Sensitivity of hydrological model to temporal and spatial resolutions of rainfall data

Yingchun Huang[1], András Bárdossy[2], and Ke Zhang[1,3]

[1]College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
[2]Institute for Modelling Hydraulic and Environmental Engineering, University of Stuttgart, Stuttgart D-70596, Germany
[3]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

Correspondence:
Ke Zhang (kzhang@hhu.edu.cn); Yingchun Huang (yingchunhuang@hhu.edu.cn)

Abstract. Rainfall is the most important input for rainfall-runoff models. It is usually measured at specific sites on a daily or sub-daily time scale and requires interpolation for further application. This study aims to evaluate if a higher temporal and spatial resolution of rainfall can lead to improved model performance. Four different gridded hourly and daily rainfall datasets with a spatial resolution of 1km×1km for the state of Baden-Württemberg in Germany were constructed using a combination

5   of data from a dense network of daily rainfall stations and a less dense network of sub-daily stations. A lumped and a spatially distributed HBV models were used to investigate the sensitivity of model performance to the spatial resolution of rainfall. The four different rainfall datasets were used to drive both lumped and distributed HBV models to simulate daily discharges in four catchments. Main findings include (1) a higher temporal resolution of rainfall improves the model performance if the station density is high; (2) a combination of observed high temporal-resolution observations with disaggregated daily rainfall

10   leads to further improvement of the tested models; (3) for the present research, the increase of spatial resolution improves the performance of the model insubstantially or only marginally in most of the study catchments.

1   Introduction

Rainfall is a primary driver of hydrological models and can impact catchment runoff response significantly (Obled et al., 1994; Ly et al., 2013). Rainfall is usually measured by standard rain gauges or wireless telemetering pluviometers over a

15   period of time (e.g. daily, sub-daily). Uncertainties in rainfall estimation for a catchment can occur due to instrument errors as well as spatial and temporal variability of rainfall. The latter are the main sources of uncertainties in model simulation and flood forecasting (Beven, 1998; Berne et al., 2004). The spatial variability of rainfall strongly influences the timing and shape of hydrograph, while the temporal variability mainly affects the peak of flood wave (Singh, 1997). The improvement of flood simulation requires understanding the sensitivity of the rainfall-runoff models to rainfall input data. Over the past

20   decades, various methods have been used to obtain the spatial distributions of rainfall based on rain gauge data and catchment characteristics (Goovaerts, 2000; Jeffrey et al., 2001; Hofierka et al., 2002; Haylock et al., 2008; Ly et al., 2013). Kobold and Brilly (2006) derived hourly areal rainfall interpolated from various numbers of rain gauges to quantitatively assess the sensitivity of peak flow to the uncertainty of rainfall data using an HBV model. They found that the error in rainfall may lead

to even greater error in flood peak. Bardossy and Das (2008) also studied the impact of spatial variability of rainfall by varying the distribution of rain gauge network. They found that the transferabilities of model parameters calibrated based on sparse and dense rainfall information are very different. Das et al. (2008) used four different model structures to simulate daily runoff in central Europe. Results indicated that the semi-distributed and semi-lumped models outperform the lumped and distributed model structures. They suggested that the lack of spatial information is responsible for the low efficiency of distributed model. Xu et al. (2013) indicated that the increase of rain gauge network density can improve the model performance, but no apparent improvement was observed when the number of rain gauges exceeded a threshold. Lobligeois et al. (2014) found that simi-distributed models outperform the lumped models when rainfall is highly variable over simulation catchment, but they perform similarly when rainfall is relatively uniform. Emmanuel et al. (2015) proposed rainfall variability indexes to characterize the influence of spatial variability rainfall and implemented this approach in the model simulation for the Cevennes catchment in France (Emmanuel et al., 2017). They found that higher spatial resolution of rainfall could achieve better model performance. We can learn from these researches that the sensitivity of model performance to the spatial resolution of rainfall seems different for some of the case studies. However, increasing spatial resolution in model simulation can lead to considerable complexity of model structure and require much more data than using a lumped version.

The rainfall-runoff response of a catchment is also strongly impacted by the temporal variability of rainfall (Bárdossy and Pegram, 2016). High temporal resolution rainfall data are collected at pluviometer stations with telemetry at sub-daily time resolutions. Sub-daily data often have poor quality caused by equipment malfunction or misreading. Compared with sub-daily rainfall data, daily data tend to be more available and reliable, cover a longer duration of time periods. Disaggregating daily into sub-daily data offers a potential solution to accurately capture the temporal variability of rainfall (Parkes et al., 2013; Bardossy and Pegram, 2016). Pui et al. (2012) compared three different approaches for disaggregating daily rainfall into sub-daily series and found the resampling method is the best one for rainfall disaggregation. Bárdossy and Pegram (2016) used Gaussian Copula-based model for disaggregating daily data to infill the gap of sub-daily data, and they found that this conditional disaggregation of rainfall is reliable and applicable in various regions. Breinl and Di Baldassarre (2019) applied a spatial method of fragments to disaggregate daily rainfall into hourly values. Kobold and Brilly (2006) found that calibrating hydrological models with sub-daily time steps can significantly improve the accuracy of flood forecasting.

Some studies focus on both the spatial and temporal resolution of rainfall. Bruneau et al. (1995) found that the temporal and spatial resolutions of rainfall used as the inputs of hydrological models can have considerable influence on the model efficiency and parameter values. Booij (2002) found the influence of rainfall spatial resolution is greater than temporal resolution in terms of simulation of extreme flows. Meselhe et al. (2009) pointed out that physically based models are more sensitive to the spatial and temporal resolution of rainfall data than conceptual models. Zhu et al. (2018) found that the spatial variability of rainfall is much more sensitive to model performance for catchments larger than $2000km^2$ under dry soil condition, and floods in small catchments are more influenced by the temporal variability of rainfall. So far, more efforts have been invested in improving the spatial or temporal resolution of rainfall, but there are less studies on quantification and direct comparison of the catchment dynamic responses driven by different rainfall temporal and spatial resolutions.

The overarching aim of this study is to understand the dependency of hydrological model performance on the rainfall data. The specific research objectives are three-fold: (1) investigate the effects of rainfall data quality on model performance, (2) examine the sensitivity of model performance to different spatial and temporal resolutions of rainfall data using two different model spatial configurations, and (3) explore the possibility of improving model performance on a daily scale. The paper will be followed by section 2 to describe the study area and the rainfall datasets used in this research. In section 3, the hydrological model and the calibration method are explained. Section 4 presents the results and discussion of this work. The conclusions and outlook are in section 5.

2   Study area and hydrometeorological datasets

This study area is located in a semi-humid region in the Baden-Württemberg state of Germany (Figure 1) with temperate monsoon climate of mild winter and warm summer. Elevation of this region ranges from 85m to 1493m above sea level. The heterogeneity of climate characteristics is mainly due to the great variability of elevations within the study area. The annual mean air temperature in Baden-Württemberg is about 10.2 °C. Rainfall is evenly distributed throughout the year. However, its seasonality shows a weak trend. The monthly rainfall is highest in June and lowest in October. The meteorological data used in this study were provided by the German Weather Service (DWD). Daily air temperature data required for the rainfall-runoff model were interpolated on a 1×1 km^2 grid from the observations using the External Drift Kriging algorithm (Ahmed and De Marsily, 1987). The topographical elevation was taken as external drift (Hundecha and Bárdossy, 2004; Das et al., 2008). The long term monthly potential evapotranspiration and the average air temperature were used to compute the daily potential evapotranspiration using the Hargreaves and Samani method (Hargreaves and Samani, 1985).

Rainfall data from a dense network of daily rainfall stations (62 km^2/station in 1991) and from a less dense network of sub-daily stations (144 km^2/station in 1991) with high resolution rainfall observations were used for this study. All data are available for the time period 1991-2010. The number of available daily stations and sub-daily stations varies according to different time period. Figure 2 illustrates the number of available observation stations in Baden-Württemberg between 1991 and 2010. It can be seen from the graph, more than 430 daily stations but only 30 sub-daily stations were available in 1991. The total number of daily stations decreased to 250 around 2003 and remained stable for the subsequent years. The number of sub-daily stations has been increasing throughout this period and experienced a sharp increase from 100 to 200 in 2005. Four different rainfall datasets were generated and explained as follows.

1. High temporal resolution observed rainfall was aggregated to hourly and then interpolated to 1×1 km^2 grids using the ordinary Kriging algorithm (Matheron, 1963). The correlation function obtained from the cross-correlations of the hourly time series was used as a basis for the variogram. This set is referred to as Sparse Hourly (SH) set.

2. Observed daily rainfall combined with the daily aggregations of the high temporal resolution data were used to create 1×1 km^2 gridded datasets using the ordinary Kriging algorithm. The variogram was based on the cross-correlations of the daily time series. This set is referred to as Dense Daily (DD) set.

3. High resolution rainfall was aggregated to daily time steps and interpolated subsequently for a 1×1 km^2 grid using the ordinary Kriging. The variogram was based on the cross-correlations of the aggregated daily time series. This set will be referred as Sparse Daily (SD) set.

4. Observed daily rainfall combined with the hourly aggregations of the high temporal resolution data were used to create a 1×1 km^2 grid using the disaggregation method rescaled ordinary Kriging (Bárdossy and Pegram, 2016). The variogram was based on the cross-correlations of the hourly time series. This set is referred to as Dense Hourly (DH) set.

Figure 3 shows the flow chart of the data collection and process. The DD and SD sets are the daily aggregations of the DH and SH sets. Note that DH is a dataset combining hourly and disaggregated daily gauge data. One of the research questions raised here is to find out if disaggregation leads to an improvement of model performance. Comparisons of the model performance using the inputs of (SD, SH) and (DD, DH) pair will reveal the effect of temporal resolution. Meanwhile, comparison between (SD, DD) and (SH, DH) will show the influence of the rainfall observation network density on the model performance.

Four mesoscale catchments (Figure 1), namely Rottweil, Schwaibach, Pforzheim and Kocherstetten, were selected from the upstream region of the state for testing the sensitivity of model performance to the four different rainfall datasets. The daily streamflow record of these catchments was collected for the period 1991-2010. The basic characteristics for the study catchments are listed in Table 1. These catchments range in size from 417km^2 to 1300km^2, along with large differences in elevation and annual precipitation. It can be seen clearly from the map that these four catchments have different rain gauge densities, the Schwaibach catchment located in the mountainous area with elevations ranging from 190m to 1028m has the lowest density of rain gauge network and the highest annual precipitation. Rottweil and Kocherstetten have similar climate conditions in terms of annual precipitation and runoff, but the catchment size of Kocherstetten is almost three times of Rottweil. Pforzheim has the smallest drainage area and the lowest amount of precipitation.

