Point-by-point response to Reviewers

Dear editor and reviewers,

We would like to thank all of you for the review of our manuscript and the constructive suggestions. All the comments have been considered and a point by point response has been provided below.

For Referee 2's additional question about the structure of the manuscript, we provided a further explanation on why we prefer to keep the subsection describing the impact of spatial resolution of rainfall on model performance. The manuscript has been thoroughly revised and polished carefully with the reviewers' help.

The point-by point response is formatted as follows:

- the referees' comments are shown in black

- authors' response are shown in blue

----

1. Response to Anonymous Referee #1

The authors have satisfactorily addressed my reviewer comments and have made a number of revisions which have improved the manuscript. I recommend that the manuscript is accepted with a few minor revisions detailed below.

1. The introduction is much improved with a wider range of references but the English needs to be carefully checked as there are a couple of sentences that don't make sense as currently written. Specifically:

a. P3 L1-3 "As most of the hydrological models are flexible and can be easily adjusted to different time steps, which makes the sensitivity analysis of model output to the temporal variability of rainfall easy."

b. P3 L13-15 "This could properly lead to a better understanding of the sensitive of rainfall inputs and help to identify relatively economical ways to improve tremendously the model behavior."

Response: We sincerely thank you for the valuable comments. We have rewritten the above sentences to make them easier to be understood. The revised manuscript has been sent to two professionals for proofreading. According to their suggestions, we have thoroughly corrected the grammar and improved the clarity of the sentences. All the corrections are marked in the revised version.

2. I liked the addition of Figure 7 but I think the x-axis is wrong. If you have plotted simply the top 10th percentile of flows then surely the x-axis should go from 0 to 10 rather than 0 to 100?

Response: Thank you for point it out. Yes, the x-axis should go from 0 and 10. We have added the corrected flow duration curve for flows higher than the 10th percentile of flow in the revised manuscript.

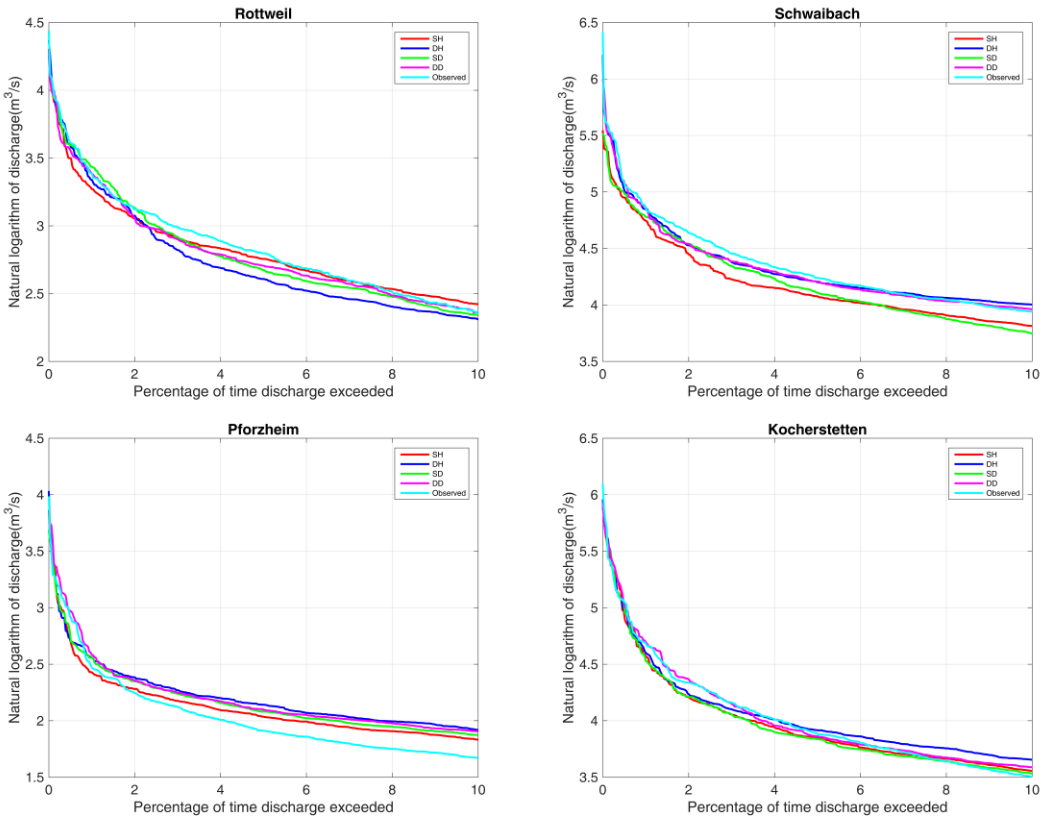

Figure 7. Comparison of the flow duration curve for flows higher than the 10th percentile of flow.

3. Figures 8 - 11 - although the authors have made changes to these plots, I still find it difficult to distinguish between the colours (particularly the red and pink dots). It would be worth changing the pink dot to a green dot so it is easier to identify the catchments.

Response: Thanks for the suggestion. We have changed the data point color schemes in Figure 8-11 (change the pink and cyan dots to green and yellow, respectively). As shown in Figure 8, it can be easier to identify the catchments than the original figure.

