https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

Why our rainfall-runoff models keep underestimating the peak flows?

András Bárdossy[1] and Faizan Anwar[1]
[1]Institute for Water and Environmental System Modeling, University of Stuttgart, 70569 Stuttgart, Germany
Correspondence: Faizan Anwar (faizan.anwar@iws.uni-stuttgart.de)

Abstract.

In this paper the question of how interpolation of precipitation in space by using various spatial gauge densities affects the rainfall-runoff model discharge if all other input variables are kept constant is investigated. This was done using a physically-based model as the reference with a reconstructed spatially variable precipitation and a conceptual model calibrated to match the reference model output. Both models were run with distributed and lumped inputs. Results showed that all considered interpolation methods resulted in underestimation of the total precipitation volume and that the underestimation was directly proportional to the amount. The underestimation was very severe for low observation densities and disappeared only for very high density precipitation observation networks. This result was confirmed by using observed precipitation with different observation densities. Model runoffs showed worse performance for their highest discharges. Using lumped inputs for the models showed deteriorating performance for peak flows as well even when using simulated precipitation.

1 Introduction

Hydrology is to a very large extent driven by precipitation, which is highly variable in space and time. Point precipitation is interpolated in space that is subsequently used as the true precipitation input for hydrological models without any further adjustments. A common problem with most interpolation schemes is that the variance of an interpolated field is always lower than the variance of the individual point data that were used to interpolate it. Another problem which is very important in the case of high precipitation events is that the absolute maximum precipitation coordinates in the interpolated field are at one of the observation locations. Contrary to reality, it is very unlikely that our observation networks persistently catch the maximum precipitation. Accounting for conditional changes with additional variables helps only a little but never enough. New interpolations schemes are introduced on a regular basis but the main drawback is inherent to all of them. The reduction comes mainly from the underestimation of precipitation for the high values and also because the distribution of the interpolated values tends to Gaussian the more neighbors one takes for interpolation for each grid cell, even when it is known that precipitation

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

has an exponential-like distribution in space for a given time step and for a point in time. The problem gets worse as the number of stations reduces. From experience in previous studies, Bárdossy et al. (2022, 2020); Bárdossy and Das (2008), it was clear that model performance was dependent on the number and the configuration of observation station locations but no decisive conclusion was made. It should not come as a surprise that models perform better, in terms of cause and effect, with increasing data quantity and quality. To investigate the effects of sampling finite points in space on model performance and the reproduction of flood peaks this study tries to answer the following questions:

1. How strong is the effect of sampling density the estimation of areal precipitation for intense precipitation events? Is there a systematic signal?
2. How does interpolation quality and hydrological model performance change with precipitation observation density?
3. How strong is the effect of an error in precipitation observations on discharge estimations?
4. What is the role of spatial variability of hydrological model performance? How much information is lost if areal mean precipitation is used i.e. sub-scale variability is ignored?

Some of the aforementioned points, among others, were recently discussed in a review paper by Moges et al. (2021). They highlighted four sources of hydrological model uncertainties: parameter, input, structural and calibration data uncertainty. Their conclusion was that out of the four, only parameter uncertainty got the most attention. Even though, all of these sources contribute to the final results in their unique ways.

Kavetski et al. (2003) concluded that addressing all types of uncertainty will force fundamental shifts in model calibration/verification philosophy. So far, the rainfall-runoff modeling community has not been able to put such ideas into common practice. The reason being that we can never find a true estimate of the uncertainty of the forever changing system that we are trying to model. The so called epistemic uncertainty will always exist. Moreover, there is no consensus on how to model these uncertainties. Following Kavetski et al. (2003), Renard et al. (2010) showed that taking into account both input and structural uncertainty is an ill-posed problem as combinations of both affect the output and the performance of the model and addressing both simultaneously is not possible.

Balin et al. (2010) tried to assess the impacts of having point rainfall uncertainty on model discharge by taking a 100 km^2 catchment at a spatial resolution of 200 m at a daily time scale. A normally distributed error of 10% was added to the observed time series to produce new time series. This was done multiple times independently. By running the model with the erroneous data (among other things), they found that the resulting performances were not so different than the original case. The only noticeable difference was that the uncertainty bounds on discharge were slightly wider resulting in more observed values contained by it. The conclusion was that the causes of model output uncertainty may not be due to erroneous data as the measurement errors everywhere compensated for the runoff error in a way that the final model performance stayed the same. More interestingly, they found out that using observed rainfall for modeling resulted in under estimation of the peak flows, a problem that this study deals with. Furthermore, for the same catchment Lee et al. (2010) arrived at similar conclusions by using a different approach and a lumped rainfall-runoff model. Yatheendradas et al. (2008) investigated uncertainty of flash floods

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

using a physically-based distributed model by considering a mixture of parameter and input data uncertainty. They concluded that their model responses were heavily dependent on the properties of the precipitation estimates using radar, among other findings.

This study is a special case of model input uncertainty. It specifically deals with the effects of using interpolated precipitation data as a rainfall-runoff model input that comes from point observations. It does not deal with input uncertainty as it is meant to be. To do so very strong assumptions have to be made. Assumptions which are likely to remain unfulfilled, which was also pointed out in Beven (2021). The idea that various forms of uncertainties exist and should be dealt with is not disputed. This research is not intended as a review, hence only relevant works are cited and discussed.

The rest of the study is organized as follows: Sect. 2 shows the investigation area and Sect. 3 shows the model setup and inputs and especially the main idea of this study in detail. Sect. 4 discusses the two models used in this study and the methods of calibration. Sect. 5 discusses the results in detail where the questions posed in the beginning of this section are answered. Finally, in Sect. 6, the study ends with the summary, main findings and implications of the results.

2 Investigation area

The study area is the Neckar catchment situated in the South-West of Germany in the federal state of Baden-Württemberg (Fig. 1). It has a total area of 14,000 km^2. From the East it is bounded by the Swabian Alps and by the Black forest on the West. The maximum elevation is ca. 1000 m in the Swabian Alps that goes down to 170 meters at the confluence of the Neckar and Rhine in the North. Mean recorded temperature is about 9.1 °C while minimum and maximum ever recorded are -28.5 and 40.2 °C respectively. Annual precipitation sums reach to about 1800 mm in the Black Forest while the rest of the area receives somewhere between 700 to 1000 mm on the mean. The chances of precipitation per day are about 50% with a mean precipitation of 2.5 mm per day over the catchment.

The entire area of the Neckar was not modeled by the considered rainfall-runoff models in this study. Rather, the large headwater catchments were modeled where times of concentration were long enough to allow for model runs on a daily resolution. Namely, the Enz, Kocher and Jagst with catchment areas of 1656, 1943 and 1820 km^2 respectively. Another reason for the selection of these was that they are relatively intact as compared to the main river which is modified for transportation.

3 Model input data preparation

Point meteorological data time series (daily precipitation, and temperature) were downloaded from the Deutscher Wetter Dienst's (DWD, German Weather Service) open access portal (DWD, 2019). The daily discharge time series was downloaded from the open access portal of the Landesanstalt für Umwelt Baden-Württemberg (LUBW, Environmental agency of the federal state of Baden-Württemberg) (LUBW, 2020). The considered time period is from 1991 to 2015.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

3.1 Precipitation interpolation using various gauge densities

The effects of taking a sample from a population on the final distribution of precipitation can be visualized even before the start of the experiment. Imagine one was to observe precipitation at each point in space. For any given time step, when it is raining at some of the points, the distribution of values is exponential-like. If one were to sample from this a finite number of N values, the chances that points are sampled from the upper tail become smaller and smaller as N approaches 1, at which point the probability of sampling from the upper tail corresponds directly to the probability density function. This also implies that samples from the lower tail are more likely. The consequence at the end is that low values are overestimated while high values are underestimated on the mean when interpolating the whole grid.

To further prove this point, the time series of the existing precipitation network for the time period of 1991-2015 was taken. There are a total of 343 gauges. Only a subset of these is active at any given time step as old stations are decommissioned and new ones are commissioned. Out of the total, random samples of sizes 10, 25, 50, 100 and 150 gauge time series were selected and gridded precipitation was interpolated using these for each catchment that was subsequently lumped into a single value for each time step. This was done 100 times. For comparison, the same was done by using all the gauges. Furthermore, two interpolation schemes i.e. Nearest neighbor (NN) and Ordinary Kriging (OK) were used to show that the problem was not interpolation scheme dependant. A stable variogram fitting method that was described in Bárdossy et al. (2021) was used for OK.

3.2 Reference precipitation

In the previous case preciitation interpolations were not compared to reality, as one does not know what the real precipitation was at unmeasured locations. To obtain a complete coverage of rainfall simulated precipitation fields have to be considered.

A realistic virtual dataset was created to investigate the effect of precipitation observation density instead of using interpolated precipitation. For this purpose, a 25 year long daily precipitation dataset corresponding to the time period 1991-2015 was used. This dataset contains gridded precipitation on a 147x130 km grid with 1 km resolution. It was created so that the precipitation amounts were the same as the observed precipitation at the locations of the weather service observation stations. Additionally, their empirical distribution function of the entire field for any selected day is the same as that of the observations and its spatial variability (measured as the variogram) is the same as the observed as well. This precipitation is considered as *reality*. It is called *reference or reconstructed precipitation* throughout this text. Full details of the procedure are described in Bárdossy et al. (2022).

3.3 Precipitation interpolation using the reference

To demonstrate the effects of sparse sampling of data and the resulting model runoff error, the following method was used:

N number of points were sampled from the reference grid. Time series for each point (1x1 km cell) was then extracted and taken as if it were an observed time series. Care was taken to sample points such that the density was nearly uniform over the study area. N was varied to obtain a given amount of station density. Here, densities of 1 in 750, 400, 200 and 130 km^2

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

were used. These correspond to 25, 50, 100 and 150 cells out of the 19,110 respectively. Labels of the form MN, are used to refer to these in the figures, where M is only a suffix to signify interpolation with N being the number of points used to create the interpolation. For reference, Germany has around 2000 active daily precipitation stations for an area of 360,000 km^2 which is about 1 station per 180 km^2. The sampling was performed 10 times for each N in order to see the effects of different configurations later on in the analysis of the results.

This way many time series were sampled from the reference for various gauge densities. From here on the same procedure was applied that is normally used in practice i.e. Spatial interpolation. To keep things simple the Ordinary Kriging (OK) method was used to interpolate fields on the same spatial resolution as the reference at each time step. The use of OK is arbitrary. One could very well use any preferred method. Use of other methods that interpolate in space will not help much as all of them tend to result in fields that have reduced variance as compared to the variance of the observed values. Moreover, it is also very unlikely that an interpolation scheme predicts an extreme at locations where no measurements were made.

3.4 Temperature and potential evapotranspiration

External Drift Kriging (EDK) (Ahmed and De Marsily, 1987) was used to interpolate daily observed minimum, mean and maximum temperature for all cells at a resolution similar to that of the precipitation with elevation from the SRTM (Farr et al., 2007) as the drift. For the potential evapotranspiration, the Hargreaves-Samani (Hargreaves and Samani, 1982) equation was used with the interpolated temperature data at each cell as input. It was assumed that temperature and potential evapotranspiration are much more continuous in space as compared to precipitation which is more like a semi-Markov process in space-time that has a much larger effect on the hydrograph in short term.