3 Model and methodology

3.1 Model structure

The conceptual HBV model was developed in the 1970s by the Swedish Meteorological and Hydrological Institute (SMHI) (Bergström and Forsman, 1973). Thanks to its simplicity, low demand of inputs and small number of model parameters, the HBV model has been widely used for rainfall-runoff simulation and flood forecasting. Figure 4 represents the structure diagram of the HBV model (Singh, 2010). There are three main modules in the HBV model, namely snow routine, soil moisture routine and runoff routine (Hartmann, 2007; Singh, 2010).

In the snow routine, the snow accumulation and melting process is estimated by the relatively simple degree-day method (Rango and Martinec, 1995) with two parameters: degree day factor (*DD*) and threshold temperature for snow/rain (*TT*) (as shown in Equation 1). The measured precipitation is supposed to be solid (snowfall) if the air temperature is lower than threshold temperature, otherwise, precipitation appears liquid state (rainfall) if the weather is warmer than the threshold value.

$$Snowmelt = DD \cdot (T - TT), \quad \text{if} \quad T > TT \tag{1}$$

In the HBV model, soil moisture storage is decided by balancing rainfall and evapotranspiration according to two soil moisture constants: permanent wilting point (*PWP*) and field capacity (*FC*). The soil wetness index, defined as the ratio of direct runoff to effective precipitation ($\frac{\Delta Q}{\Delta P}$), is expressed as:

$$\frac{\Delta Q}{\Delta P} = (\frac{SM}{FC})^{Beta} \tag{2}$$

5   where *SM* denotes the actual soil moisture and *Beta* the proportion of effective precipitation contributing to runoff for given soil moisture. The Penman equation is used to estimate the potential evapotranspiration according to the long-term monthly mean air temperature (T_M) and long-term monthly averaged potential evapotranspiration (PE_M) (Penman, 1948):

$$E_{tp} = (1 + C(T - T_M))PE_M \tag{3}$$

where C is the evapotranspiration coefficient. The actual evapotranspiration (E_{ta}) can be estimated as:

10   $$E_{ta} = \begin{cases} E_{tp} & \text{if } SM > PWP \\ \frac{SM}{PWP} \cdot E_{tp} & \text{else} \end{cases} \tag{4}$$

As shown in Equation 2, runoff is calculated by a non-linear function based on excessive effective precipitation and actual soil moisture. The runoff concentration process consists of upper and lower reservoirs with five parameters:

$$Q_0 = K_0(S_1 - HL) \tag{5}$$

15   $$Q_1 = K_1 S_1 \tag{6}$$

$$Q_d = K_d S_1 \tag{7}$$

$$Q_2 = K_2 S_2 \tag{8}$$

20   The runoff is divided into surface flow (Q_0), interflow (Q_1) and base flow (Q_2) with three recession coefficients K_0, K_1 and K_2, along with a conceptual threshold water level (*HL*) for generating surface flow. The two parallel reservoirs are connected in the form of percolation storage (Q_d) from upper reservoir to the lower one with the parameter of percolation constant K_d. A transformation function with the triangular weighting parameter *MAXBAS* is used to smooth the total runoff ($Q_0 + Q_1 + Q_2$) to obtain discharge at the outlet.

25   In this study, for investigating the sensitivity of model performance on the spatial resolution of input variables, two HBV models with different spatially configurations were applied: lumped HBV and spatially distributed HBV, respectively. In the

lumped model, precipitation, temperature and potential evapotranspiration were assumed uniformly distributed within a catchment and all the processes were calculated for the whole catchment. Previous studies have indicated that the elevation is an important reason for the spatial differentiation of meteorological variables, including temperature, precipitation, evapotranspiration and snow melt are in reality not uniformly distributed within a catchment. They often exhibit dependence with elevation.

The spatially distributed HBV model used in this study divides a catchment into several zones based on elevation. The 1×1 km^2 grid based rainfall and temperature data were averaged for each elevation zone. In the spatially distributed model, the parameters associated with the snowmelt and soil moisture modules were calibrated for each elevation zone. The parameters associated with the runoff response module were calibrated for each catchment similarly to the lumped model (Das et al., 2008).

Out of the 15 parameters within the HBV model, 9 parameters were calibrated in this study. Table 2 lists the initial upper and lower limit of the to-be-calibrated parameters using historical data. The data depth based parameter optimization method- Robust Parameter Estimation (ROPE) algorithm (Bárdossy and Singh, 2008) was applied for model parameter optimization. The ROPE approach could lead to a certain number of model parameters with ideal model performance (Bárdossy et al., 2016). For this study, each simulation results in 10,000 heterogeneous parameter sets with similarly acceptable model performance.

3.2   Performance criteria

Previous studies have shown that model performance strongly depends on the selection of performance criteria (Gupta et al., 2009). The model simulations corresponding to the model parameters using different objective functions differ considerably as they have different focuses (Bárdossy et al., 2016). The purpose of this study is to investigate the sensitivity of conceptual model to rainfall variability, and according find effective ways to improve the precision of flood forecasting. Since high flow is extremely important for flood forecasting, the Nash-Sutcliffe (NS) coefficient (Nash and Sutcliffe, 1970), one of the widely used indicators, was used in this study to assess the model performance based on observed discharge. NS coefficient focuses on high flow as it evaluates the squared difference between simulated and measured streamflow. It can be calculated using the following equation:

$$NS = 1 - \frac{\sum_{t=1}^{T} \left(Q_o(t) - Q_m(t) \right)^2}{\sum_{t=1}^{T} \left(Q_o(t) - \bar{Q}_o \right)^2} \tag{9}$$

where $Q_o(t)$ and $Q_m(t)$ are the observed and simulated discharge, respectively, and $\bar{Q}_o$ is the mean of observed discharge.

The Mean Square Error (MSE) of the flow for the time period of the observed discharge higher than the 10th percentile of flow was used to assess the flood forecasting ability of the models:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Q_0(i) - Q_m(i))^2 \tag{10}$$

where $Q_o(i)$ and $Q_m(i)$ are the observed and modeled discharges when the observed discharge is higher than the 10th percentile of flow.

3.3 Model calibration experiments

A split sample calibration methodology was applied in this study to divide the data into two 10-year periods: 1991-2000 and 2001-2010. Model calibration was carried out for both time periods and then a cross-validation analysis was performed. For each calibration run, the first water year data was used as a warm-up period and was not used to evaluate the model performance.

In this study we investigated the impacts of using different methods for spatial interpolation of hourly rainfall data on model performance. The four rainfall datasets were used as input variables for model calibration and validation. In all modeling experiments, daily mean temperature and potential evapotranspiration were used as inputs. This is to isolate the effects of different rainfall inputs on the model performance. The effects of the temporal and spatial resolutions of the rainfall inputs on the model performance were assessed in terms of NS coefficient and the MSE of the high flow. We conducted experiments of model calibration for a lumped and a spatially distributed HBV model using hourly and daily input variables, respectively. For the spatially distributed model structure, a contour interval of 100m was used to divide a catchment into different elevation zones. Note that all the model calibrations were performed on the basis of simulating daily discharge.

We investigated if the combination of daily and hourly models can lead to better prediction in streamflow. It is interesting to investigate the similarities of different temporal resolution. Therefore, the common calibration approach was used to calibrate the daily and hourly models simultaneously. This approach may identify robust model parameters that are applicable using different temporal resolutions. The common calibration approach is a multi-objective optimization function and the compromise programming method (Zeleny, 1981) was used to formulate the objective function:

$$O(\theta) = \sum_{i=1}^{n} (NS_i^* - NS_i(\theta))^p \tag{11}$$

Here index i denotes the type of temporal resolution, NS_i^* means the optimal model performance which can be represented by the individual calibrated model performance. Here we aim to minimize the value of objective function $O(\theta)$. For the balancing factor p, a moderately high $p = 4$ was given in this study. More details about the common calibration of hydrological models' strategy can be found in Bárdossy et al. (2016).

4 Results and discussion

4.1 Comparison of the rainfall datasets

The quality of the rainfall datasets was assessed and compared for the four selected catchments. As the SD and DD sets are daily aggregations of the SH and DH sets, we only compared the daily rainfall sets SD and DD for both calibration periods (Figure 5). It can be seen clearly from the figures that the interpolated rainfall datasets display some difference for all study catchments. The asymmetry of the scatterplots is evident for the period 1991-2000. In general, the DD dataset leads to higher value than the SD dataset. It is mainly because the low density of sub-daily observations during the period of 1991-2000 leads to large errors in the spatial interpolation of rainfall. This is especially the case for Schwaibach catchment which varies strongly

in geographical elevation (from 190m to 1028m. For the period of 2001-2010, the SD and DD sets are in closer agreement due to higher density of sub-daily gauges.

4.2 Results of calibration and validation

As described in section 3.3, for the selected catchments, model calibrations were carried out using four rainfall datasets for both lumped and spatially distributed HBV models. Two 10-year time periods 1991-2000 and 2001-2010 were used for calibration and cross-validation. In total 16 calibration runs and 16 validation runs were performed for each catchment. As mentioned before, each simulation obtained 10,000 parameter sets with similar model performance. We then used the mean value of the 10,000 model performances to quantify the model performance.

Table 3 lists the average value of the NS model performance for the four selected catchments using lumped HBV model and Table 4 lists the simulated NS performance for spatially distributed version of the model, respectively. The results show that all four datasets can reproduce relatively accurate historical daily streamflow series for all selected catchments. Results also show that the model performances vary across catchments. The Kocherstetten catchment generally performs the best with an average NS value of 0.84 for all simulations, while Pforzheim has the worst mean NS performance of 0.58 for all calibration runs. Moreover, for a specific catchment, the calibrated models perform differently for different data period. For the Schwaibach and Pforzheim catchment, the calibrated model performance for the period of 2001-2010 is better than the performance for the time period 1991-2000 for most of the datasets. This might be due to the increasing rain gauge density inside or near the catchment and the quality of rainfall data with the development of time and technological progress. In particular, the model calibrations for the period 1991-2000 of the Schwaibach catchment using the sets SH and SD perform very poorly for both calibration and validation. NS coefficient using SH and SH inputs is about 0.3 less than the results of the sets DH and DD. This indicates that systematic interpolated rainfall errors have critical influence on model calibration.