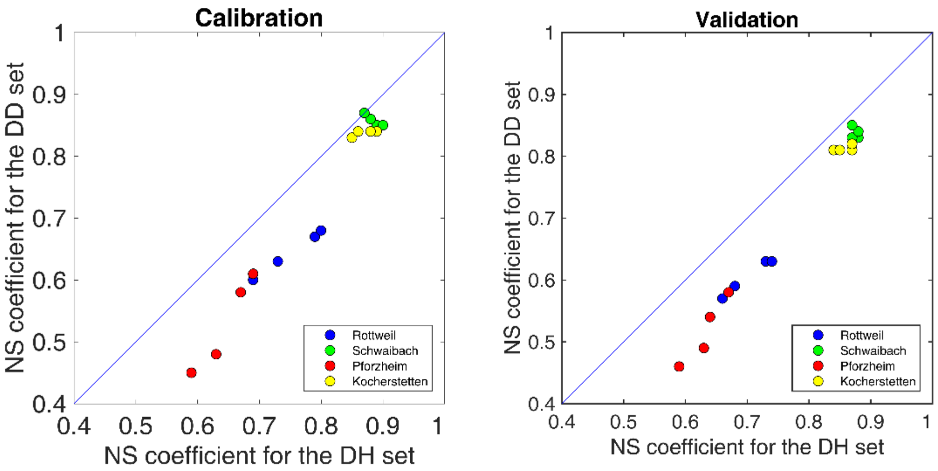

Figure 8. Comparison of NS model performance for using hourly and daily rainfall as model input for the DH and DD sets.

---

2. Response to Anonymous Referee #2

The literature review was extended, that is a good point. However it now reads like a very long list and the authors should help the readers by adding a few summarizing sentences, to underline that the literature does not have a “ready” answer for the question they ask and even that the papers they review do not agree with each other (which is a further reason for writing this paper).

Response: We sincerely appreciate you for reviewing our manuscript. Your valuable comments and suggestions led to an improved version of the manuscript. We have reorganized the introduction part, and the literature review is expanded with a summary of previous studies.

As far a the restructuration/simplification of the paper I had suggested the authors did not do it. I understand that this is a lot of work. However, in my opinion, it would have made the paper simpler and easier to understand. I believe the paper still reads more like an exhaustive report than as a selection of the most interesting results.

Response: We thank the reviewer for your suggestion on simplifying the paper structure as mentioned in your previous comments: “I suggest removing the part on the different rainfall densities, and only keeping the densest network (high density daily disaggregated into hourly)”. We think that the imperfection of the sub-section titles in the earlier version hinder the reviewer from better understanding the logic flow of our result section. Therefore, we revised the sub-section titles of the result section in the revised manuscript. The revised sub-sections are:

“4.1 Comparison of the rainfall dataset”, “4.2 Results of calibration and validation”, “4.3 Model performance using different temporal resolutions of rainfall data ”, “4.4 Model performance in terms of observation density”, “4.5 Model performance in terms of spatial resolution of rainfall data”, and “4.6 Common model calibration with different temporal resolutions”. As you may tell, the revised sub-section titles are easy to follow and each sub-section are closely related to the objectives and unique to each other. In addition, six sub-sections in a result section are not too many. We therefore believe the current structure works fine.

In addition, we would like to answer in more detail for the questions you raised during your first review of the manuscript:

(1) The purpose of this study is to find the effective ways for improving model performance for flood forecasting. It requires understanding the sensitivity of the rainfall-runoff modes to rainfall input data. The spatial variability of rainfall strongly influences the timing and shape of hydrograph, while the temporal variability mainly affects the peak of flood wave. As increasing the temporal and spatial resolutions of model are two common methods in hydrological modeling, we believe that the comparison of temporal and spatial variability of rainfall to runoff simulation is very important. With the testing of different spatial resolutions, we hope to answer the specific question that which one is more efficient to improve model performance: Increasing temporal resolution or spatial resolution of rainfall? We can conclude from this study that higher temporal resolution of rainfall can lead to a significant improvement of model performance, while higher spatial resolution of rainfall does not always enhance model performance. It suggests that compared with increasing the model spatial resolution that comes at a cost of increasing the complexity of model structure and parameters, increasing the temporal resolution of precipitation inputs with disaggregation method can be easier and more efficient to improve model performance. We think it is worth preserving the results based on different spatial resolutions of rainfall in the manuscript.

(2) This study aims to increasing the accuracy of flood prediction and thus pays more attention on the high flow. The HBV model performance to different performance criteria has been investigated by our previous study. Result shows that the model sensitivity can be different is the model performance is measured differently. Result also shows that for most of the cases, the model performance for different objective functions have same tendency of changes under different catchments. In this study, each calibration process requires 90000 running of HBV model to obtain 10000 best parameter sets. Due to the heavy computation, we only tested the sensitively of model performance based on the most widely used performance criterion- NS coefficient. NS coefficient represents the squared difference between the observed and discharge

series and mainly focuses on the high flow, which could satisfy the needs of improving the accuracy of flood prediction.

(3) The HBV model used in this study is relatively simple. There is no interception routine in this version of HBV model. Currently, the process of interception is simulated implicitly in the evapotranspiration part of the model and the comparison of different temporal resolutions is based on the simulation of daily runoff. In our further study, the interception routine will be included to investigate the impact in hourly simulation.

Last, I still found a number of typos (an example the authors write “in additional” instead of “in addition” in the conclusion).

Response: We have carefully proofread the manuscript and corrected the typos and grammar errors in the revised manuscript.