4 Model setup

Two rainfall-runoff models, namely SHETRAN (Ewen et al., 2000) and HBV (Bergström, 1992), were considered in this study. Same gridded inputs were used for both and at a spatial resolution of 1x1 km and at a daily temporal resolution. Except for precipitation, all other inputs stayed the same. Setup specific to each model is discussed in the following two subsections.

4.1 SHETRAN

SHETRAN is a physically based distributed hydrological model which simulates the major flows (including subsurface) and their interactions on a fine spatial grid (Ewen et al., 2000). It includes components for vegetation interception and transpiration, overland flow, variably saturated subsurface flow and channel–aquifer interactions. The corresponding partial differential equations are solved using a finite-difference approximation. Being a physically-based model, SHETRAN requires much more input as compared to lumped models such as HBV. The model parameters were not calibrated. Instead, available data such as the digital elevation model, soil and land use maps were used to estimate the model parameters at a 1x1 km spatial resolution (Lewis et al., 2018; Birkinshaw et al., 2010). It was considered as a theoretically correct transformation of rainfall to runoff. This way, combined with the reference precipitation, a realistic virtual reality was created in which the effect of differ-

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

ent sampling densities could be investigated. Same SHETRAN settings were used for the various precipitation interpolations. Furthermore, the model and settings are the same as those in Bárdossy et al. (2022).

4.2 HBV

HBV is one of the most widely used models that needs no introduction. It requires very little input data i.e. precipitation, temperature and potential evapotranspiration. Each grid cell of HBV was assumed to be a completely independent unit. All cells shared the same parameters, only the inputs were different. The runoff produced by all cells was summed up at the end to produce the final simulated discharge value for each time step. It was calibrated for the reference precipitation and each of the precipitation interpolation. DE (Storn and Price, 1997) was used to find the best parameter vector. It is one of the genetic-type optimization schemes to find the global optimum by updating a given sample of parameter vectors successively by mixing with each other and perturbing them individually. A population of 400 was used to find the global optimum. Overall, it needed 150 to 200 iterations to converge for 11 parameters. 50% Nash-Sutcliffe (NS) (Nash and Sutcliffe, 1970) and 50% NS using the Natural logarithm (Ln-NS) was used as the objective function for calibration. Ln-NS was chosen because NS alone concentrates too much on the peak flows during calibration and disregards, almost 95% of, the remaining flows. Ln-NS helped to mitigate this flaw to some degree but not completely.

5 Results

Before showing the results, some terms specific to the upcoming content are defined first. They are put here for the readers' ease.

1. The term *reference precipitation* refers to the reconstructed precipitation that is taken as if it were the observation. *Reference model* refers to SHETRAN with the reference precipitation as input and the resulting discharge of this setup is the *reference discharge*.

2. *Interpolated discharge* refers to the discharge of SHETRAN or HBV with interpolated precipitation as input. The model is mentioned specifically for it. The term *subsampling* refers to extraction of a subset of points (and their time series) from the entire grid.

3. *Model performance*, refers to the value of the objective function whose maximum, and optimum, value is 1.0, anything less is *less* performance. The performance of the reference discharge is 1.

4. Furthermore, *discharge* and *runoff* are used interchangeably. They both refer to the volume of water coming out of the catchment per unit time which is cubic meters per second in this study. The terms *observation station*, *gauge* and *station* are used interchangeably. These refer to the meteorological observation stations.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

5.1 Metrics used for evaluation

To compare the change in precipitation or model runoff, scatter plots of reference values on the horizontal versus their corresponding values after interpolation on the vertical scale are shown. Each point represents lumped precipitation for all the cells i.e. mean of all the cells per time step.

The largest five aerially lumped values for precipitation in the reference and the values at the corresponding time steps in interpolation are compared by showing them as a percentage of the reference. This produces a number of points that is the product of the number of interpolations and number of events used. For comparing discharge, 5 largest values in the reference discharge are compared against the values at the same time steps using interpolated discharge. This results in 5 points per interpolation. Violin plots are used to show these points as densities.

Furthermore, figures comparing a high discharge event using all the interpolations are shown at the end of each section wherever relevant.

Tables summarizing over- and under-estimations as percentages relative to the reference are shown at the end of each subsection.

5.2 Comparison of interpolations using fewer vs. all gauges

Comparison of the largest five precipitation events' depth using various number of gauges taken from the entire network are shown in Fig. 2 and 3 for Enz. These are normalized with respect to the values computed using the entire network (343 gauges). There are many cases where the fewer gauges interpolations have larger values than the ones with the entire network. The more important point to observe is the bias. By using a lower number of gauges, underestimation of the largest precipitation events is more likely. NN is more spread out in terms of under- and over-estimations compared to OK. Using more gauges shows that the deviations reduce significantly. Another aspect that should be kept in mind is that here *interpolations* are compared to *an interpolation*. Even by using all the gauges, there is still a good chance of missing the absolute maximum precipitation at a given time step. Keeping this in mind one should be aware that runoff predicted by a model using this *smoothed* precipitation with all the gauges will still produce, on average, smaller peaks. This will become clearer by the results in the next sections where the reference precipitation is used. Tables 1 and 2 summarize in numbers the cases with under- and over-estimations using various number of gauges with respect to using all of them for interpolation for the three catchments using NN and OK respectively.

5.3 Effects of subsampling from reference on precipitation

Fig. 4 shows an exemplary event with very high daily precipitation for the reference and various interpolation cases. For the lowest number of gauges the field looks very smooth and has a smaller variance as compared to the field with the most stations which is much closer to the reference.

In Fig. 5, the scatter plot of lumped reference against one of the lumped interpolation values for all time steps for Kocher is shown. Here, it is interesting to see that overall, the larger the value in reference, the more it is reduced by the interpolation.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

	10	25	50	100	150
Enz	60-09-31	51-16-33	45-26-29	35-41-24	25-52-23
Kocher	55-13-32	49-20-31	40-28-32	32-38-30	22-53-25
Jagst	52-15-33	48-20-32	40-28-32	32-36-32	25-53-22

Table 1. Relative percentages of under- and over-estimations of the top 5 precipitation events using various gauge densities (columns) with respect to the top 5 values using interpolation with all the gauges for the three considered catchments (rows) using NN. The values are of the format *percentage of underestimations (below 97.5%), within a threshold of ±2.5%, and overestimations (above 102.5%) with respect to the interpolation using all gauges*. For example, 60-09-31 in the first row and first column means that out of the 100 interpolations (with 5 events per interpolation) using 10 gauges for Enz, 60% of the events were below 97.5%, 9% were within 97.5% and 102.5%, and 31% were above 102.5% of the top 5 events using the interpolation with all the 343 gauges.

	10	25	50	100	150
Enz	66-09-25	56-18-26	50-29-21	36-43-21	24-56-20
Kocher	61-15-24	62-17-21	52-25-23	35-42-23	20-57-22
Jagst	60-12-28	57-20-23	50-29-21	39-40-22	24-60-15

Table 2. Under- and over-estimation percentages of the top 5 precipitation events using OK. Caption of Table 1 shows how to interpret the numbers here.

The other way around, the interpolation increases the magnitude of the small values. Events in the mid range values are underestimated by a significant margin. Most points are below the ideal line. Now imagine the subsequent mass balance problems that would arise by such a consistent bias. Over the long term, one would adjust the model to have lower evapotranspiration. Over the short term, the peak flows would almost always be underestimated. It is important to keep in mind that high discharge values are the result of a threshold process in the catchment where the water moves in larger volumes towards the stream once the soil saturates or when the infiltration cannot keep up with the rainfall/melt intensity. To match the peak in such scenarios, it is important to get the correct estimates of precipitation.

Fig. 6 shows the relative change of the largest five peak precipitation values. These were computed by dividing the interpolated precipitation by the reference at the time steps of the top 5 events. A consistent bias, i.e. underestimation, is clear. Especially, for the coarsest interpolation (25 points). Such a figure seems small but imagine the extra volume over a 1000 km^2 catchment that is not intercepted by the soil. Another interesting thing to note is that for the other interpolations there are some overestimations as well. All the relative under- and over-estimations for the three catchments with various densities are summarized in Table 3.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

	M025	M050	M100	M150
Enz	74-20-06	60-20-20	45-38-18	44-47-09
Kocher	80-06-14	60-18-22	50-15-35	31-44-24
Jagst	84-04-12	42-24-34	35-35-30	22-53-24

Table 3. Percentages of under- and over-estimations of the top 5 precipitation events using various gauge densities (columns) with respect to the top 5 values using reference precipitation for the three considered catchments (rows). The values are of the format *percentage of underestimations (below 97.5%), within a threshold of ±2.5%, and overestimations (above 102.5%) with respect to the reference precipitation*. For example, 31-44-24 in the second row and last column means that out of the 10 interpolations (with 5 events per interpolation) using 150 points for Kocher, 31% of the events were below 97.5%, 44% were within 97.5% and 102.5%, and 24% were above 102.5% of the top 5 events using the reference precipitation.

5.4 Effects of subsampling from reference on discharge of SHETRAN

Similar to the precipitation, the systematic bias in model runoff was investigated as well. Fig. 7 shows the resulting runoff by using the same precipitation (Fig. 5) as input to SHETRAN. What is immediately clear is that there are almost no overestimations of discharge values when using interpolated precipitation. The largest peak is reduced by almost 50%.

Looking at Fig. 8, the mean of the largest five peaks is reduced significantly while using the least number of points for Kocher. The other thing to note is that the peaks drop on average for other interpolations (except for the last one) much more as compared to the reduction in precipitation. To see the effects more in detail, Fig. 9 and 10 show hydrographs obtained using various gauging densities for two events. It is very clear that as the gauging density rises, the underestimation decreases proportionally and the hydrographs become similar. All the under- and over-estimations are summarized in the Table 4 for all the catchments.

	M025	M050	M100	M150
Enz	80-02-18	78-06-16	60-28-12	49-27-24
Kocher	91-04-04	84-10-06	56-24-20	27-49-24
Jagst	76-04-20	62-18-20	56-28-16	38-38-24

Table 4. Percentages of under- and over-estimations of the top 5 discharge events using various gauge densities (columns) with respect to the top 5 values using reference discharge for the three considered catchments (rows) using SHETRAN. The values are of the format *percentage of underestimations (below 97.5%), within a threshold of ±2.5%, and overestimations (above 102.5%) with respect to the reference discharge*. For example, 38-38-24 in the third row and last column means that out of the 10 interpolations (with 5 events per interpolation) using 150 points for Jagst, 38% of the events were below 38%, 38% were within 97.5% and 102.5%, and 24% were above 102.5% of the top 5 events using the reference discharge.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

5.5 Effects of subsampling from reference on discharge of HBV

Looking at the scatter of reference and interpolated precipitation discharge in Fig. 11, HBV shows a different behaviour as compared to SHETRAN. First of all, overestimations from low to high flows exist except for the largest high flows which are underestimated as well. Again, not as much as that by SHETRAN. This is due to the recalibration, where the new parameters compensate for the missing precipitation by decreasing evapotranspiration. This aspect will be investigated thoroughly in future research.