We then analyzed if the model is robust for simulating high flows. Tables 5 and 6 list the mean square errors of the top 10th percentile of flows for the lumped and spatially distributed model, respectively. Figure 6 shows the flow duration curve for the natural logarithm of simulated and observed discharge for all study catchment for the years between 2001 and 2010. Figure 7 shows the corresponding results for flows higher than the 10th percentile of flow. Results indicate that for most of the calibration runs, the set DH performs the best for the high flow, followed by set SH, set DD performs a little weaker than set SH, while set SD has the worst performance in flood simulation.

4.3 Model performance using different temporal resolutions of rainfall data

Firstly, the model performance of different temporal resolutions of rainfall was compared for four datasets and two model spatial configurations. For the pairwise comparison, all the conditions are the same in the model except for the rainfall temporal resolution (hourly and daily). The results of the sparse sets and dense sets are separated here. Figure 8 compares the model performance of using hourly and daily rainfall inputs that were interpolated using only high-resolution rainfall observations (SH, SD). Figure 9 compares the results from the rainfall inputs that incorporated observed daily value with high-resolution observations (DH, DD). The result shows that all the scatters are lying below the diagonal line for the different level of

observation density. For both calibration and validation periods, the simulations using hourly input data outperform the ones based on the daily resolution. For the dataset with low observation network density, the averaged NS of set SH is about 0.73 for the calibration period and 0.68 for the validation period, while the mean NS coefficient calibrated using SD set is 0.67 and 0.6, respectively. The higher observation density datasets show a similar tendency. The mean NS of using DH set is around 0.79 for calibration and 0.77 for validation, while that of set DD is 0.72 and 0.69, respectively. The fact that the hourly scale model performs better than the daily model suggests that the dynamic runoff of catchment could be better simulated with a higher temporal resolution of rainfall. According to the distances from the diagonal to the scatter plots, we can observe that the difference in model performance for different temporal resolutions is larger for the catchments with relatively low NS model performance, such as Schwaibach and Pforzheim. For Rottweil and Kocherstetten, the performance of hourly calibrated model is only slightly better than the daily model.

4.4 Model performance in terms of observation density

The rainfall gauge network density has significant impact on model simulation and parameter optimization. Figure 10 shows the simulated NS coefficient of the model simulations using the daily input data interpolated using different densities of rain gauge networks. It shows obviously from the location of points that the simulated model performance using the DD set is generally better than that using the SD set for both calibration and validation period. The averaged NS model performance of DD and SD sets are 0.71 and 0.64, respectively. The model performance using hourly inputs shows similar trend as that using daily inputs. As shown in Figure 11, the model using the DH set outperforms the one using the SH set. These results demonstrate that high rain gauge density lead to improvement of model performance at both daily and hourly time resolution.

Figure 12 illustrates the cumulative distribution function of NS coefficient using sets SD, SH and DH for model calibration (left) and validation (right). As can be seen clearly from the curves, if rainfall data comes from a sparse network of sub-daily stations, use of higher temporal resolution datasets (the SH set) leads to better model performance than using lower resolution ones (the SD set). Simulation of daily streamflow can benefit from running the model at a higher temporal resolution. In addition, the combination of observed sub daily rainfall with disaggregated daily rainfall (the DH set) leads to a further improvement of daily streamflow simulation.

4.5 Model performance in terms of spatial resolution of rainfall data

The model performance was compared between the lumped and spatially distributed HBV model when they were driven by different rainfall datasets. Figure 13 compares the NS model performance for calibration (left) and validation (right) periods. The correlation between model performance and the spatial resolution of model seems not clear for the study catchments. For some simulations, the spatially distributed model outperforms the lumped, especially for the catchments having high NS coefficient, despite the increase in model performance being only marginal. However, for the catchments with relatively poorer model performance, the lumped model could even lead to slightly better performance than the semi-distributed model, especially for the validation period when the difference seems larger than the calibration period. It indicates that for model validation, the model parameters estimated by distributed HBV model shows weaker transferability. Possible explanation

for this case could be that the distributed model has a larger number of parameters to be calibrated and the parameters are underestimated during the calibration period. We conclude that the improvement in spatial resolution of model structure did not enhance the model performance, which is surprising since higher spatial resolution and more model parameters are expected to improve the model performance. Our results confirm the findings of Das et al. (2008) that the distributed models do not necessarily improve model performance.

The distributed model did not perform better than the lumped model in this study. This could be because the catchment underlying surface information and/or the calibration procedure was not sufficient for identifying optimal distributed model parameters. A second reason could be the temporal resolution of the rainfall inputs is not sufficient for the distributed model.

4.6 Common model calibration with different temporal resolutions

As shown before, the combination of hourly and daily gauge data leads to the improvement of data quality as the model using sets DH and DD has better performance than using sets SH and SD. Furthermore, common calibration of the lumped HBV model was performed for the sets DH and DD to identify model parameters good for both hourly and daily time steps. It is important to note that the parameters (DD, K_0, K_1, K_d and K_2) that are dependent on time steps should be converted according to the simulation step of the model. The common calibration was performed for the two time periods separately, and a cross-validation analysis was performed as well. The common calibration and validation results were compared with the individual calibration (Figure 14). For the calibration period, the common calibration always leads to slightly weaker performance for all datasets. For three of the DD datasets, model performances of common parameters are similar to individual calibration results. The average loss of NS coefficients over all catchments is about 0.02 for set DH and 0.01 for set DD. For the validation period, it can be seen from the scatter plots that the common parameters outperform the individual ones for about half of the all simulations. It suggests that parameters values obtained using the common calibration approach based on different time steps can improve the temporal transferability of models. The reason for the robustness of common parameters might be that common calibration strategy can provide more information for identifying model parameters.

The calibrated model parameters using daily rainfall, hourly rainfall and common calibration strategy were also compared. Figure 15 and Figure 16 show the distributions of the optimized model parameters for Rottweil and Pforzheim, respectively. Note that all parameters are normalized by the initial ranges in Table 2. Form the box plots we could find that some model parameters strongly depend on the selected rainfall dataset. This is very evident with the shape factor ($Beta$) and the threshold water level for surface runoff (L).

5 Conclusions and outlook

In this study, we investigated the impacts of temporal and spatial variability of rainfall in model simulation and parameter estimation. We also explored the question whether higher temporal and spatial resolutions of rainfall data lead to any improvement of model performance. Both the lumped and spatially distributed HBV model were applied to simulate daily runoff for four mesoscale catchments driven by four different rainfall datasets which were constructed using a combination of data from high

density of daily stations and relatively low density sub-daily stations. The impacts of rainfall variability on model simulations were evaluated using the NS coefficient and the mean squared error of flows higher than the 10th percentile of flow. The model sensitivities to the temporal and spatial resolutions of rainfall were compared. In addition, the common calibration approach was proposed to calibrate the models with different time steps simultaneously for finding robust model parameters.

5 For the study catchments, the results indicate that the temporal variability of rainfall data has direct impact on dynamic response of a catchment. For both lumped and spatially distributed models, if the observation density is the same, the hourly based simulation outperforms the daily based simulation, indicating that higher temporal resolution can significantly improve the model performance. Disaggregating high density daily observations into relatively low density sub-daily values could lead to considerable model improvement, especially for the catchment with a sparse rain gauge network. Rainfall disaggregation

10 approach is an effective way of increasing the temporal resolution of rainfall data and the model performance. However, the lumped and spatially distributed HBV model perform very similarly, indicating that higher spatial resolution does not or only marginally improve the model performance for the study catchments. The result agrees with the general findings of Lobligeois et al. (2014) and Zhu et al. (2018), where insignificant improvement was observed using higher spatial resolution of rainfall. The reason that the spatially distributed model does not outperform the lumped model could be due to the fact the study

15 catchments are smaller than 2000km^2 with relatively uniform rainfall.

As stated at the beginning of this paper, we aim to investigate the sensitivity of model to rainfall variability and to find effective ways for improving the model performance. This study shows that rainfall data disaggregation can lead to a significant improvement of model performance, while higher spatial resolution of rainfall does not always enhance model performance. Most of the hydrological models can be easily adjusted to use different time steps. The study suggests that increasing the

20 temporal resolution of rainfall inputs with disaggregation method can be an easier and more efficient way to improve model performance, compared to increasing the model spatial resolution at a cost of increasing the complexity of model structure and parameters.

This study focuses on high flows and uses only the NS coefficient as a quantitative measure of model sensitivity. As model performance highly depends on the selection of objective functions, the model sensitivity can be different if the model perfor-

25 mance is measured differently. In addition, all the hourly simulated runoff was aggregated into daily, the hydrological response was evaluated based on daily discharge. Sub-daily response of a catchment is more sensitive to the temporal and spatial variability of rainfall, which should be considered in the future if the hourly discharge observation is available.

Acknowledgements. This study was partially supported by the National Key Research and Development Program of China (2018YFC1508101, 2018YFC1508102, 2016YFC0402701), National Natural Science Foundation of China (51879067), Fundamental Research Funds for the

30 Central Universities of China (2018B42914), the China Postdoctoral Science Foundation (2017M621614), the Postdoctoral Research Supporting Program of Jiangsu province (2018K128C), and the Priority Academic Program Development of Jiangsu Higher Education Institutions. The authors gratefully acknowledge Bettina Schaefli and two anonymous reviewers for their invaluable and constructive suggestions. Special thanks to Yi He, Jingfeng Wang and Gebdang Biangbalbe Ruben for proofreading of the manuscript.

References

Ahmed, S. and De Marsily, G.: Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity, Water Resour. Res., 23, 1717–1737, https://doi.org/10.1029/wr023i009p01717, 1987.

Bardossy, A. and Das, T.: Influence of rainfall observation network on model calibration and application., Hydrol. Earth Syst. Sci., 12,1(2008-01-25), 12, 77–89, https://doi.org/10.5194/hess-12-77-2008, 2008.

Bardossy, A. and Pegram, G.: Combination of radar and daily precipitation data to estimate meaningful sub-daily point precipitation extremes, J. Hydrol., 544, 2016.

Bárdossy, A. and Pegram, G.: Space-time conditional disaggregation of precipitation at high resolution via simulation, Water Resour. Res., 52(2):920-937, https://doi.org/10.1002/2015wr018037, 2016.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters., Hydrol. Earth Syst. Sci., 12, 1273–1283, https://doi.org/10.5194/hess-12-1273-2008, 2008.

Bárdossy, A., Huang, Y., and Wagener, T.: Simultaneous calibration of hydrological models in geographical space, Hydrol. Earth Syst. Sci., 20, 2913–2928, https://doi.org/10.5194/hess-20-2913-2016, 2016.