Sensitivity of hydrological model to ~~the~~temporal and spatial resolutions of rainfall ~~input~~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. ~~As~~Rainfall is the most important input for rainfall-runoff models~~,~~. ~~precipitation~~It is usually ~~observed~~measured at specific sites on a daily or sub-daily time scale and requires interpolation for further application. This study aims to ~~explore that for a given objective function, whether~~evaluate if a higher temporal and spatial resolution of ~~precipitation could provide an improvement in~~rainfall can lead to improved model performance. Four different gridded hourly and daily ~~precipitation~~rainfall datasets ~~,~~with a spatial resolution of 1km×1km for the state of Baden-Württemberg ~~state of~~in Germany were constructed using a combination of data from a dense network of daily rainfall stations and a less dense network of sub-daily stations. ~~Two different types of HBV models with different model structures,~~A lumped and a spatially distributed HBV models were used to investigate the sensitivity of model performance ~~on~~to the spatial ~~variability of precipitation.~~resolution of rainfall. ~~For four selected mesoscale catchments, these~~The four different rainfall datasets were used to ~~simulate the daily discharges using both lumped and semi-distributed HBV models.~~drive both lumped and distributed HBV models to simulate daily discharges in four catchments. ~~Different possibilities of improving the accuracy of daily streamflow prediction were investigated.Three main results were obtained from this study:~~Main findings include (1) a higher temporal resolution of ~~precipitation improved~~rainfall improves the model performance if the ~~observation~~station density ~~was~~is high; (2) a combination of observed high temporal-resolution observations with ~~disaggregate~~disaggregated daily ~~precipitation~~rainfall leads to ~~a~~further improvement ~~in the model performance~~of the tested models; (3) for the present research, the increase of spatial resolution ~~improved~~improves the performance of the model insubstantially or only marginally ~~for~~in most of the study catchments.

1 Introduction

Rainfall is ~~one of the most important driving forces in hydrological modeling and produces a direct impact on~~a primary driver of hydrological models and can impact catchment runoff response significantly (Obled et al., 1994; Ly et al., 2013). ~~In general, r~~Rainfall is usually measured by standard rain gauges or wireless telemetering pluviometers over a period of time (e.g. daily, sub-daily). ~~The~~Uncertainties in rainfall estimation for a catchment can occur due to instrument ~~measuring error~~errors as well as ~~and the representativeness of point rainfall causes a certain amount of uncertainty in precipitation estimation for a specific catchment. The~~spatial and temporal variability of ~~precipitation~~rainfall. ~~is one of~~The latter are the main sources of ~~uncertainty~~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). ~~Therefore, it is of great significance to investigate the sensitivity of hydrological models to rainfall input and find an effective way to improve the accuracy of model simulation and flood forecasting.~~The improvement of flood simulation requires understanding the sensitivity of the rainfall-runoff models to rainfall input data. ~~In recent~~Over the past decades, ~~extensive efforts have been put on investigating the influence of rainfall spatial variability in hydrological models.Different interpolation~~various methods have been used to obtain the spatial ~~distribution structure~~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). ~~These approaches can potentially improve the spatial resolution of rainfall that is used as input for rainfall-runoff models, thereby reducing the uncertainty of hydrological models.Singh (1997) found that the spatial variability of rainfall can have significant influence on the timing and shape of hydrograph, while the temporal variability can affect the peak of flood wave.~~Kobold and Brilly (2006) ~~used a different number of rain gauge stations to derive areal rainfall and quantified uncertainties of rainfall inputs using HBV model in hourly time step.~~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 ~~precipitation~~rainfall may lead to even greater error in ~~the peak of flood~~flood peak. Bardossy and Das (2008) also ~~invested~~studied the impact of spatial variability of rainfall by varying the distribution of ~~the~~rain gauge network. They found that the ~~transferability~~transferabilities of model parameters calibrated based on ~~spare~~sparse and ~~density~~dense rainfall information ~~is~~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~~, and they naturally concluded~~. 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 ~~gradually improves~~can improve the model performance~~up to some threshold~~, but no apparent improvement was observed when the number of rain gauges exceeded ~~the~~a threshold. Lobligeois et al. (2014) ~~investigated the impacts of rainfall spatial variability by implementing diverse representations of model for a considerable number of catchments. They typically found that for the region with variable precipitation, the semi-distributed models outperform the lumped one, but these two models perform similar for the catchments that having relatively uniform precipitation.~~ 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 ~~carefully evaluate~~characterize the influence of ~~rainfall~~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, ~~the~~ increasing ~~of~~spatial resolution in model simulation ~~leads~~can lead to considerable complexity of model structure and ~~requires for~~require much more data than using a lumped version.