Fig. 12 shows the scaling of the five highest peaks compared to the reference discharge. Here, a similar reduction for the least amount of stations can be seen. Even for the highest number of stations, the discharges are still underestimated. This signifies that even a distributed HBV with full freedom to readjust its parameters cannot fully mimic the dynamics of the flow produced by SHETRAN. Hydrographs for the same events shown in the previous section for HBV are shown in Fig. 13 and 14. It is interesting to note that the first event is overestimated by all interpolations and that the hydrographs become similar as the gauging density increases. The second event is estimated better as the gauging density increases. All the under- and over-estimations are summarized in the Table 5 for all the catchments.

	M025	M050	M100	M150
Enz	78-04-18	78-04-18	78-02-20	71-00-29
Kocher	100-00-00	96-04-00	90-08-02	89-11-00
Jagst	74-06-20	78-00-22	80-00-20	80-00-20

Table 5. Percentages of under- and over-estimations of the top 5 discharge events using various gauge densities (columns) with respect to the top 5 values using reference discharge for the three considered catchments (rows) using HBV. Caption of Table 4 shows how to interpret the numbers here.

5.6 Effects of removing subscale variability of precipitation on SHETRAN discharge

Effects of using lumped precipitation, by taking its mean, on the resulting discharge i.e. same precipitation value at each time step for all cells were also investigated. The aim was to see the effects of subscale variability on runoff. While considering 10 largest peaks per catchment for the entire time period, NS of these dropped to 0.77, 0.78 and 0.90 for Enz, Kocher and Jagst respectively. Almost all peaks were reduced in their magnitudes to 84%, 85% and 93% with respect to the ones produced by the model on average with the distributed reference precipitation. Most of the underestimation of the peaks were during winter that are likely to be snowmelt events. These are location/elevation dependent and it makes sense that using a lumped value of precipitation results in incorrect melt behavior. Overall, the tendency was towards reduced discharge when using lumped precipitation. This tendency is likely to be much higher when a single cell is used to represent the catchment i.e. a fully lumped model.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

5.7 Effects of measurement error in precipitation on runoff

To test how measurement error affects the model discharge, precipitation with a measurement error of 10% of each observed value having a standard Normal distribution was used and then interpolated as well. There, it was observed that magnitude of under- and over-estimation of the peaks becomes more variable as compared to the reference but the bias stayed the same as that compared to using precipitation with no error. Table 6 shows the effects for precipitation. Comparing these to Table 3, we can observe that the trends are not so different. Similar to interpolations with no errors, the ones with the largest number of samples have more values closer to the reference precipitation. These results corroborate the conclusions by Balin et al. (2010); Lee et al. (2010).

These results have an interesting consequence. Given that gauges have measurement errors that are Normally distributed i.e. cheaper gauges, such as the Netatmo personal weather stations, can be used to close the gap of missing precipitation due to sparse distribution networks. Studies, such as those by de Vos et al. (2017, 2019); Bárdossy et al. (2021), have shown that these alternative sources of data can augment the existing networks while still having some drawbacks. The type of measurement error by these can be further studied to validate their actual usefulness for rainfall-runoff modeling.

	M025	M050	M100	M150
Enz	71-11-18	46-32-22	54-22-24	24-50-26
Kocher	80-09-11	66-18-16	46-34-20	38-34-28
Jagst	82-11-07	60-16-24	48-30-22	30-40-30

Table 6. Percentages of under- and over-estimations of the top 5 precipitation events using various gauge densities (columns) with respect to the top 5 values using reference precipitation for the three considered catchments (rows). The values are of the format *percentage of underestimations (below 97.5%), within a threshold of ±2.5%, and overestimations (above 102.5%) with respect to the reference precipitation*. For example, 38-34-28 in the second row and last column means that out of the 10 interpolations (with 5 events per interpolation) using 150 points for Kocher, 38% of the events were below 97.5%, 34% were within 97.5% and 102.5%, and 28% were above 102.5% of the top 5 events using the reference precipitation.

6 Summary and conclusions

An often ignored problem with model discharge in rainfall-runoff modeling by using interpolated data was investigated in this study. To do so, data were interpolated using different gauge densities. It was shown how interpolated precipitation differs from reference precipitation. SHETRAN was used as a reference model that was assumed to represent reality with reconstructed precipitation as input. The other model was HBV. Runoff from SHETRAN was chosen as a reference to avoid inherent mismatch of mass balances as compared to using observed discharge series.

The results showed that:

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022


Hydrology and Earth System Sciences
Discussions

1. Interpolation produced a consistent underestimation of high precipitation values and overestimation of the low ones. The underestimation increased with the increase in magnitude of precipitation. Possible reasons were also discussed. This bias resulted in much higher error in model discharge i.e. consistent underestimation of peak flows.

2. Removing subscale variability lead to underestimation of peaks. This is most likely due to the redistribution of water to cells with less water from the cells with more water. As the drier cells hold the additional water, which would have otherwise overflowed, from the cells that were saturated. Incorrect snowmelt due to spatial averaging is also a culprit.

Further conclusions that can be derived from the above mentioned results are:

1. While modeling in hydrology, variables should be modeled in space and time at the correct resolution to have usable results. Disregarding spatial characteristics leads to problems that cannot be solved by any model or finer resolution temporal data.

2. Cheaper networks may prove valuable where observations are sparse given that their measurement error are Normally distributed.

In future research, the following issues could be addressed:

1. Underestimation of intense precipitation due to interpolation.

2. Sensitivity of other variables, such as temperature, to interpolation and their effects on runoff. Especially, catchments with seasonal or permanent snow cover.

3. The magnitude of performance compensation that recalibration introduces due to the missing precipitation.

4. The effects of using different density networks on calibrated model parameters and regionalization of model parameters.

5. Determining the error distributions of cheaper precipitation gauges to establish their usefulness in rainfall-runoff modeling.

Author contributions. BA: conceptualization. BA, FA: Data preparation. BA: SHETRAN modeling. FA: HBV modeling. BA and FA: Final manuscript preparation.

Competing interests. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements. Python (van Rossum, 1995), Numpy (Harris et al., 2020), Scipy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), Pandas (The pandas development team, 2020; Wes McKinney, 2010), Pathos (McKerns et al., 2012), PyCharm IDE, Eclipse IDE and PyDev.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

The authors acknowledge the financial support of the research group FOR 2416 “Space-Time Dynamics of Extreme Floods (SPATE)” by the German Research Foundation (DFG) and the University of Stuttgart for paying the article processing charges of this publication.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Open Access
Discussions

References

Shakeel Ahmed and Ghislain De Marsily. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resources Research*, 23(9):1717–1737, 1987. https://doi.org/https://doi.org/10.1029/WR023i009p01717.

Daniela Balin, Hyosang Lee, and Michael Rode. Is point uncertain rainfall likely to have a great impact on distributed complex hydrological modeling? *Water Resources Research*, 46(11), 2010. https://doi.org/https://doi.org/10.1029/2009WR007848. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009WR007848.

A. Bárdossy and T. Das. Influence of rainfall observation network on model calibration and application. *Hydrology and Earth System Sciences*, 12(1):77–89, 2008. https://doi.org/10.5194/hess-12-77-2008. URL https://hess.copernicus.org/articles/12/77/2008/.

A. Bárdossy, J. Seidel, and A. El Hachem. The use of personal weather station observations to improve precipitation estimation and interpolation. *Hydrology and Earth System Sciences*, 25(2):583–601, 2021. https://doi.org/10.5194/hess-25-583-2021. URL https://hess.copernicus.org/articles/25/583/2021/.

S. Bergström. *The HBV Model: Its Structure and Applications*. SMHI Reports Hydrology. SMHI, 1992. URL https://books.google.de/books?id=u7F7mwEACAAJ.

Keith Beven. An epistemically uncertain walk through the rather fuzzy subject of observation and model uncertainties1. *Hydrological Processes*, 35(1):e14012, 2021. https://doi.org/https://doi.org/10.1002/hyp.14012. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.14012.

Stephen J Birkinshaw, Philip James, and John Ewen. Graphical user interface for rapid set-up of shetran physically-based river catchment model. *Environmental Modelling & Software*, 25(4):609–610, 2010.

András Bárdossy, Faizan Anwar, and Jochen Seidel. Hydrological modelling in data sparse environment: Inverse modelling of a historical flood event. *Water*, 12(11), 2020. ISSN 2073-4441. https://doi.org/10.3390/w12113242. URL https://www.mdpi.com/2073-4441/12/11/3242.

András Bárdossy, Ehsan Modiri, Faizan Anwar, and Geoffrey Pegram. Gridded daily precipitation data for iran: A comparison of different methods. *Journal of Hydrology: Regional Studies*, 38:100958, 2021. ISSN 2214-5818. https://doi.org/https://doi.org/10.1016/j.ejrh.2021.100958. URL https://www.sciencedirect.com/science/article/pii/S2214581821001877.

András Bárdossy, Chris Kilsby, Stephen Birkinshaw, Ning Wang, and Faizan Anwar. Is precipitation responsible for the most hydrological model uncertainty? *Frontiers in Water*, 4, 2022. ISSN 2624-9375. https://doi.org/10.3389/frwa.2022.836554. URL https://www.frontiersin.org/article/10.3389/frwa.2022.836554.

L. de Vos, H. Leijnse, A. Overeem, and R. Uijlenhoet. The potential of urban rainfall monitoring with crowdsourced automatic weather stations in amsterdam. *Hydrology and Earth System Sciences*, 21(2):765–777, 2017. https://doi.org/10.5194/hess-21-765-2017. URL https://hess.copernicus.org/articles/21/765/2017/.

Lotte Wilhelmina de Vos, Hidde Leijnse, Aart Overeem, and Remko Uijlenhoet. Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. *Geophysical Research Letters*, 46(15):8820–8829, 2019. https://doi.org/https://doi.org/10.1029/2019GL083731. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083731.

DWD. https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/, 2019 2019.

John Ewen, Geoff Parkin, and Patrick Enda O'Connell. Shetran: distributed river basin flow and transport modeling system. *Journal of hydrologic engineering*, 5(3):250–258, 2000.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Tom G. Farr, Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, David Seal, Scott Shaffer, Joanne Shimada, Jeffrey Umland, Marian Werner, Michael Oskin, Douglas Burbank, and Douglas Alsdorf. The shuttle radar topography mission. *Reviews of Geophysics*, 45(2), 2007. https://doi.org/https://doi.org/10.1029/2005RG000183. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005RG000183.

G.H. Hargreaves and Z.A. Samani. Estimating potential evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108(3): 225–230, 1982.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. https://doi.org/10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. https://doi.org/10.1109/MCSE.2007.55.

Dmitri Kavetski, Stewart W. Franks, and George Kuczera. *Confronting Input Uncertainty in Environmental Modelling*, pages 49–68. American Geophysical Union (AGU), 2003. ISBN 9781118665671. https://doi.org/https://doi.org/10.1029/WS006p0049. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WS006p0049.

Hyosang Lee, Daniela Balin, Rajesh Raj Shrestha, and Michael Rode. Streamflow prediction with uncertainty analysis, weida catchment, germany. *KSCE Journal of Civil Engineering*, 14(3):413–420, 2010.

Elizabeth Lewis, Stephen Birkinshaw, Chris Kilsby, and Hayley J. Fowler. Development of a system for automated setup of a physically-based, spatially-distributed hydrological model for catchments in great britain. *Environmental Modelling & Software*, 108:102–110, 2018. ISSN 1364-8152. https://doi.org/https://doi.org/10.1016/j.envsoft.2018.07.006. URL https://www.sciencedirect.com/science/article/pii/S1364815216311331.

