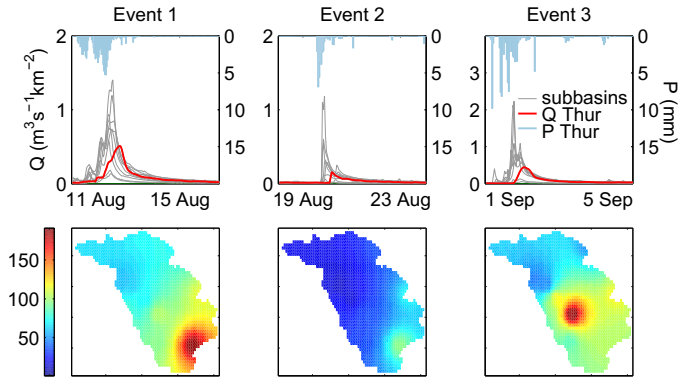

Figure 6: The three flood events, shown with their relative discharge for the Thur basins and the nine sub-basins (upper panels), and the distribution of precipitation (in mm) during these events (lower panels).

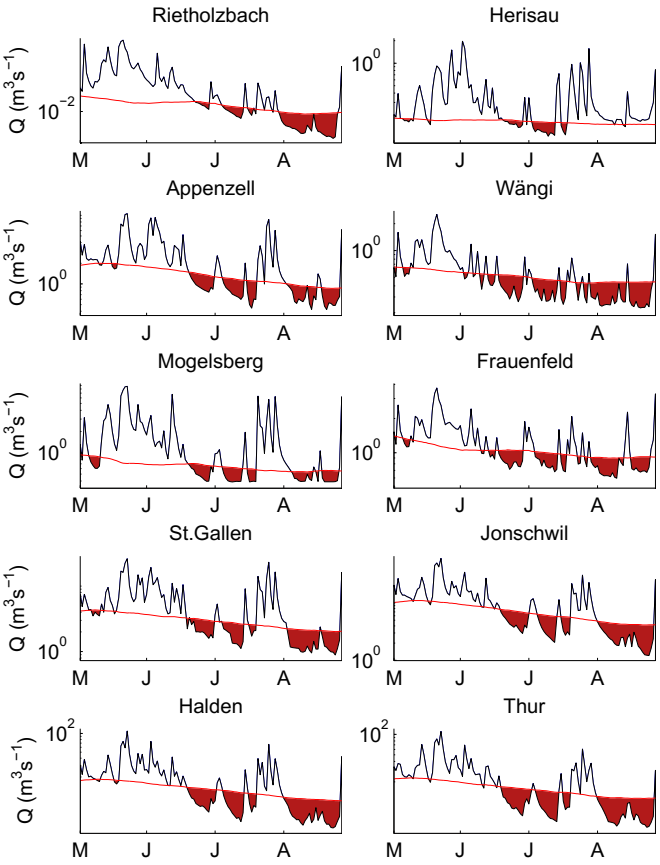

Figure 7: The daily observed discharge (on a logarithmic axis) for the summer months in 2003 in each of the ten basins, and the drought threshold level Q_{90} in light red. The basins is assumed to experience a hydrological drought as soon as the discharge drops below the threshold level. The areas in dark red are defined as the drought deficit; the difference between the observed discharge and the threshold level.

S2. Model performance expressed in performance metric

In total, twelve different model configurations were employed, with varying resolution, spatial representation of forcing, and calibration period. The model performance for the different configurations is expressed in terms of NSE(Q) with an hourly time interval for the flood events, and the NSE(logQ) with a daily time interval for the drought event. An overview of the model performance is provided in Figure 8.

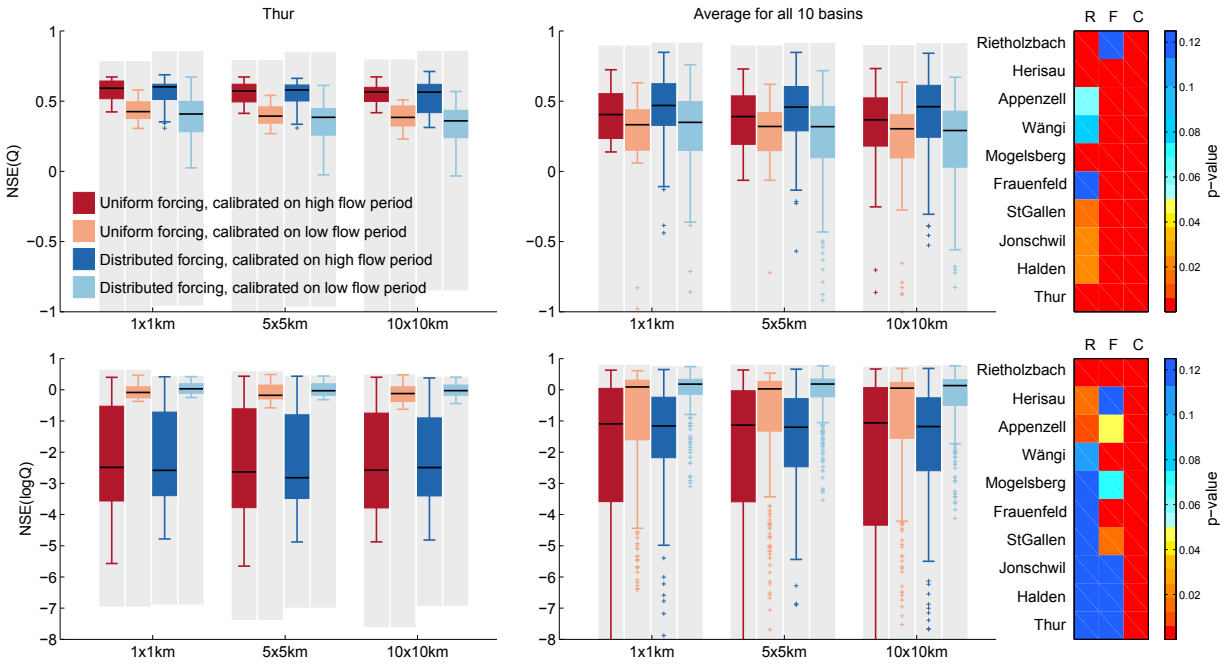

Figure 8: Model performance expressed in NSE(Q) at an hourly time step (upper panels) and NSE(logQ) at a daily time step (lower panels) for the validation of the flood period and the drought period respectively (see Figure 2), and ANOVA significance level to show the impact of modeling decisions on model performance. Left: The model performance of the selected behavioral sets (1%) in the validation period for the different model configurations. The left panel refers to the Thur basin only, the middle panel shows the model performance averaged for all ten basins. The colored boxes show the 25–75% quantile. The grey boxes show the model performance of the complete parameter sample in the validation period. Right panel: ANOVA significance level of the impact of Resolution (R), Forcing (F), and Calibration period (C) on the model performance for the 10 considered basins. The basins are ordered from small to large.