~~Simultaneously, the~~The rainfall-runoff response of a catchment is also strongly impacted by the temporal variability of rainfall (Bárdossy and Pegram, 2016). ~~The high~~High temporal resolution rainfall data ~~is typically measured by~~are collected at pluviometer stations ~~(wireless instruments recording at sub-daily intervals, be called sub-daily data in the following),~~with telemetry at sub-daily time resolutions. ~~which faces the problem of poor data~~Sub-daily data often have poor quality caused by equipment malfunction or misreading. Compared with sub-daily rainfall data, ~~the daily rainfall data~~daily data ~~are more reliable and plentiful~~tend to be more available and reliable, cover a longer duration of time periods. Disaggregating daily into sub-daily ~~values~~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) ~~properly~~compared three different approaches for disaggregating daily rainfall into sub-daily series and ~~indicated~~found the resampling method is the best ~~way~~one for rainfall ~~disaggregating~~disaggregation. Bárdossy and Pegram (2016) used Gaussian Copula-based model for disaggregating daily data to infill the gap of ~~pluviometer~~sub-daily data, and they found that this conditional disaggregation of ~~precipitation~~rainfall is reliable and applicable in various regions. Breinl and Di Baldassarre (2019) applied a spatial method of fragments to disaggregate daily ~~precipitation~~rainfall into hourly values. ~~Although considerable studies have been carried out in the interpolation of sub-daily rainfall, thoroughly verification of the data quality of these products through the comparison of rainfall-runoff simulation results is required. It is extreme important to find out if the disaggregation leads to an improvement of model performance. As most of the hydrological models are flexible and can be easily adjusted to different time steps, which makes the sensitivity analysis of model output to the temporal variability of rainfall easy.~~Kobold and Brilly (2006) found that calibrating hydrological models with sub-daily time steps can significantly improve the accuracy of flood forecasting.

~~Furthermore, certain~~Some studies focus on both the spatial and temporal resolution of rainfall. Bruneau et al. (1995) ~~indicated~~found that the temporal and spatial resolutions of rainfall used ~~for~~as the inputs of ~~the~~hydrological ~~model possess a~~models can have considerable influence on the model efficiency and ~~parameters~~parameter values. Booij (2002) ~~assuredly~~found ~~that~~the influence of ~~model~~rainfall spatial resolution is ~~indeed~~greater than ~~rainfall~~temporal ~~variability on the~~resolution in terms of simulation of extreme ~~flow~~flows. Meselhe et al. (2009) ~~indicated~~pointed out that ~~the~~ physically based ~~model is~~models are more sensitive to the spatial and temporal resolution of rainfall data than ~~the~~conceptual ~~model~~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~~; while flood~~, and floods in ~~the~~small catchments ~~is controlled~~are more influenced by the temporal variability of rainfall. ~~Since a vast number of efforts had been made to improve~~So far, more efforts have been invested in improving the spatial or temporal resolution of rainfall, ~~it is important to focus on a quantitative analysis and~~but there are less studies on quantification and direct comparison of the ~~potential effect of rainfall temporal variability with the spatial variability to catchment dynamic response. This could properly lead to a better understanding of the sensitive of rainfall inputs and help to identify relatively economical ways to improve tremendously the model behavior.~~catchment dynamic responses driven by different rainfall temporal and spatial resolutions.

The ~~ultimate~~overarching aim of this study is to ~~undoubtedly gain more firsthand knowledge on~~understand the dependency of hydrological model performance on the ~~precipitation~~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 effects of rainfall data quality on model performance were investigated. The sensitivity of model performance to different spatial and temporal resolutions of rainfall data was examined using two distinctive model structures. The possibility of improving model performance on a daily scale was properly discussed.~~The ~~manuscript is organized as follows: the introduction,~~paper will be followed by section 2 ~~, which describes~~to describe the study area and the ~~precipitation~~rainfall datasets used in this research. In section 3, the hydrological model and the calibration ~~framework used in this research~~method are explained~~, while section~~. Section 4 presents the results and discussion of this work. The conclusions and outlook are ~~provided~~in section 5.

2 Study area and hydrometeorological datasets

This study ~~was tested~~area is located in a semi-humid region in the Baden-Württemberg state of Germany (Figure 1) ~~that characterized by temperate monsoon climate~~with temperate monsoon climate of mild winter and warm summer. ~~Elevations of this state range~~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. ~~Winters are mild whereas summers are warmer.~~The annual mean air temperature in Baden-Württemberg is about 10.2 °C. ~~Precipitation~~Rainfall is evenly distributed ~~through~~throughout the year. However, its seasonality shows a weak trend. The monthly rainfall ~~reaches its peak~~is highest in June ~~, whereas the month of October shows the least precipitation~~and lowest in October. The meteorological ~~observations~~data used in this study ~~was~~were provided by the German Weather Service (DWD). Daily air temperature data required for the rainfall-runoff model ~~was~~were interpolated on a 1×1 km^2 grid from the observations using the ~~algorithm of~~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).

~~Precipitation~~Rainfall data from a dense network of daily ~~precipitation~~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 ~~precipitation~~rainfall observations were used for this study. All ~~available data from~~data are available for the time period 1991-2010~~ was considered~~. The number of available daily stations and sub-daily stations varies according to different time period. Figure 2 illustrates the number of available observation ~~locations~~stations in Baden-Württemberg between ~~the years~~1991 and 2010. It can be seen from the graph, more than 430 daily stations ~~were available in 1991, while~~but only 30 sub-daily stations were available in 1991. The total number of daily stations decreased ~~dramatically~~to 250 around 2003 and remained ~~constant~~stable for the subsequent years. The number of sub-daily stations ~~kept~~has been increasing throughout ~~the whole~~this period and experienced a sharp increase from 100 to 200 in ~~the year~~2005. ~~The following~~Four different ~~precipitation~~rainfall datasets were ~~created according to the available observed data:~~generated and explained as follows.

1. High temporal resolution observed ~~precipitation~~rainfall was aggregated to hourly ~~time steps~~and then interpolated ~~subsequently~~to~~a~~ 1×1 km^2 ~~grid~~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 ~~will be~~is referred to as Sparse Hourly (SH) set.