LUBW. https://udo.lubw.baden-wuerttemberg.de/public/, 2020.

Michael M McKerns, Leif Strand, Tim Sullivan, Alta Fang, and Michael AG Aivazis. Building a framework for predictive science. *arXiv preprint arXiv:1202.1056*, 2012.

Edom Moges, Yonas Demissie, Laurel Larsen, and Fuad Yassin. Review: Sources of hydrological model uncertainties and advances in their analysis. *Water*, 13(1), 2021. ISSN 2073-4441. https://doi.org/10.3390/w13010028. URL https://www.mdpi.com/2073-4441/13/1/28.

J.E. Nash and J.V. Sutcliffe. River flow forecasting through conceptual models. 1. a discussion of principles. *Journal of Hydrology*, 10: 282–290, 1970.

Benjamin Renard, Dmitri Kavetski, George Kuczera, Mark Thyer, and Stewart W. Franks. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5), 2010. https://doi.org/https://doi.org/10.1029/2009WR008328. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008328.

Rainer Storn and Kenneth Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

The pandas development team. pandas-dev/pandas: Pandas, feb 2020. URL https://doi.org/10.5281/zenodo.3509134.

G. van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

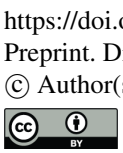
Hydrology and Earth System Sciences
Discussions
Open Access
EGU

R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. https://doi.org/10.1038/s41592-019-0686-2.

Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. https://doi.org/10.25080/Majora-92bf1922-00a.

Soni Yatheendradas, Thorsten Wagener, Hoshin Gupta, Carl Unkrich, David Goodrich, Mike Schaffner, and Anne Stewart. Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resources Research*, 44(5), 2008. https://doi.org/https://doi.org/10.1029/2007WR005940. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007WR005940.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

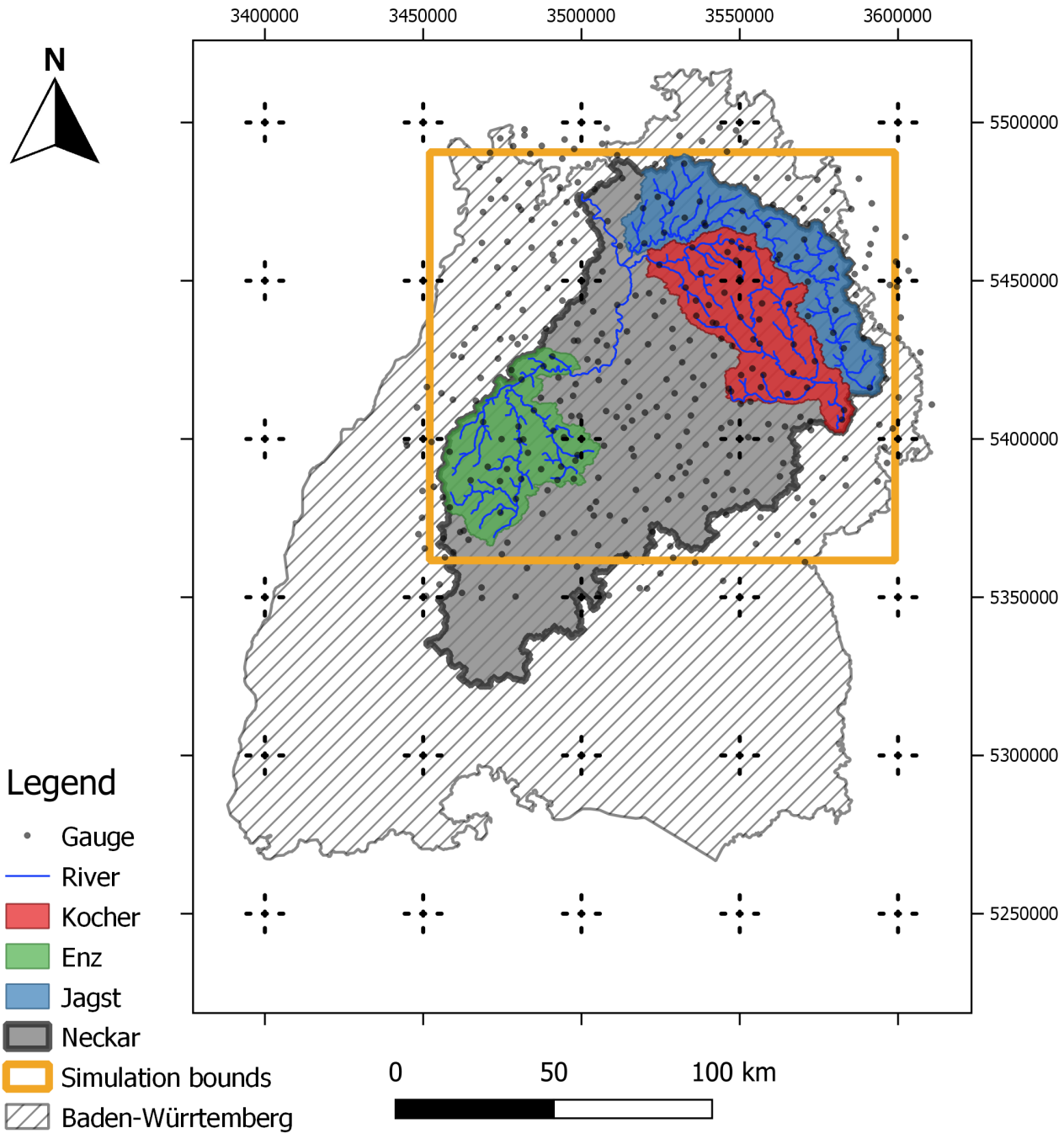


Figure 1. Study area (taken from Bárdossy et al. (2022))

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

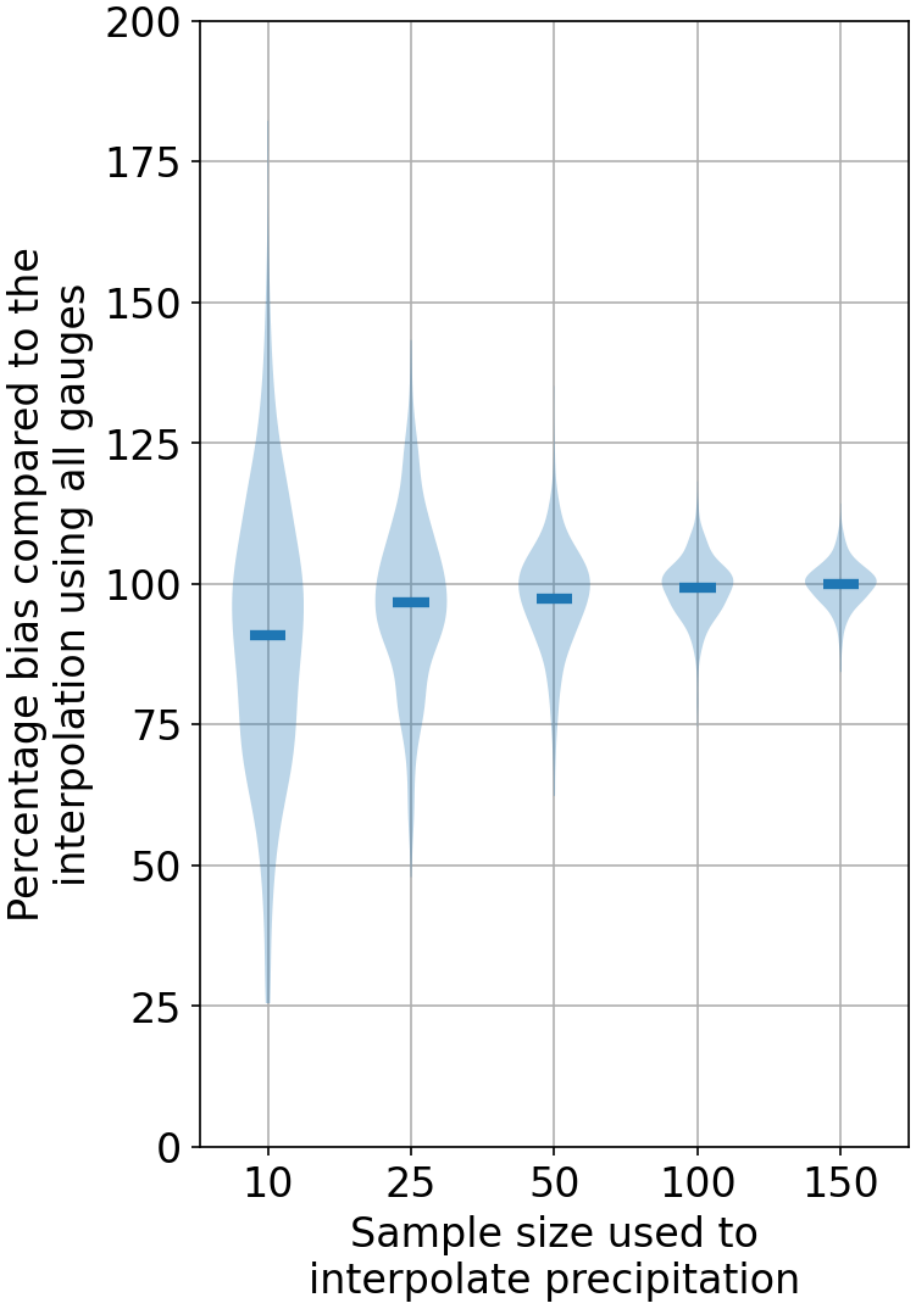


Figure 2. Precipitation bias comparison of the top five largest values due to using fewer points against an interpolation that uses all the points using NN.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