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff model., Nord. Hydrol., 4, 174–190, 1973.

Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, J. Hydrol., 299, 166–179, 2004.

Beven, K.: Model predictions: Uncertainty, Encyclopedia of Earth Science, 486–489, 1998.

Booij, M.: Modelling the effects of spatial and temporal resolution of rainfall and basin model on extreme river discharge, International Association of Scientific Hydrology Bulletin, 47, 14, 2002.

Breinl, K. and Di Baldassarre, G.: Space-time disaggregation of precipitation and temperature across different climates and spatial scales, J. Hydrol.: Regional Studies, 21, 126–146, 2019.

Bruneau, P., Gascuel-Odoux, C., Robin, P., Merot, P., and Beven, K.: Sensitivity to space and time resolution of a hydrological model using digital elevation data, Hydrol. Process., 9, 69–81, 1995.

Das, T., Bárdossy, A., Zehe, E., and He, Y.: Comparison of conceptual model performance using different representations of spatial variability, J. Hydrol., 356, 106–118, 2008.

Emmanuel, I., Andrieu, H., Leblois, E., Janey, N., and Payrastre, O.: Influence of rainfall spatial variability on rainfall–runoff modelling: Benefit of a simulation approach?, J. Hydrol., 531, 337–348, 2015.

Emmanuel, I., Payrastre, O., Andrieu, H., and Zuber, F.: A method for assessing the influence of rainfall spatial variability on hydrograph modeling. First case study in the Cevennes Region, southern France, J. Hydrol., 555, 314–322, 2017.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228, 113–129, 2000.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, 2009.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Ang. Agric., 1, 96–99, 1985.

Hartmann, G. M.: Investigation of evapotranspiration concepts in hydrological modelling for climate change impact assessment, Ph.D. dissertation No. 161, University of Stuttgart, 2007.

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res.-Atmos., 113, 2008.

Hofierka, J., Parajka, J., Mitasova, H., and Mitas, L.: Multivariate interpolation of precipitation using regularized spline with tension, T. GIS, 6, 135–150, 2002.

Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, J. Hydrol., 292, 281–295, 2004.

5   Jeffrey, S. J., Carter, J. O., Moodie, K. B., and Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environ. Modell. Softw., 16, 309–330, 2001.

Kobold, M. and Brilly, M.: The use of HBV model for flash flood forecasting, Nat. Hazard. Earth Sys., 6, 407–417, 2006.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrol. Earth Syst. Sci., 18, 575–594, https://doi.org/10.5194/hess-18-
10   575-2014, 2014.

Ly, S., Charles, C., and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review, Biotechnol. Agron. Soc., 17, 392, 2013.

Matheron, G.: Principles of geostatistics, Econ. Geol., 58, 1246–1266, 1963.

Meselhe, E. A., Habib, E. H., Oche, O. C., and Gautam, S.: Sensitivity of Conceptual and Physically Based Hydrologic Models to Temporal
15   and Spatial Rainfall Sampling, J. Hydrol. Eng., 14, 711–720, 2009.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models. 1. A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Obled, C., Wendling, J., and Beven, K.: The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data, J. Hydrol., 159, 305–333, 1994.

Parkes, B. L., Wetterhall, F., Pappenberger, F., He, Y., Malamud, B. D., and Cloke, H. L.: Assessment of a 1-hour gridded precipitation
20   dataset to drive a hydrological model: A case study of the summer 2007 floods in the Upper Severn, UK, Hydrol. Res., 44, 89–105, 2013.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193, 120–145, 1948.

Pui, A., Sharma, A., Mehrotra, R., Sivakumar, B., and Jeremiah, E.: A comparison of alternatives for daily to sub-daily rainfall disaggregation, J. Hydrol., 470, 138–157, 2012.

25   Rango, A. and Martinec, J.: Revisiting the degree-day method for snowmelt computations 1, J. Am. Water Resour. As., 31, 657–669, 1995.

Singh, S. K.: Robust parameter estimation in gauged and ungauged basins, Ph.D. dissertation No. 198, University of Stuttgart, 2010.

Singh, V.: Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, Hydrol. Process., 11, 1649–1669, 1997.

Xu, H., Xu, C. Y., Chen, H., Zhang, Z., and Li, L.: Assessing the influence of rain gauge density and distribution on hydrological model
30   performance in a humid region of China, J. Hydrol., 505, 1–12, 2013.

Zeleny, M.: Multiple Criteria Decision Making., McGraw-Hill, New York, USA, 1981.

Zhu, Z., Wright, D. B., and Yu, G.: The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis, Water Resour. Res., 54, 8983–8998, https://doi.org/10.1029/2018WR023550, 2018.

Table 1. Catchment characteristics for the 4 selected catchments.

No.	Stream gauge name	Longitude (^{o}E)	Latitude (^{o}N)	Area (km^2)	Elevation (m)	Annual rainfall (mm)	Average temperature (^{o}C)	Annual runoff (mm)
1	Rottweil, Neck	8.38	48.10	455	555-1010	929.0	9.7	363.2
2	Schwaibach,Kinzig	8.02	48.24	955	190-1028	1331.8	9.7	757.3
3	Pforzheim,Würm	8.43	48.52	417	357-583	761.7	9.3	232.9
4	Kocherstetten, Kocher	9.45	49.16	1288	292-698	930.6	9.4	401.6

Table 2. Description of HBV model parameters and parameter ranges for model calibration.

Parameter	Description	Max	Min
TT	Threshold temperature for snow melt initiation (^{0}C)	2	-2
DD	Degree-day factor	3	1.5
FC	Field capacity (mm)	600	50
Beta	Shape coefficient	8	0.2
HL	Threshold water level for near surface flow (mm)	100	1
K_0	Near surface flow storage constant	0.8	0.2
K_1	Interflow storage constant	0.25	0.1
K_d	Percolation storage constant	0.2	0.05
K_2	Baseflow storage constant	0.1	0.01

Table 3. Average NS model performance for the lumped HBV model.

Catchment	Rainfall dataset	Calibration for 1991-2000	Calibration for 2001-2010	Validation for 1991-2000	Validation for 2001-2010
Rottweil	SH	0.71	0.71	0.65	0.65
	DH	0.79	0.73	0.73	0.68
	SD	0.61	0.61	0.56	0.55
	DD	0.67	0.63	0.63	0.59
Schwaibach	SH	0.60	0.88	0.52	0.72
	DH	0.89	0.88	0.88	0.87
	SD	0.57	0.85	0.49	0.68
	DD	0.84	0.86	0.83	0.83
Pforzheim	SH	0.61	0.69	0.60	0.65
	DH	0.63	0.69	0.63	0.67
	SD	0.48	0.60	0.46	0.56
	DD	0.48	0.60	0.49	0.57
Kocherstetten	SH	0.88	0.85	0.86	0.84
	DH	0.89	0.85	0.87	0.84
	SD	0.84	0.84	0.81	0.79
	DD	0.84	0.83	0.81	0.81

Table 4. Average NS model performance for the distributed HBV model.

Catchment	Rainfall dataset	Calibration for 1991-2000	Calibration for 2001-2010	Validation for 1991-2000	Validation for 2001-2010
Rottweil	SH	0.70	0.68	0.63	0.55
	DH	0.80	0.69	0.74	0.66
	SD	0.61	0.59	0.54	0.46
	DD	0.68	0.60	0.63	0.57
Schwaibach	SH	0.59	0.88	0.50	0.76
	DH	0.90	0.88	0.88	0.87
	SD	0.55	0.86	0.47	0.72
	DD	0.85	0.86	0.84	0.85
Pforzheim	SH	0.55	0.68	0.55	0.64
	DH	0.59	0.67	0.59	0.64
	SD	0.42	0.58	0.41	0.54
	DD	0.45	0.58	0.46	0.54
Kocherstetten	SH	0.88	0.86	0.86	0.84
	DH	0.89	0.86	0.87	0.84
	SD	0.84	0.84	0.82	0.80
	DD	0.84	0.84	0.82	0.81

Table 5. Mean square error for flows higher than the 10th percentile of flow for the lumped HBV model.

Catchment	Rainfall dataset	Calibration for 1991-2000	Calibration for 2001-2010	Validation for 1991-2000	Validation for 2001-2010
Rottweil	SH	83.1	74.6	118.7	83.5
	DH	55.1	69.8	82.4	84.7
	SD	120.0	104.5	151.4	108.5
	DD	101.7	98.9	120.0	110.1
Schwaibach	SH	2511.4	338.6	3214.9	663.6
	DH	565.4	324.4	722.7	328.2
	SD	2739.9	401.1	3423.0	805.7
	DD	916.0	389.2	1048.1	448.2
Pforzheim	SH	11.8	7.3	12.4	8.3
	DH	11.2	6.9	11.8	7.3
	SD	19.1	10.6	19.6	12.0
	DD	18.9	10.3	19.5	10.9
Kocherstetten	SH	438.9	457.5	545.5	558.7
	DH	288.5	439.3	350.5	518.8
	SD	651.9	551.9	801.9	760.4
	DD	556.0	544.1	665.0	701.3

Table 6. Mean square error for flows higher than the 10th percentile of flow for the distributed HBV model.