2. Observed daily ~~precipitation~~rainfall combined with the daily aggregations of the high temporal resolution data were used to create ~~a~~ 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 ~~will be~~is referred to as Dense Daily (DD) set.

3. High resolution ~~precipitation~~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 ~~precipitation~~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 ~~denoted~~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 ~~practically~~the daily aggregations of the DH and SH sets. Note that DH is a dataset combining hourly ~~observations~~and ~~artificially disaggregate~~disaggregated daily gauge data. One of the research questions raised here is to find out if ~~a~~disaggregation leads to an improvement of model performance. Comparisons of the model ~~performances on the pairs~~performance using the inputs of (SD, SH) and (DD, DH) ~~provide information on~~pair will reveal the effect of temporal resolution. ~~While comparisons~~Meanwhile, comparison between (SD, DD) and (SH, DH) ~~, provide information on~~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~~as described previously~~. 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 ~~ranging~~range in size from 417km^2 to ~~about~~1300km^2, along with ~~a~~large ~~difference~~differences in elevation and annual precipitation. It can be seen clearly from the map that these four catchments have different rain gauge ~~density~~densities, the Schwaibach catchment ~~, which~~located in the ~~mountain~~mountainous area with ~~various~~elevations ~~(from 190m to 1028m),~~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 ~~introduced~~developed in the 1970s ~~at~~by the Swedish Meteorological and Hydrological Institute (SMHI) (Bergström and Forsman, 1973). ~~Due~~Thanks to its simplicity, low demand of inputs and ~~few~~small number of model parameters, the HBV model has been ~~a preferred model~~widely used for rainfall-runoff simulation and flood forecasting. Figure 4 represents the structure diagram of the HBV model (Singh, 2010). ~~In general,~~ There are three main modules ~~are included~~in the HBV model, namely snow routine, soil moisture routine and runoff routine (Hartmann, 2007; Singh, 2010).

~~First of all,~~In the snow routine, the snow accumulation and ~~melt~~melting process is estimated by the relatively simple degree-day method (Rango and Martinec, 1995) ~~using~~with two parameters: degree day factor (DD) and threshold temperature for snow/rain (TT) (as shown in Equation 1). ~~In this method, the~~The measured precipitation is supposed to be solid (snowfall) if the air temperature is lower than ~~the~~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, ~~which is represented by~~defined as the ratio of direct runoff to effective precipitation ($\frac{\Delta Q}{\Delta P}$), ~~can be estimated by~~is expressed as:

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

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

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

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

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

As shown in Equation 2, runoff ~~response routine~~ 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 ~~free~~parameters:

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

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

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

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

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.

~~Finally, a~~A transformation function ~~approach~~ with the triangular weighting parameter *MAXBAS* is used to smooth the ~~generated~~total runoff ($Q_0 + Q_1 + Q_2$) to obtain discharge at the outlet.

In this study, for investigating the sensitivity of model performance on the spatial resolution of input variables, two HBV models with different ~~levels of complexity~~spatially configurations were applied: lumped HBV and spatially distributed HBV, respectively. In the lumped model, precipitation, temperature and potential evapotranspiration were ~~supposed to be equally~~assumed uniformly distributed ~~among the~~within a catchment and all the processes were calculated for the whole catchment. Previous studies have indicated that the ~~altitude~~elevation is an important reason for the spatial differentiation of meteorological ~~elements, such as~~variables, including temperature, precipitation, evapotranspiration and snow melt are in reality not uniformly distributed within a catchment. They often exhibit dependence with elevation. ~~Therefore, the~~The spatially distributed HBV model ~~was constructed to separate the whole~~used in this study divides a catchment into several zones based on ~~topographic~~elevation. The 1×1 km^2 grid based ~~precipitation~~rainfall and temperature data were ~~computed averagely according to elevation zone and used as inputs for model simulation~~averaged for each elevation zone. In the spatially distributed model, the parameters associated with the snowmelt and soil moisture modules ~~related parameters can be adjusted differently~~were calibrated for each elevation zone. The parameters ~~controlling~~associated with the runoff response ~~processes~~module were ~~estimated for the whole~~calibrated for each catchment similarly to the lumped model (Das et al., 2008).

~~There are~~Out of the 15 parameters ~~describing~~within the HBV model, ~~where only~~9 parameters were ~~selected for calibration~~calibrated in this study. Table 2 lists the initial upper and lower limit of the to-be-calibrated parameters ~~that will be optimized by model calibration~~using historical data. The data depth based parameter optimization method-Robust Parameter Estimation (ROPE) algorithm (Bárdossy and Singh, 2008) was applied for model parameter ~~identification~~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 ~~similar and good~~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 ~~simulated result and~~model simulations corresponding to the model parameters using different objective functions differ considerably as they have different ~~focus~~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 ~~floods~~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 ~~efficiency~~coefficient ~~is one of the most widely used performance criteria in model simulation. It~~focuses on high flow as it evaluates the squared difference between simulated and measured streamflow. ~~NS efficiency~~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 ~~discharges~~discharge, respectively, and $\bar{Q}_o$ is the mean of observed discharge~~ series~~.