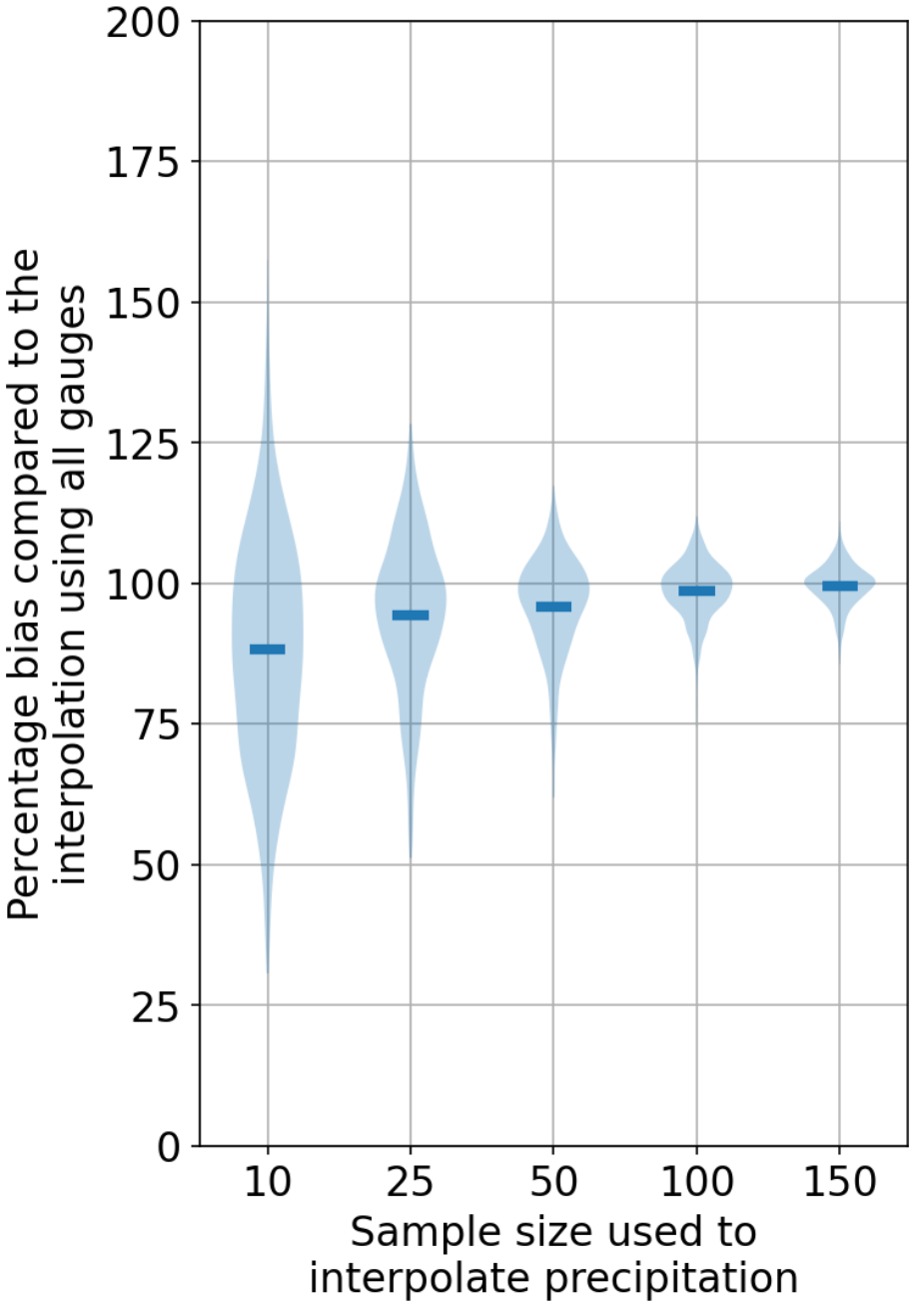


Figure 3. Precipitation bias comparison of the top five largest values due to using fewer points against an interpolation that uses all the points using OK.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

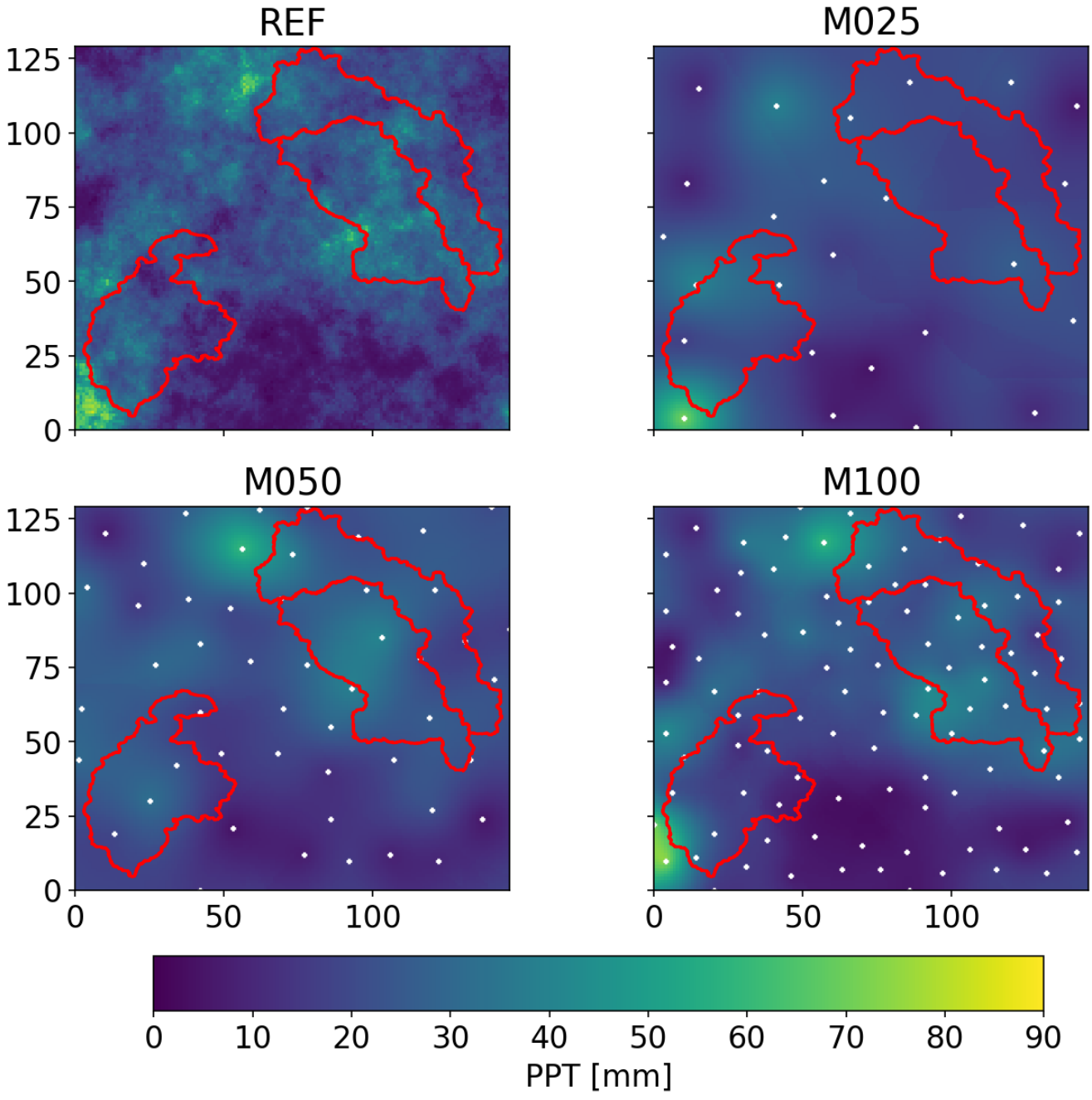


Figure 4. Comparison of precipitation interpolations for a time step with high precipitation with the reference.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

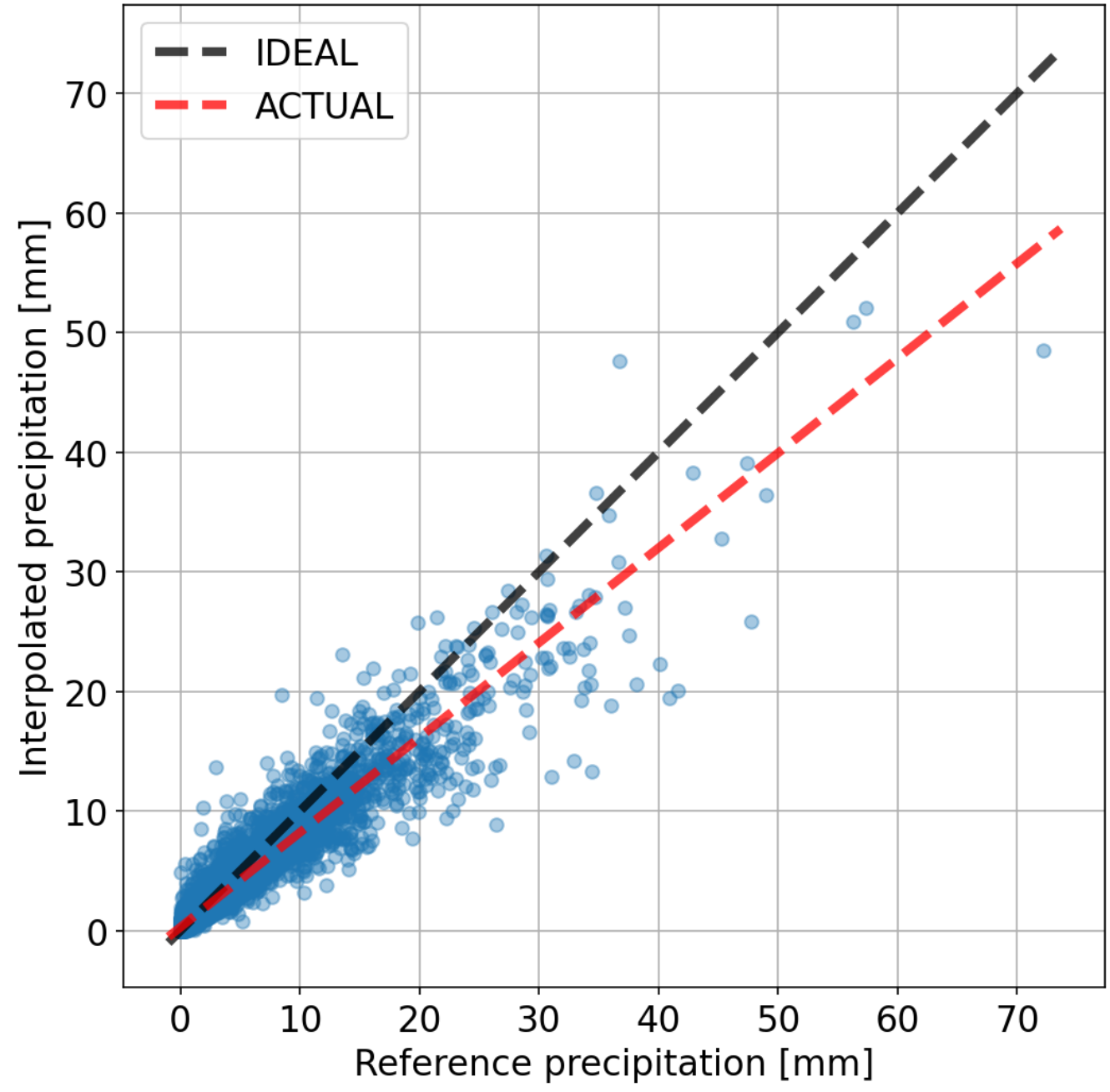


Figure 5. Scatter of reference and lumped interpolated precipitation for one catchment.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