Catchment	Rainfall dataset	Calibration for 1991-2000	Calibration for 2001-2010	Validation for 1991-2000	Validation for 2001-2010
Rottweil	SH	89.0	86.8	127.8	120.1
	DH	56.5	85.2	80.1	95.0
	SD	121.0	113.6	161.4	144.5
	DD	100.6	111.5	119.6	121.9
Schwaibach	SH	2657.1	326.9	3330.8	527.1
	DH	526.1	311.4	680.7	317.7
	SD	2869.6	387.9	3546.7	681.5
	DD	892.8	376.5	983.2	405.9
Pforzheim	SH	12.5	7.1	12.7	8.1
	DH	11.9	6.7	12.4	7.2
	SD	19.6	10.3	19.7	11.5
	DD	19.5	9.9	19.6	10.6
Kocherstetten	SH	425.7	455.1	541.2	551.5
	DH	293.5	429.1	355.3	515.1
	SD	633.3	552.0	778.6	727.3
	DD	542.4	540.8	637.0	670.9
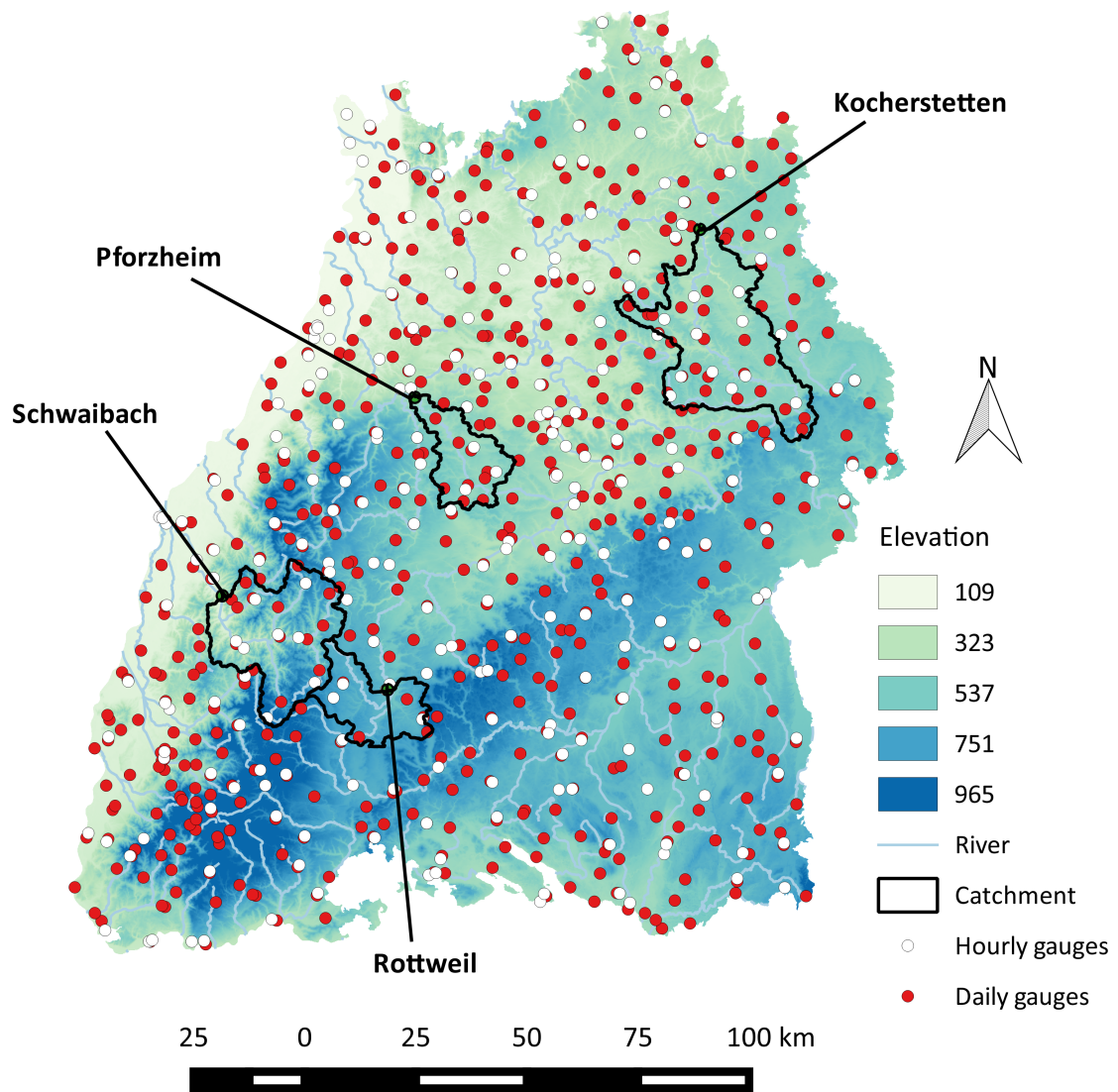
Figure 1. Locations of the sub-daily and daily rain gauges in Baden-Württemberg and the four selected catchments.
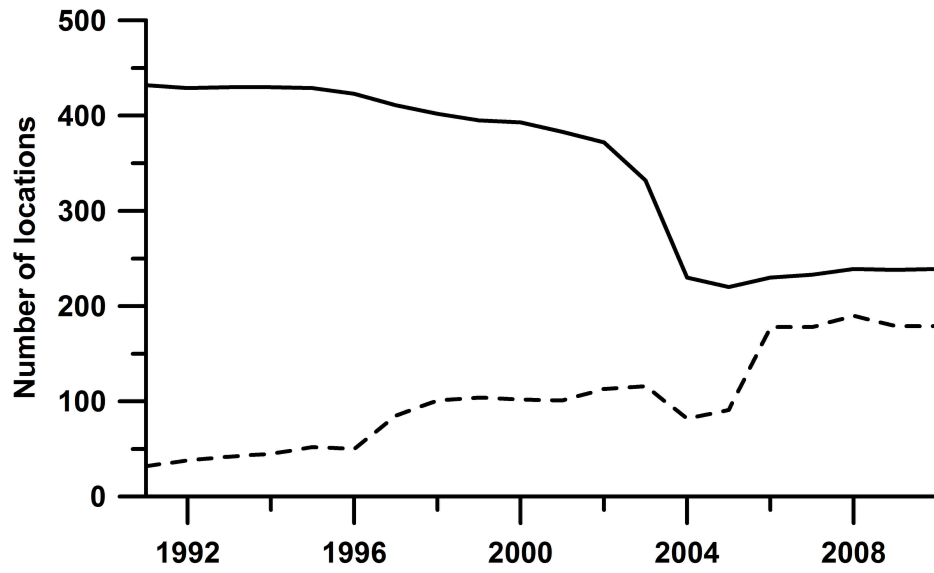