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

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

~~Here~~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 ~~has been~~was applied in this study to ~~separate~~divide the ~~whole~~data ~~series~~into two changeequal10-year periods: 1991-2000 and 2001-2010. Model calibration was carried out for both time periods and then a cross-validation analysis was performed~~ as well~~. For each calibration run, the first water year data was used as a warm-up period ~~to reduce initial errors~~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 ~~assigned~~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. ~~We also assessed the~~The effects of the temporal and spatial resolutions of the rainfall inputs on the model performance were assessed in terms of ~~Nash-Sutcliffe efficiency~~NS coefficient and the ~~mean square error~~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 ~~taken~~used to divide ~~the whole study~~a catchment into ~~several~~different elevation zones. Note that all the model calibrations were performed on the basis of simulating daily discharge. ~~Due to the limitation of observed temperature, air temperature and potential evapotranspiration were assumed to be constant over the whole day.~~

We ~~also wonder~~investigated if the combination of daily ~~scale model~~and hourly ~~scale model leads to a~~models can lead to better prediction in streamflow. It is interesting to investigate the similarities of different temporal resolution. Therefore, the common calibration ~~tragedy~~approach was ~~proposed in this study~~used to calibrate the daily ~~scale model~~and hourly ~~scale model~~models simultaneously. This ~~kind of~~approach ~~is expected to~~may identify robust model parameters ~~for the application of model in~~that are applicable using different temporal resolutions. ~~Common~~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 ~~indicates~~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

~~Firstly, the~~The quality of the rainfall ~~products~~datasets was assessed and compared for the four selected catchments. As the SD and DD sets are ~~the~~daily aggregations of the SH and DH sets, ~~here~~we only compared the daily ~~precipitation~~rainfall sets SD and DD for both calibration ~~decades~~periods (Figure 5). It can be seen clearly from the figures that the interpolated ~~precipitation~~rainfall datasets display some difference for all study catchments. The asymmetry of the scatterplots is ~~fairly obvious~~evident for the ~~first decade (1991-2000)~~period 1991-2000. In general, the DD dataset leads to higher value than the SD dataset. ~~The reason behind this~~It is mainly because the low density of sub-daily observations during the period of 1991-2000 leads to ~~big~~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 ~~become similar in magnitude along with the increasing of available sub-daily observations.~~are in closer agreement due to higher density of sub-daily gauges.

4.2 ~~Calibration and validation model performance~~Results of calibration and validation

As ~~designed~~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. ~~Data series from 1991 to 2010 were average split into two sub-periods~~Two 10-year time periods 1991-2000 and 2001-2010 were used for calibration and cross-validation. ~~This leads to~~In total 16 calibration runs and 16 validation runs were performed for ~~every~~each catchment. As ~~mention~~mentioned before, each simulation ~~could obtain~~obtained 10,000 parameter sets with similar model performance. ~~To make it simple, we took~~We then used the mean value of the ~~corresponding~~10,000 model performances to ~~represent~~quantify the model ~~efficiency~~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 ~~indicate~~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 ~~the~~Pforzheim ~~catchment~~ has the worst mean NS performance of 0.58 for all calibration runs. Moreover, for a specific catchment, the calibrated ~~model performances~~models perform differently for different data period~~ are also different~~. For the Schwaibach and Pforzheim ~~catchments~~catchment, the calibrated model performance for the ~~time~~period of 2001-2010 is ~~obviously~~better than the performance for the time period 1991-2000 for most of the datasets. This might be due to the increasing ~~of the~~rain gauge density inside or ~~nearby~~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 ~~weak~~poorly for both calibration and validation~~; the loss in NS~~

~~coefficient~~. NS coefficient using SH and SH inputs is about 0.3 ~~when compared to the corresponding~~less than the results of the sets DH and DD. This indicates that systematic interpolated ~~precipitation~~rainfall errors have critical influence on model calibration.

~~The flexibility of model in flood prediction is analyzed with the behavior of high flow~~We then analyzed if the model is robust for simulating high flows. Tables 5 and 6 list the mean square ~~error of flows higher than the~~errors of the top 10th percentile of ~~flow~~flows for the lumped ~~model~~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~~, while~~. Figure 7 shows the corresponding results for flows higher than the 10th percentile of flow. Results indicate ~~obviously~~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 ~~the~~flood simulation.

4.3 ~~Comparison of the~~Model performance ~~corresponding to the temporal resolution~~using different temporal resolutions of rainfall data

Firstly, the model performance of different temporal ~~resolution~~resolutions of rainfall was compared for ~~all~~four datasets and two model ~~structures~~spatial configurations. For the pairwise comparison, all the conditions are the same in the model except for the rainfall temporal resolution ~~of input variables~~(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 ~~variables as model input for the precipitation sets~~inputs that were interpolated using only high-resolution ~~precipitation~~rainfall observations (SH, SD). Figure 9 compares the ~~corresponding~~results ~~for~~from the rainfall ~~datasets~~inputs that incorporated observed daily value with high-resolution observations (DH, DD). The result shows that all the scatters are ~~laying~~lying below the diagonal line for the different level of observation density. For both calibration and validation periods, the simulations using hourly input data ~~as model input~~outperform the ~~one that~~ones based on the daily resolution. For the dataset with low observation network density, the ~~average~~averaged NS ~~value~~of set SH is about 0.73 for the calibration period and 0.68 for the validation period, while the mean NS coefficient ~~that was~~calibrated using SD set is 0.67 and 0.6, respectively. The higher observation density datasets show a similar tendency. The mean NS ~~value~~of using DH set is around 0.79 for calibration and 0.77 for validation, while ~~the result~~that of set DD is 0.72 and 0.69, respectively. The fact that the hourly scale model performs better than the daily model ~~indicating~~suggests that the dynamic runoff of catchment could be better simulated with a higher temporal resolution of ~~input variables~~rainfall. According to the distances from the diagonal to the scatter plots, we ~~could find~~can observe that the difference in model performance for different temporal ~~resolution~~resolutions is larger for the catchments with relatively low NS model performance, such as Schwaibach and Pforzheim. For Rottweil and Kocherstetten, the ~~model~~performance of hourly calibrated model is only slightly better than the daily ~~based~~model.