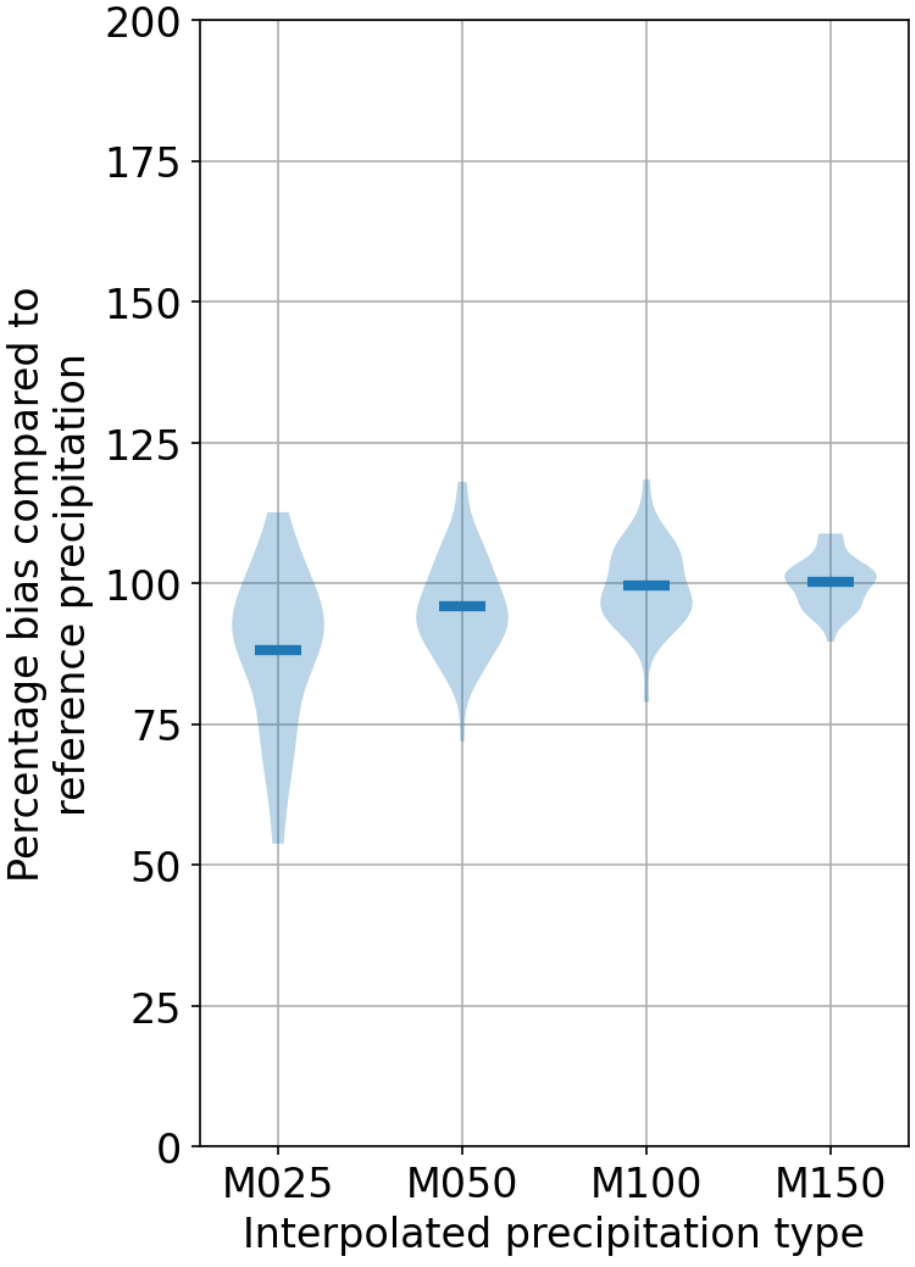


Figure 6. Precipitation bias comparison of various interpolations with respect to the reference for the top five largest values.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

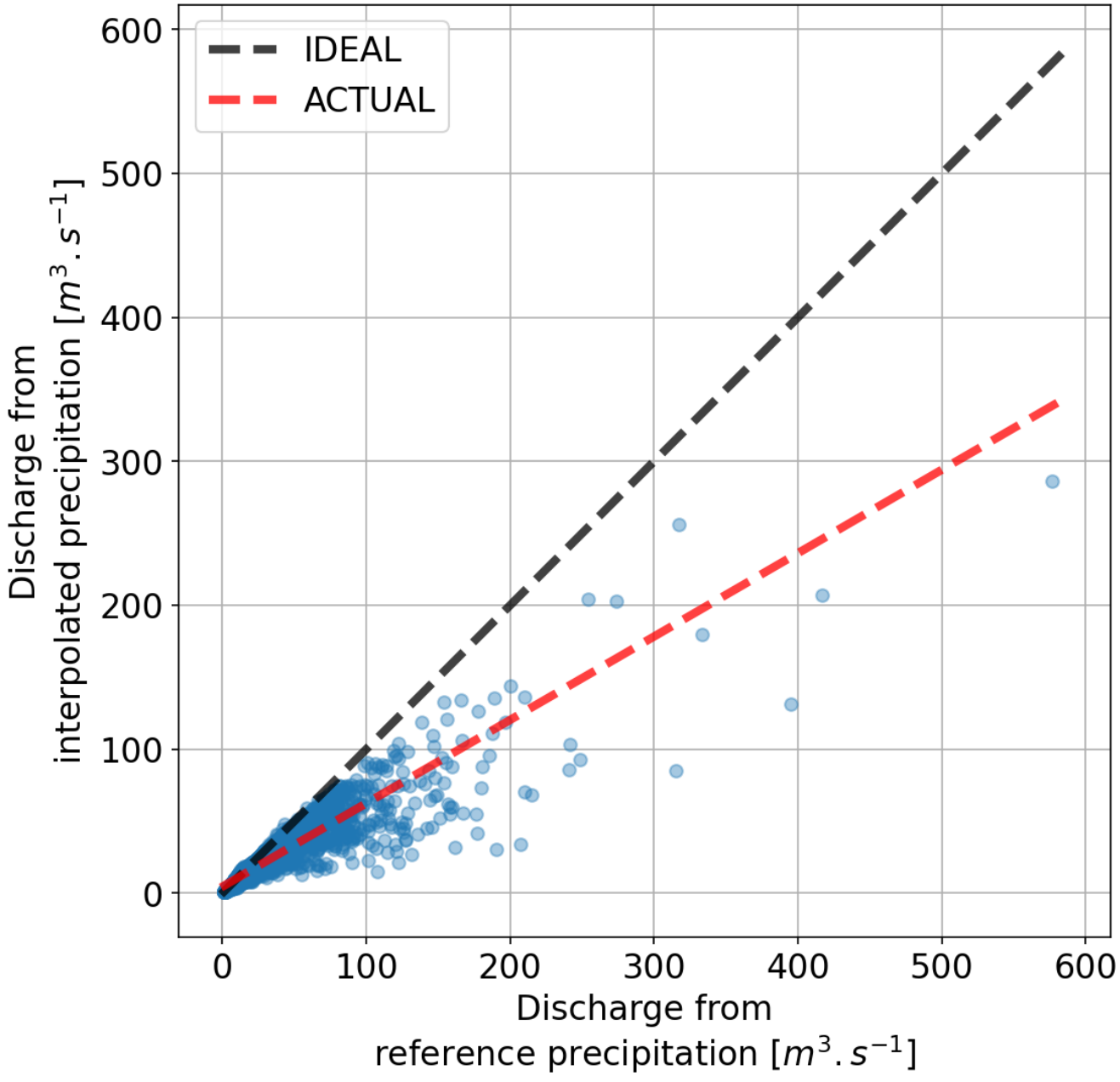


Figure 7. Scatter of discharge using reference and interpolated precipitation for one catchment using SHETRAN.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

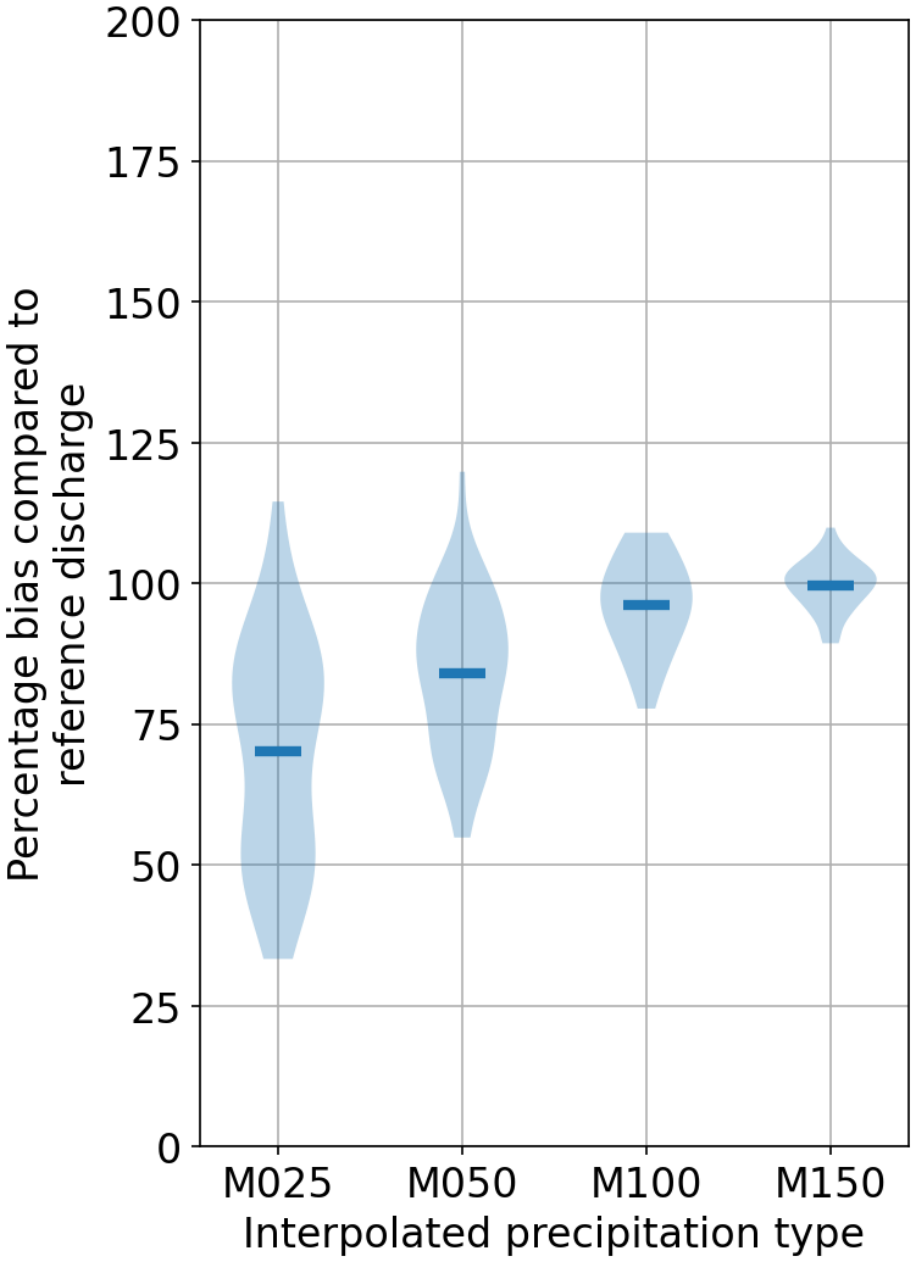


Figure 8. Discharge bias comparison of various interpolations with respect to the reference for the top five largest values using SHETRAN.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