Figure 2. The number of available observation locations. Daily stations - solid line, Sub-daily stations - dashed line.
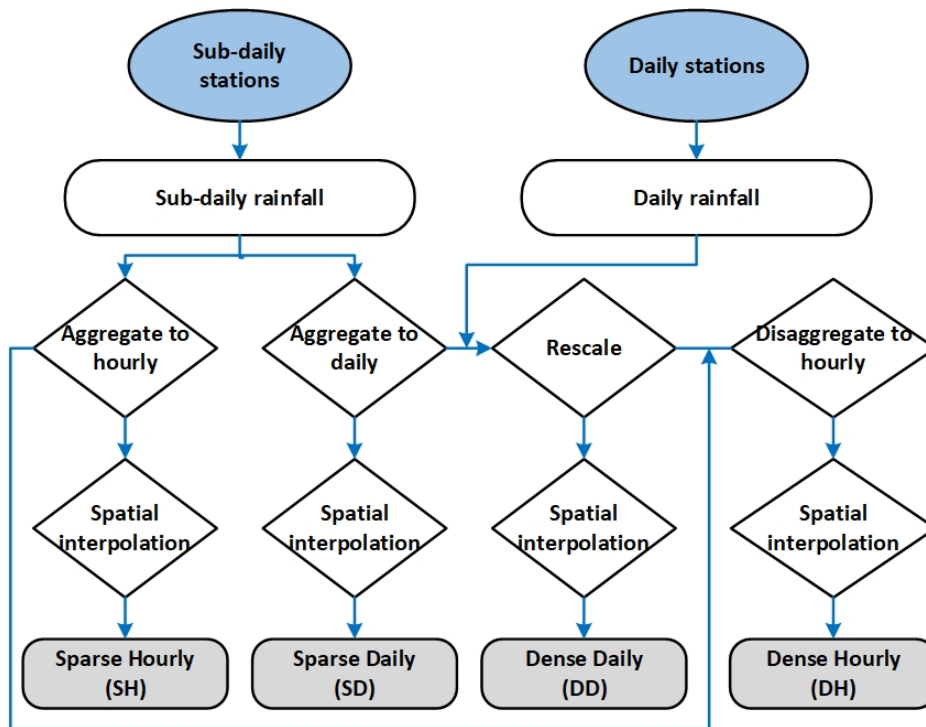
Figure 3. Schematic representation of four different rainfall datasets.
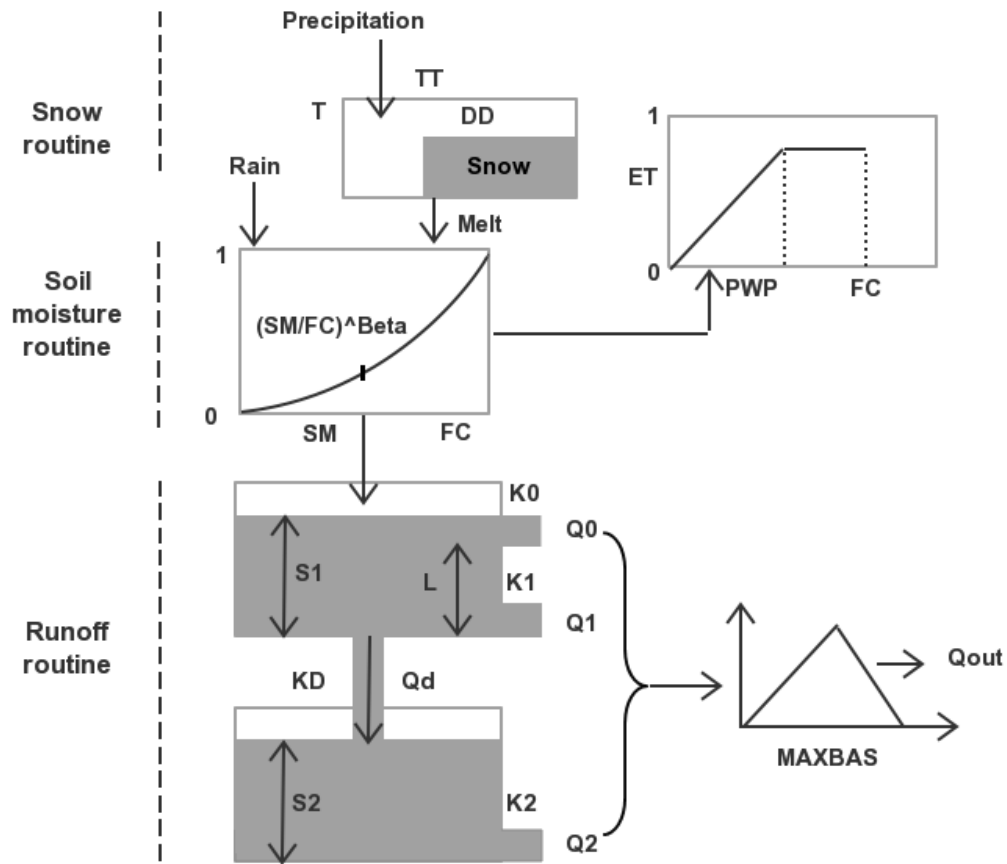
Figure 4. Schematic representation of HBV model.
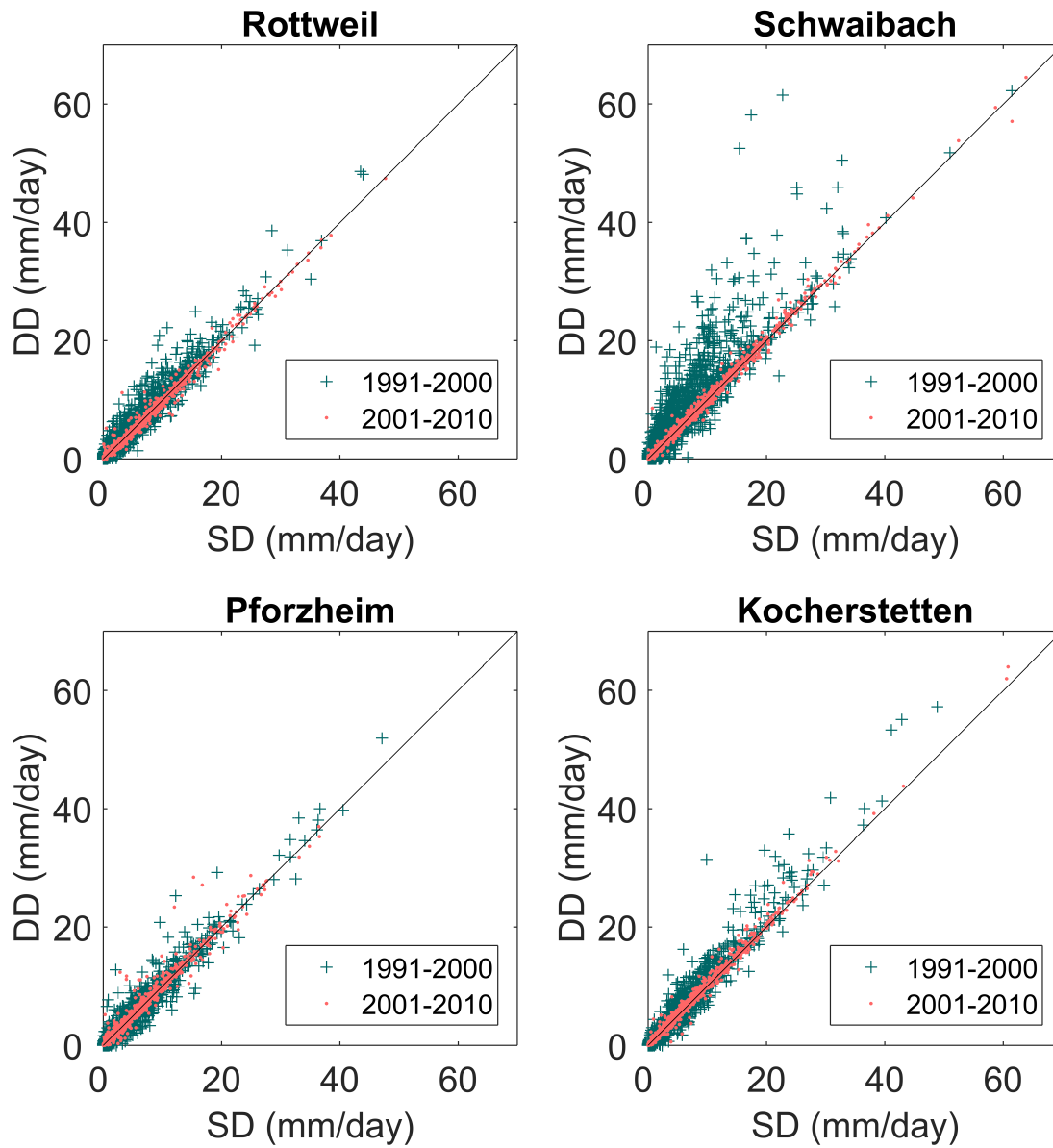
Figure 5. Comparison of the daily rainfall data that interpolated using different densities of rain gauge network.
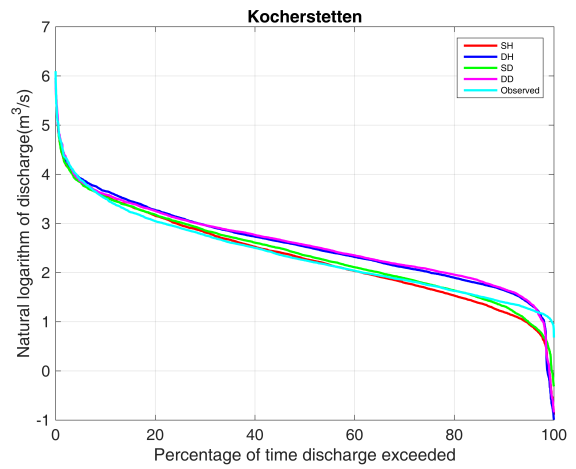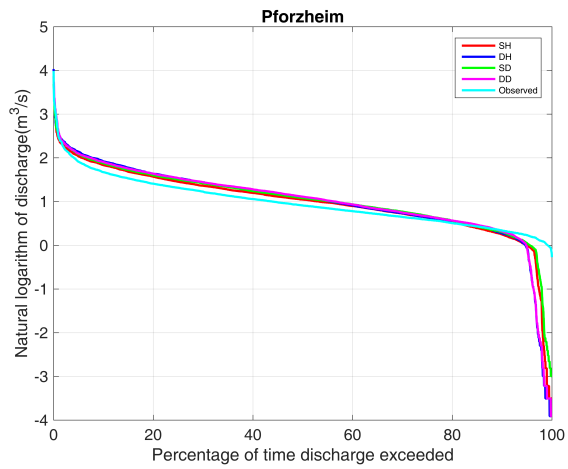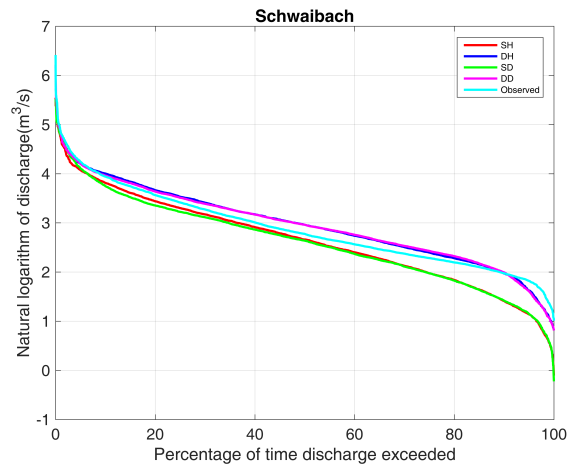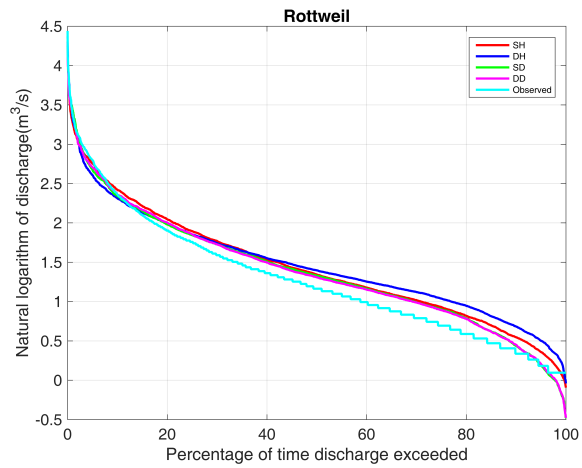
Figure 6. Comparison of the simulated flow duration curve.
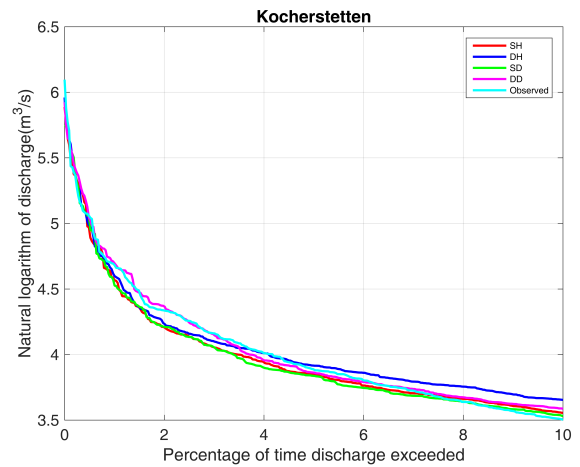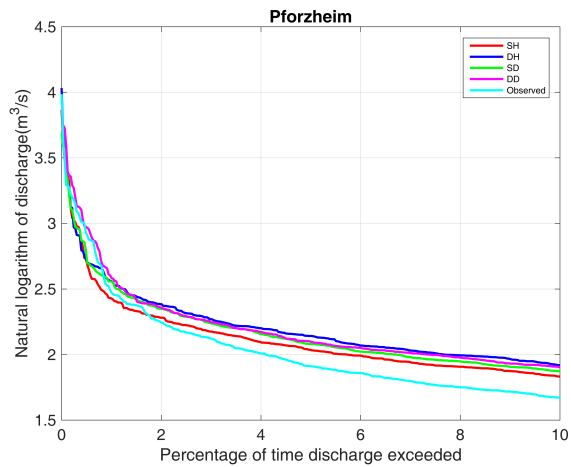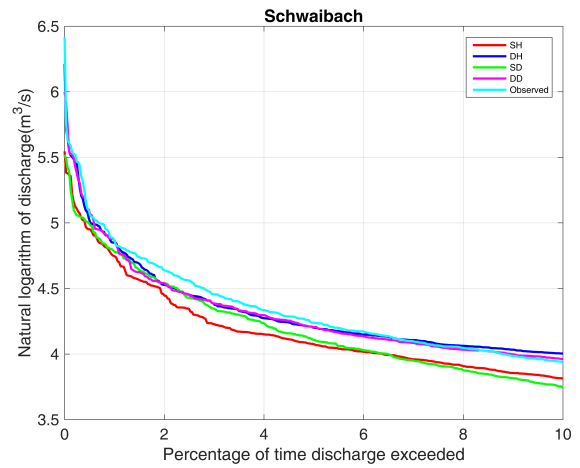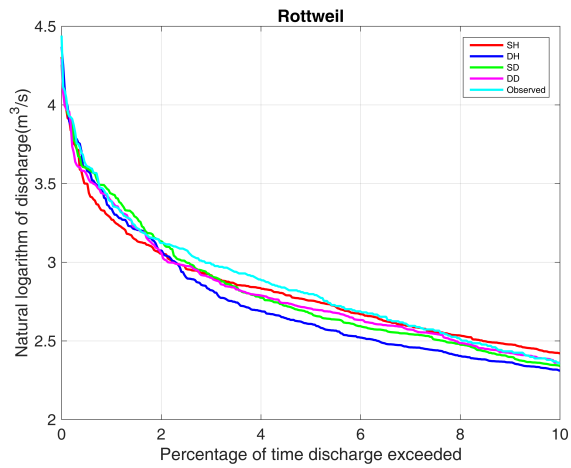