4.4 ~~Comparison of the~~Model performance ~~corresponding to~~in terms of observation density

~~Results also indicate that~~The rainfall ~~data~~gauge network density has significant impact on model simulation and parameter optimization. Figure 10 ~~plots the simulated NS coefficient for the daily datasets that was interpolated using different density of rainfall observation network.~~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 ~~of~~using the DD set is generally better than ~~the result of~~that using the SD set for both calibration and validation ~~periods~~period. The ~~average~~averaged NS model performance of DD ~~set over all simulations is about~~and SD sets are 0.71 ~~while the value for SD set is~~and 0.64, respectively. The model performance ~~for the hourly based simulation~~using hourly inputs shows similar trend as ~~the model performance for the daily time step~~that using daily inputs. As shown in Figure 11, the model ~~calibration of~~using the DH set ~~outperform~~outperforms the ~~result of~~one using the SH set. ~~The~~These results demonstrate that ~~the high observation density could lead to considerable~~high rain gauge density lead to improvement of model performance ~~for~~at both daily and hourly time ~~scales~~resolution.

Figure 12 illustrates the cumulative distribution function of NS ~~model performance~~coefficient using sets SD, SH and DH for model calibration (left) and validation (right). As can be seen clearly from the curves, if ~~precipitation~~rainfall data comes from a sparse network of sub-daily stations, use of higher temporal resolution datasets (~~as represented by set SH~~the SH set) ~~can achieve~~leads to better model performance than ~~the lower ones (as represented by set SD)~~using lower resolution ones (the SD set). ~~Decreasing the length of time step in model simulation could provide a better fit~~Simulation of daily streamflow can benefit from running the model at a higher temporal resolution. In addition, the combination of observed ~~high-resolution observations~~sub daily rainfall with ~~disaggregate~~disaggregated daily ~~precipitation~~rainfall (~~as represented by set DH~~the DH set) leads to a further improvement of daily streamflow ~~prediction~~simulation.

4.5 ~~Comparison of the~~Model performance ~~corresponding to the~~in terms of spatial resolution of rainfall data

The model performance ~~of different model structures in terms of different spatial resolutions was assessed by comparing performance for lumped HBV model and spatially distributed HBV model~~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 these two model structures~~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 ~~elevation zone based~~spatially distributed ~~models outperform~~model outperforms the lumped, ~~ones~~ especially for the catchments having high NS coefficient~~. Despite~~, despite the increase in model performance being only marginal. However, for the catchments with relatively ~~weak~~poorer model performance, the lumped model could even lead to slightly better performance than the semi-distributed model, ~~structure,~~especially for the validation period ~~that~~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 ~~structure raises the~~has a larger number of parameters to be ~~identified~~calibrated and the parameters are underestimated during the calibration period. We ~~can~~conclude ~~from this comparison~~that the improvement in spatial resolution of model structure did not ~~clearly~~enhance the model performance, ~~. However, it is surprising since we expected a better model performance with a higher spatial resolution of model and a complicated set of parameters~~which is surprising since higher spatial resolution and more model parameters are expected to improve the model performance. ~~The~~Our results ~~support~~confirm the findings of Das et al. (2008) that the distributed ~~model structures does~~models do not ~~significantly~~necessarily improve model performance.

The ~~complex structure version of~~distributed model did not perform better than the lumped model in ~~current research~~this study. This ~~might be due to the lack of~~could be because the catchment underlying surface information and/or the calibration procedure was not ~~enough~~sufficient for ~~the identification of~~identifying optimal distributed model parameters. A second ~~explanation~~reason could be ~~that~~the temporal resolution of the ~~force~~rainfall inputs is not sufficient for the distributed model ~~structure~~.

4.6 Common ~~calibration of models~~model calibration with different temporal resolutions

As shown before, the combination of hourly ~~observations~~and daily ~~observations lead~~gauge data leads to the improvement of data quality as the model using sets DH and DD ~~show~~has better ~~model~~performance than ~~the~~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 ~~value of time step dependent~~parameters (DD, K_0, K_1, K_d and K_2) that are dependent on time steps should be converted according to the ~~temporal resolution of~~simulation step of the model. The common calibration was performed for ~~two decades~~the two time periods separately, and ~~the~~a cross-validation analysis was performed as well. The common calibration and validation results were compared with the individual calibration ~~cases~~(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 ~~rather~~similar to individual calibration results. The average loss of NS ~~model performance~~coefficients over all catchments is about 0.02 for set DH and 0.01 for set DD. For the validation period, ~~from the scatter plots,~~it ~~is clearly seen~~can be seen from the scatter plots that the common parameters outperform the individual ones for about half of the all simulations. It ~~indicates~~suggests that ~~common calibrated parameters based on different time steps could be a feasible approach for increasing~~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 ~~tragedy could~~strategy can provide more information for identifying model parameters.