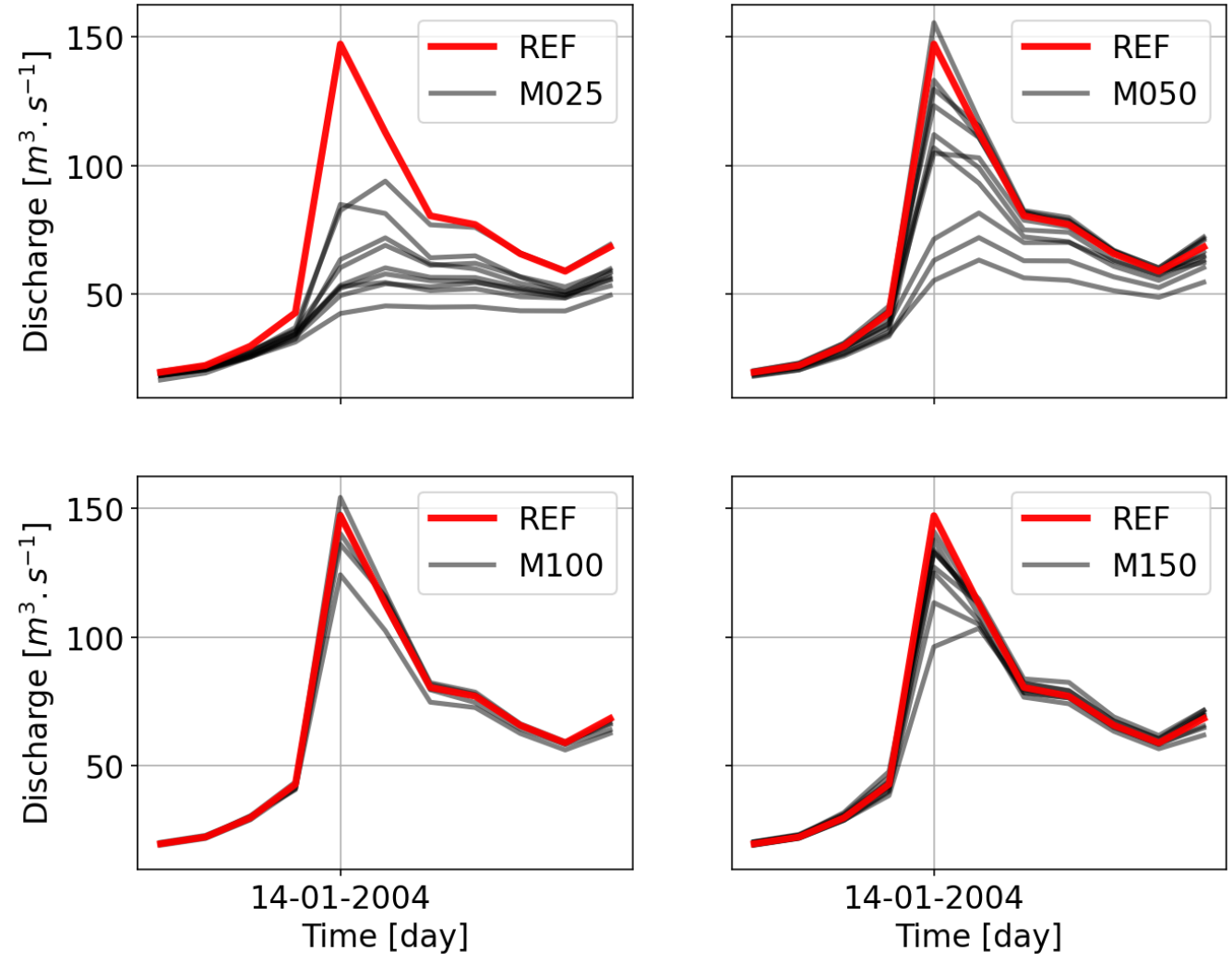


Figure 9. Event hydrograph comparison for various gauging densities using SHETRAN.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

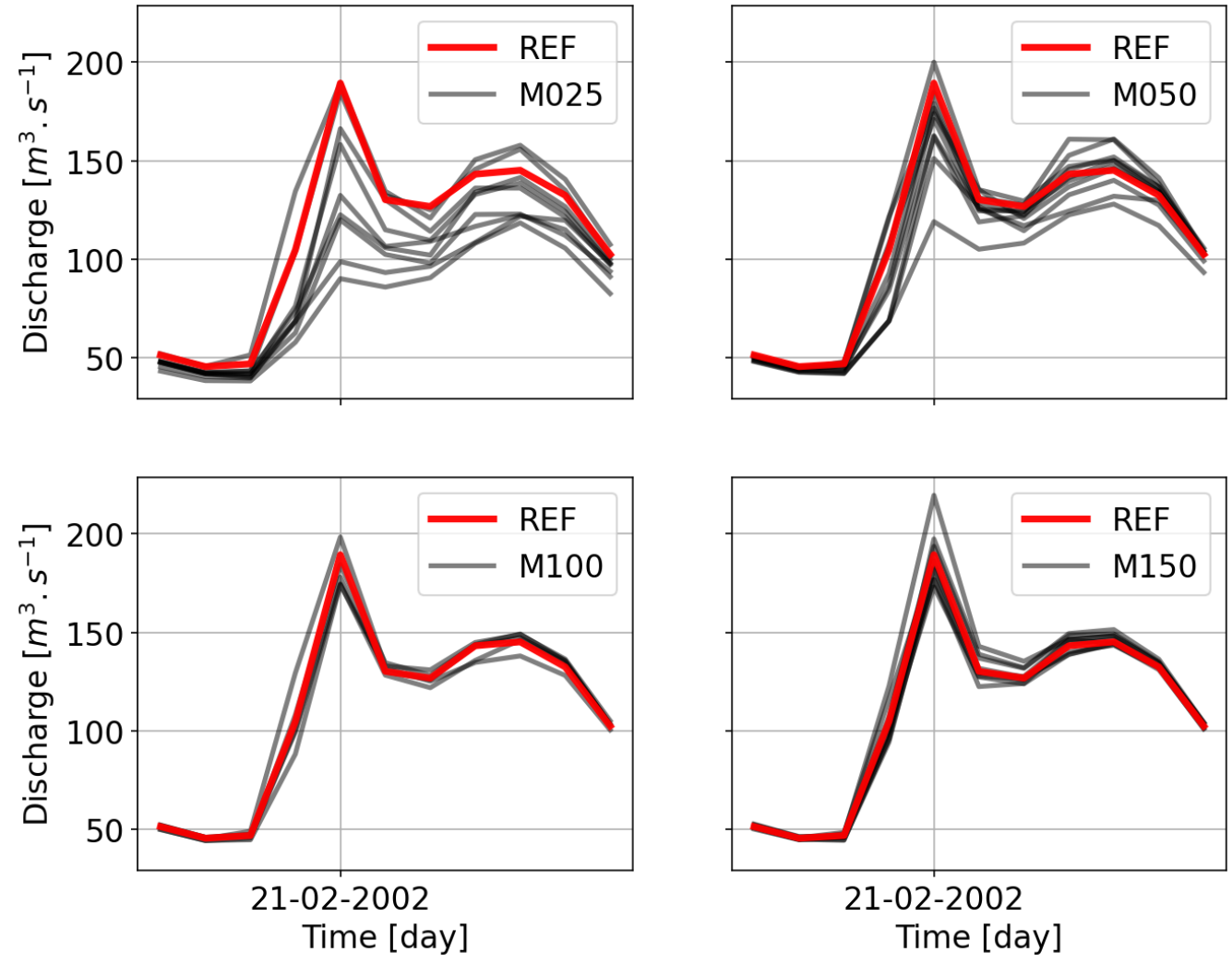


Figure 10. Event hydrograph comparison for various gauging densities using SHETRAN.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

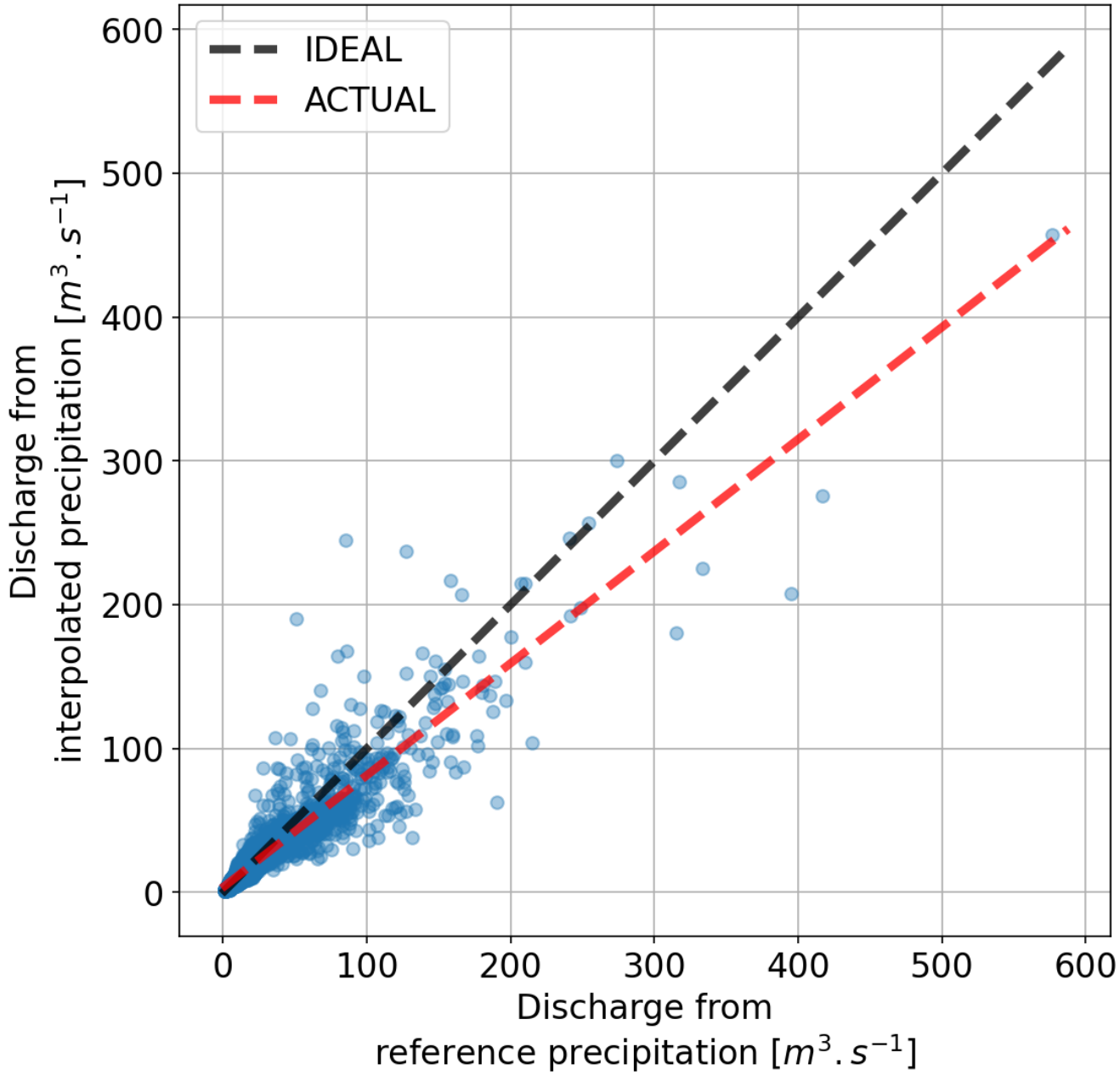


Figure 11. Scatter of discharge using reference and interpolated precipitation for one catchment using HBV.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