Figure 7. Comparison of the simulated flow duration curve for flows higher than the 10th percentile of flow.
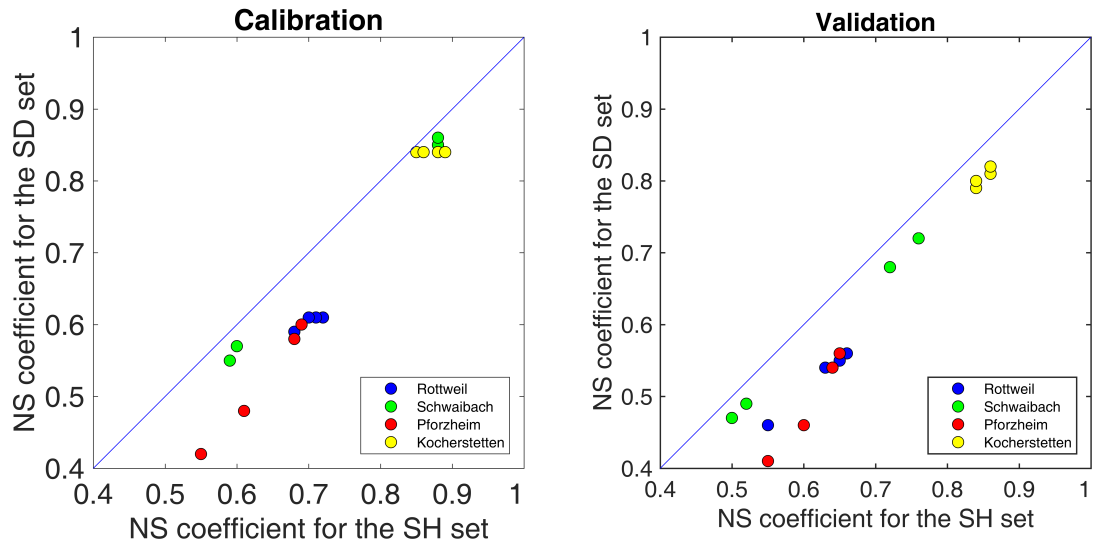
Figure 8. Comparison of NS coefficient for using hourly and daily rainfall as model input for the SH and SD sets.
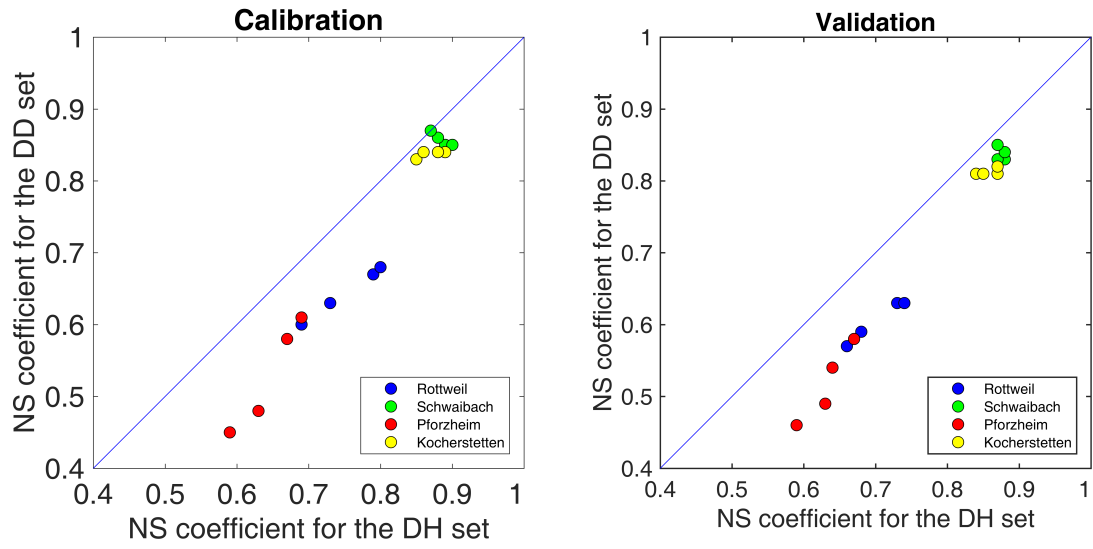
Figure 9. Comparison of NS coefficient for using hourly and daily rainfall as model input for the DH and DD sets.
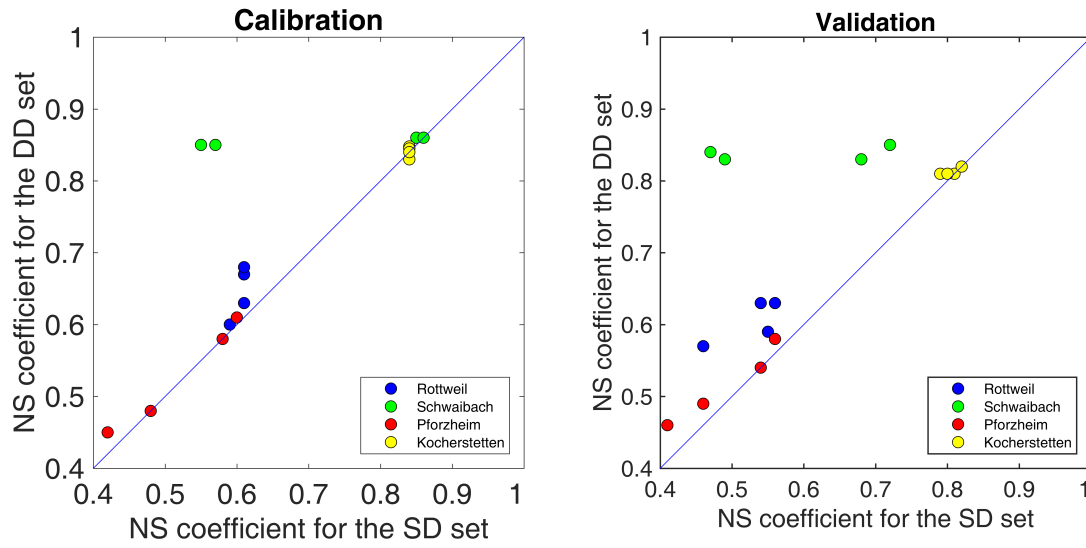
Figure 10. Comparison of NS coefficient for different densities of rain gauge network, models were simulated based on daily time step.
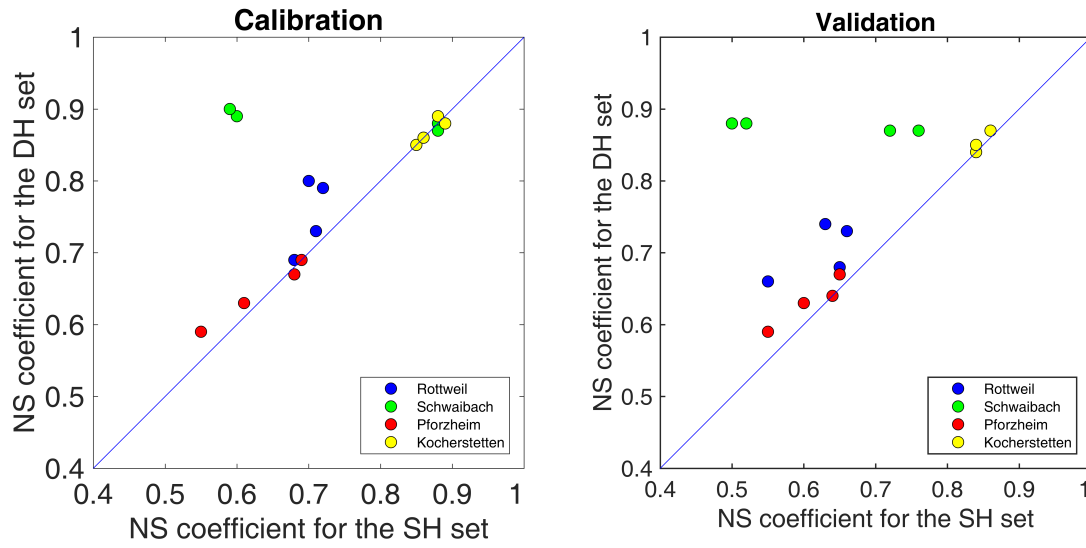
Figure 11. Comparison of NS coefficient for different densities of rain gauge network, models were simulated based on hourly time step.