Hydrology and Earth System Sciences
Discussions

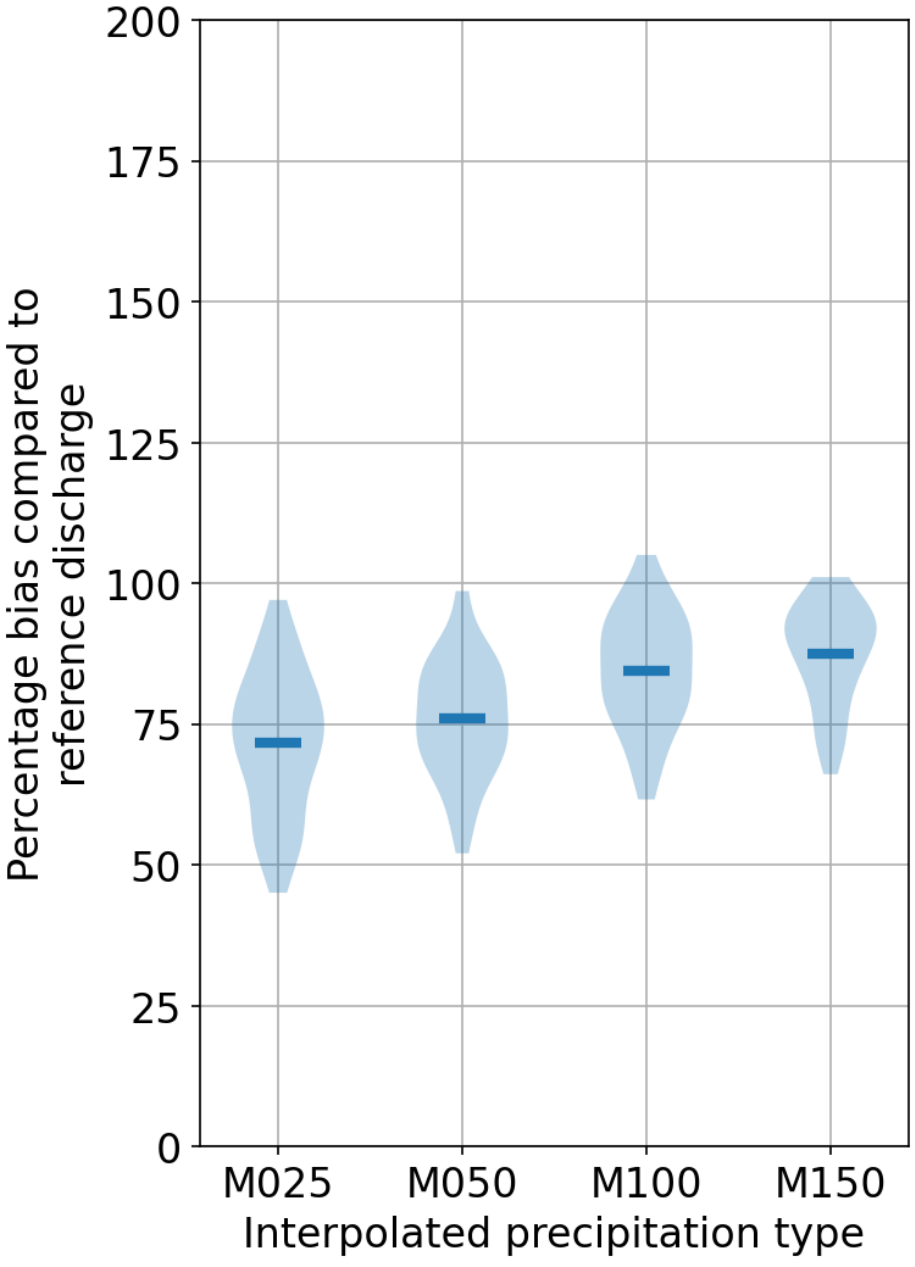
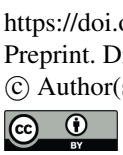


Figure 12. Discharge bias comparison of various interpolations with respect to the reference for the top five largest values using HBV.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022



Hydrology and Earth System Sciences
Discussions

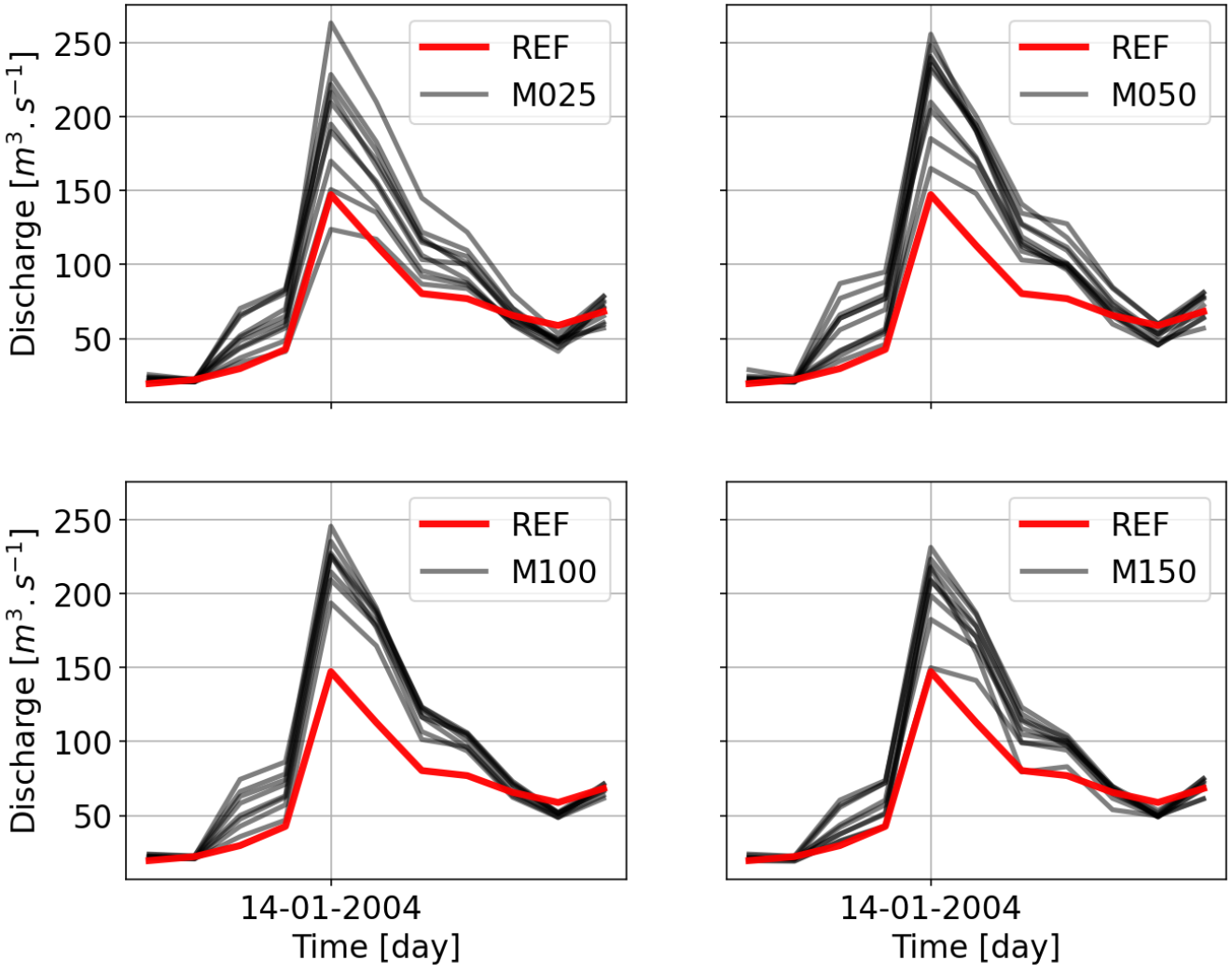


Figure 13. Event hydrograph comparison for various gauging densities using HBV.

https://doi.org/10.5194/hess-2022-281
Preprint. Discussion started: 5 August 2022

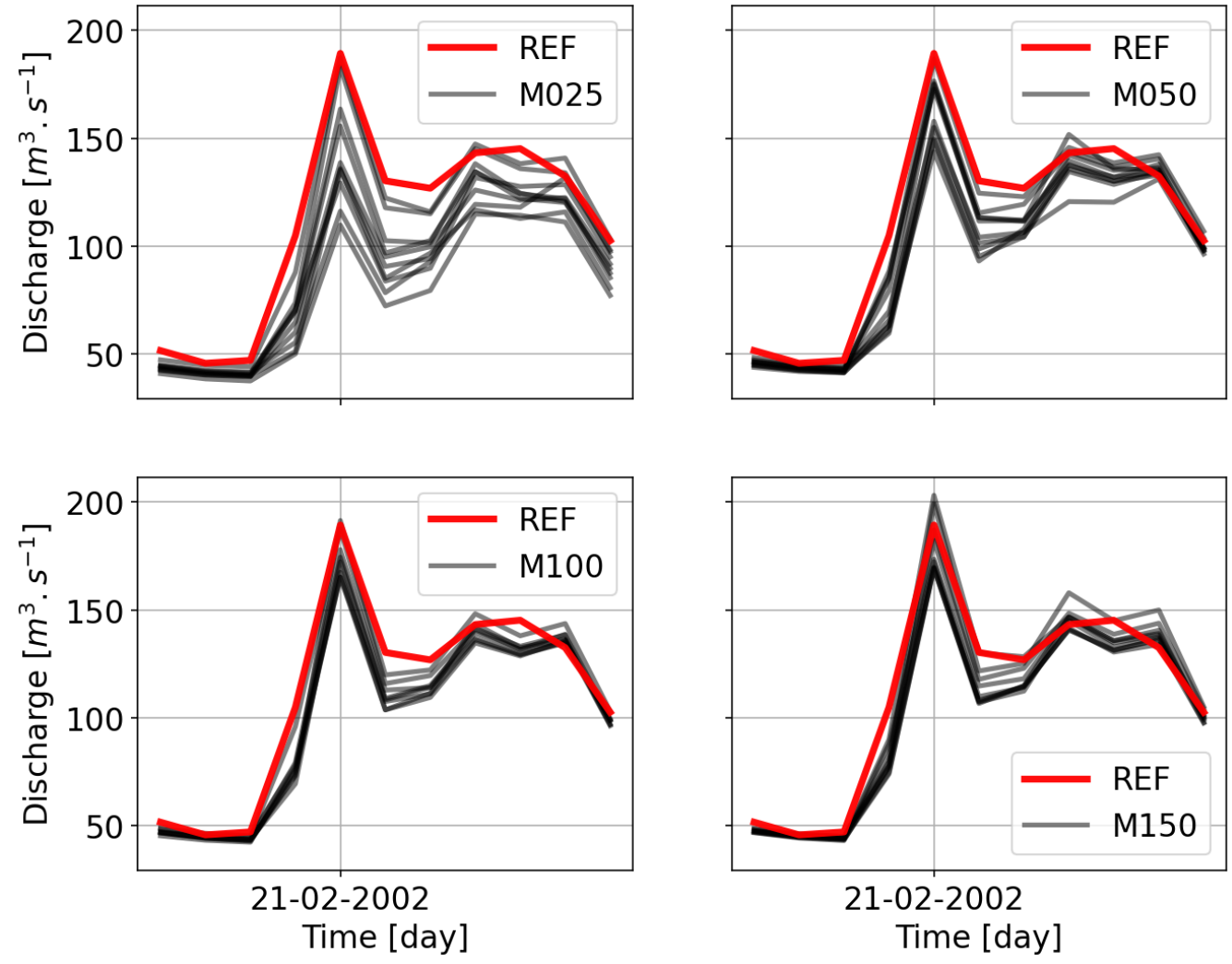


Figure 14. Event hydrograph comparison for various gauging densities using HBV.