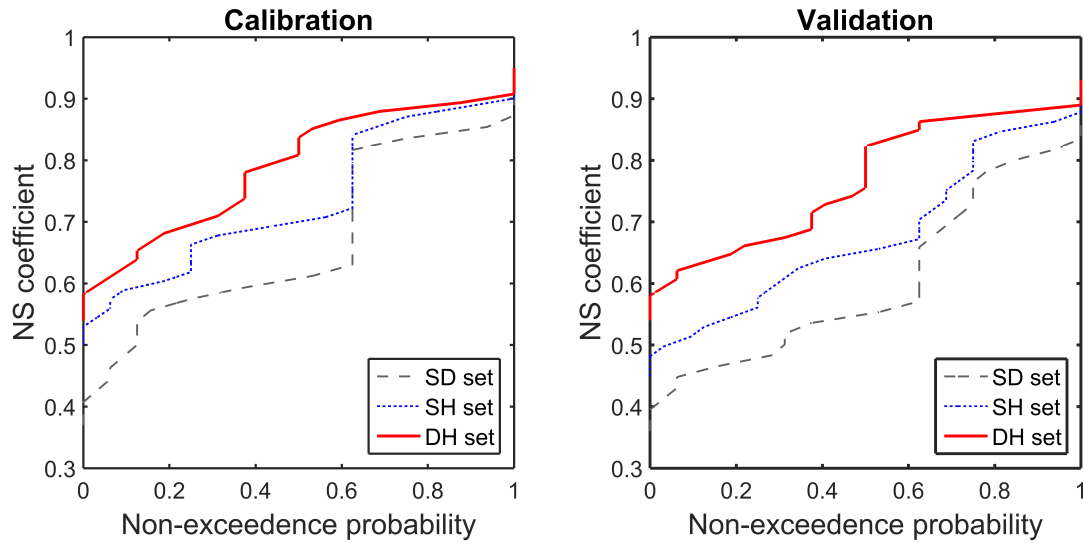
Figure 12. Cumulative distribution of NS coefficient for model calibration using different rainfall datasets .
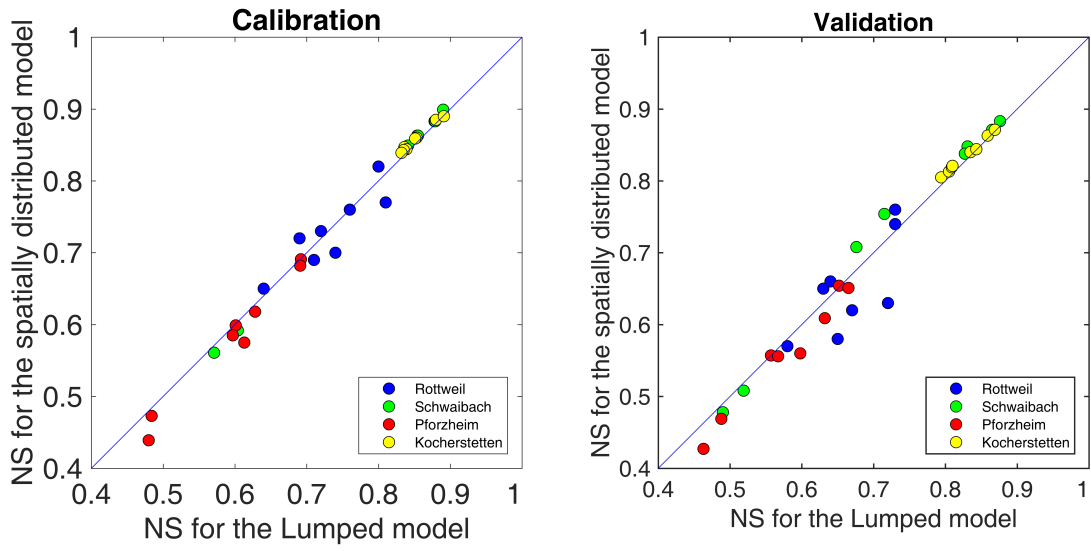
Figure 13. Comparison of NS coefficient for different spatial resolution of model structure.
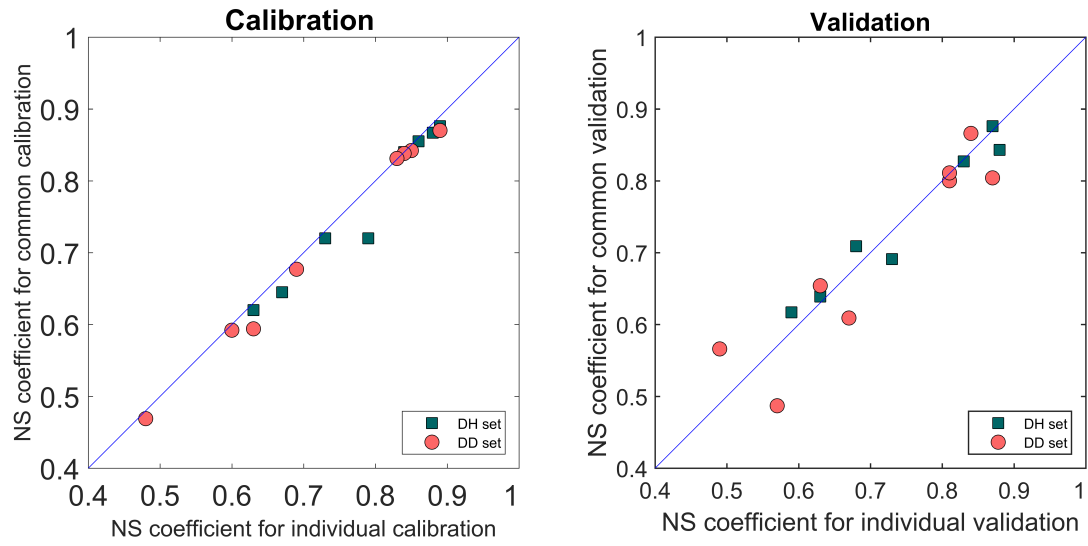
Figure 14. Comparison of NS coefficient for individual calibration and common calibration using datasets with different temporal resolutions.
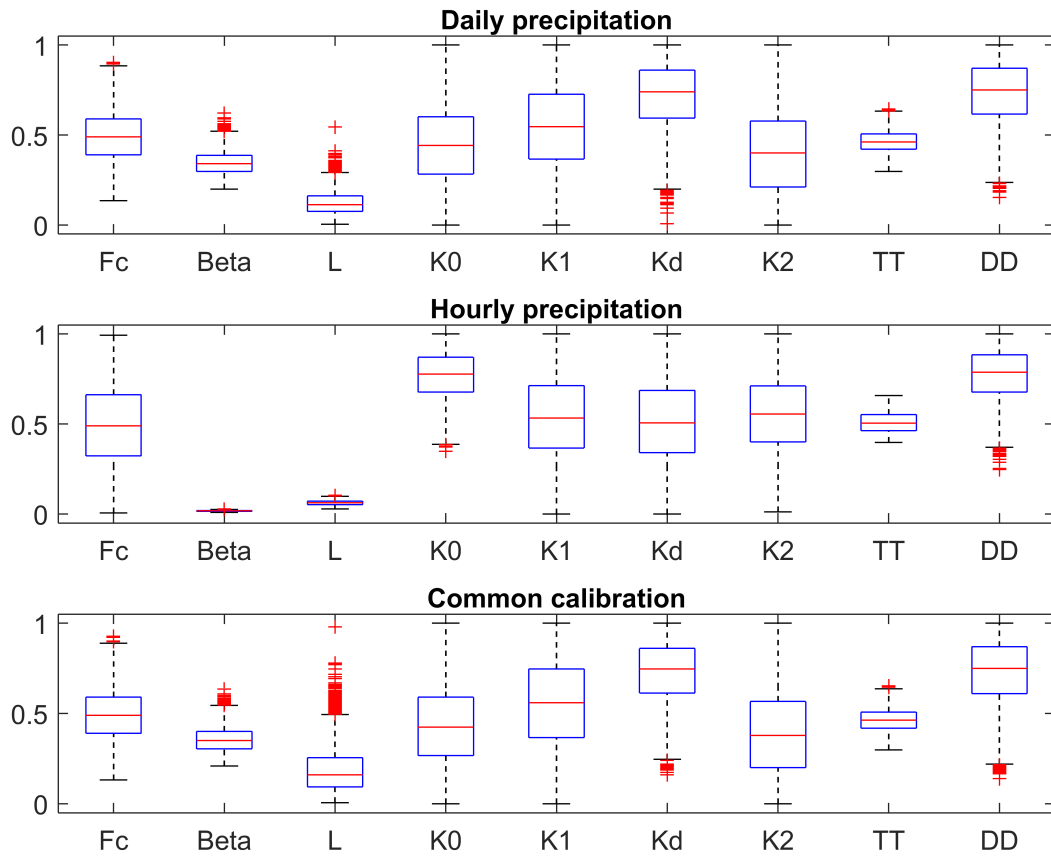
Figure 15. Comparison of model parameters for different temporal resolutions for Rottweil catchment.
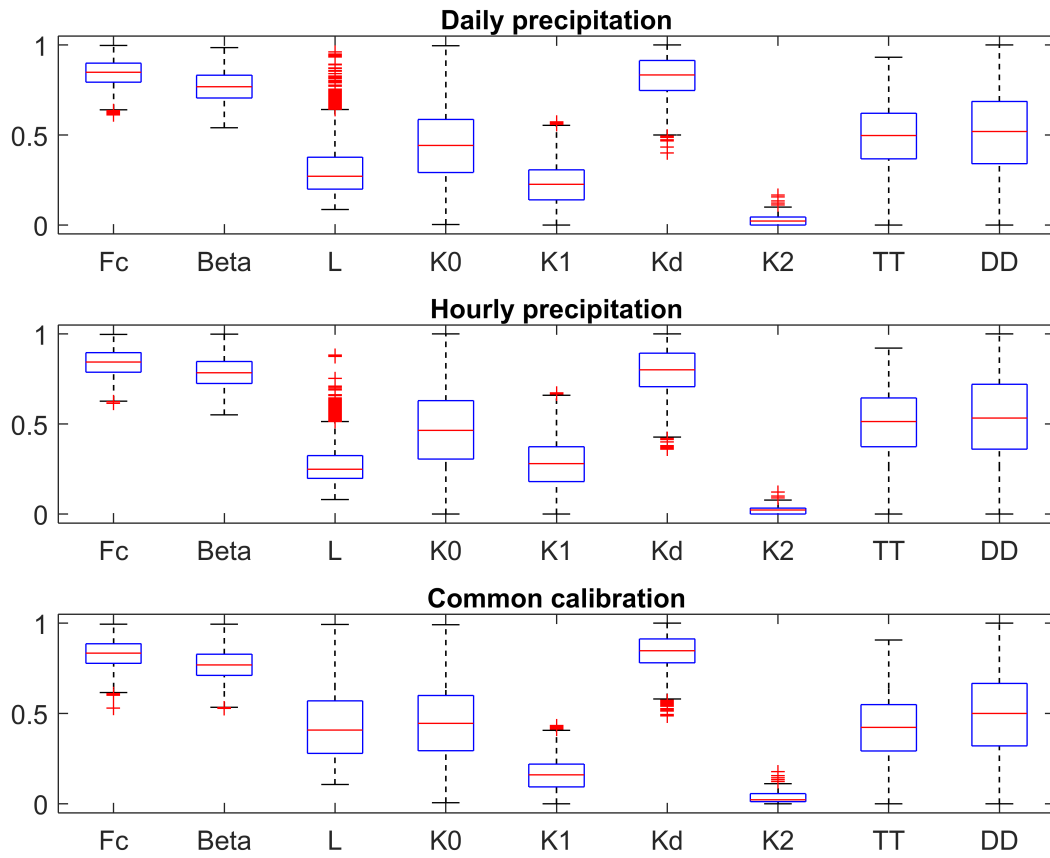
Figure 16. Comparison of model parameters for different temporal resolutions for Pforzheim catchment.