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

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 ~~HBV~~and spatially distributed HBV ~~models~~model were applied to simulate ~~the~~daily runoff for four mesoscale catchments driven by four different ~~types of~~~~precipitation~~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 ~~simulation~~simulations were evaluated using ~~Nash-Sutcliffe efficiency~~the NS coefficient and the mean squared error of flows higher than the 10th percentile of flow. The ~~sensitivity of model~~model sensitivities to the temporal and spatial resolutions of rainfall ~~was~~were compared. In ~~additional~~addition, the common calibration approach was proposed to calibrate the models with different time steps simultaneously for ~~seeking~~finding robust model parameters.

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 ~~completely~~outperforms the daily based simulation, indicating that higher temporal resolution ~~could~~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 ~~disaggregating~~disaggregation approach ~~provides~~is an effective way ~~for~~of increasing the temporal resolution of rainfall data and the model performance~~ of model simulation~~. However, the lumped and spatially distributed HBV model perform very similarly, indicating that higher ~~model~~spatial resolution does not or only marginally improve the model performance for the study catchments. The result ~~supports~~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 catchments are smaller than 2000km^2 ~~and have~~with relatively uniform ~~precipitation~~rainfall.

As ~~discussed~~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 ~~research indicates~~study shows that ~~rainfall~~data disaggregation ~~approach could~~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 temporal resolution of ~~precipitation~~rainfall inputs with disaggregation method ~~could~~can be an easier and more efficient way to improve model performance, compared ~~with~~to increasing the model spatial resolution ~~that comes~~at a cost of increasing the complexity of model structure and parameters.

This study focuses on high flows and uses only the ~~Nash-Sutcliffe efficiency~~NS coefficient as ~~the objective function to investigate the~~a quantitative measure of model sensitivity. As model performance highly depends on the selection of objective functions, the model sensitivity can be different if ~~using different performance criteria~~the model performance 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 ~~could~~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 (2016YFC0402701), National Science Foundation of China (51879067), Fundamental Research Funds for the Central Universities of China (2015B28514), the China Postdoctoral Science Foundation(2017M621614), 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 resources research, 23, 1717–1737, 1987.

Bardossy, A. and Das, T.: Influence of rainfall observation network on model calibration and application., Hydrology and Earth System Sciences,12,1(2008-01-25), 12, 77–89, 2008.

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

Bárdossy, A. and Pegram, G.: Space-time conditional disaggregation of precipitation at high resolution via simulation, Water Resources Research, 2016.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters., Hydrology and Earth System Sciences, 12, 1273–1283, 2008.

Bárdossy, A., Huang, Y., and Wagener, T.: Simultaneous calibration of hydrological models in geographical space, Hydrology and Earth System Sciences, 20, 2913–2928, 2016.

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff model., Nordic Hydrology, 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, Journal of Hydrology, 299, 166–179, 2004.

Beven, K.: Model predictions: Uncertainty, Encyclopedia of Earth Science, pp. 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, Journal of Hydrology: 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, Hydrological Processes, 9, 69–81, 1995.

Das, T., Bárdossy, A., Zehe, E., and He, Y.: Comparison of conceptual model performance using different representations of spatial variability, Journal of Hydrology, 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?, Journal of hydrology, 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, Journal of Hydrology, 555, 314–322, 2017.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of hydrology, 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, Journal of Hydrology, 377, 80–91, 2009.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Applied engineering in agriculture, 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, Journal of Geophysical Research: Atmospheres, 113, 2008.

Hofierka, J., Parajka, J., Mitasova, H., and Mitas, L.: Multivariate interpolation of precipitation using regularized spline with tension, Transactions in 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, Journal of Hydrology, 292, 281–295, 2004.

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, Environmental Modelling & Software, 16, 309–330, 2001.

Kobold, M. and Brilly, M.: The use of HBV model for flash flood forecasting, Natural Hazards and Earth System Science, 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, Hydrology and Earth System Sciences, 18, p–575, 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, Biotechnologie, Agronomie, Société et Environnement, 17, 392, 2013.

Matheron, G.: Principles of geostatistics, Economic geology, 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 and Spatial Rainfall Sampling, Journal of Hydrologic Engineering, 14, 711–720, 2009.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models. 1. A discussion of principles, Journal of Hydrology, 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, Journal of hydrology, 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 dataset to drive a hydrological model: A case study of the summer 2007 floods in the Upper Severn, UK, Hydrology Research, 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, Journal of Hydrology, 470, 138–157, https://doi.org/10.1016/j.jhydrol.2012.08.041, <GotoISI>://WOS:000312686200012, 2012.

Rango, A. and Martinec, J.: REVISITING THE DEGREE-DAY METHOD FOR SNOWMELT COMPUTATIONS1, JAWRA Journal of the American Water Resources Association, 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, Hydrological Processes, 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 performance in a humid region of China, Journal of Hydrology, 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 Resources Research, 54, 8983–8998, 2018.

Table 1. Catchment characteristics for the 4 selected catchments.

No.	Stream gauge name	Longitude (^oE)	Latitude (^oN)	Area (km^2)	Elevation (m)	Annual ~~precipitation~~rainfall (mm)	Average temperature (^oC)	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	~~Precipitation~~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	~~Precipitation~~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 ~~of the flow~~for flows higher than the 10th percentile of flow for the lumped HBV model.

Catchment	~~Precipitation~~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~~of the flow~~for flows higher than the 10th percentile of flow for the distributed HBV model.

Catchment	~~Precipitation~~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

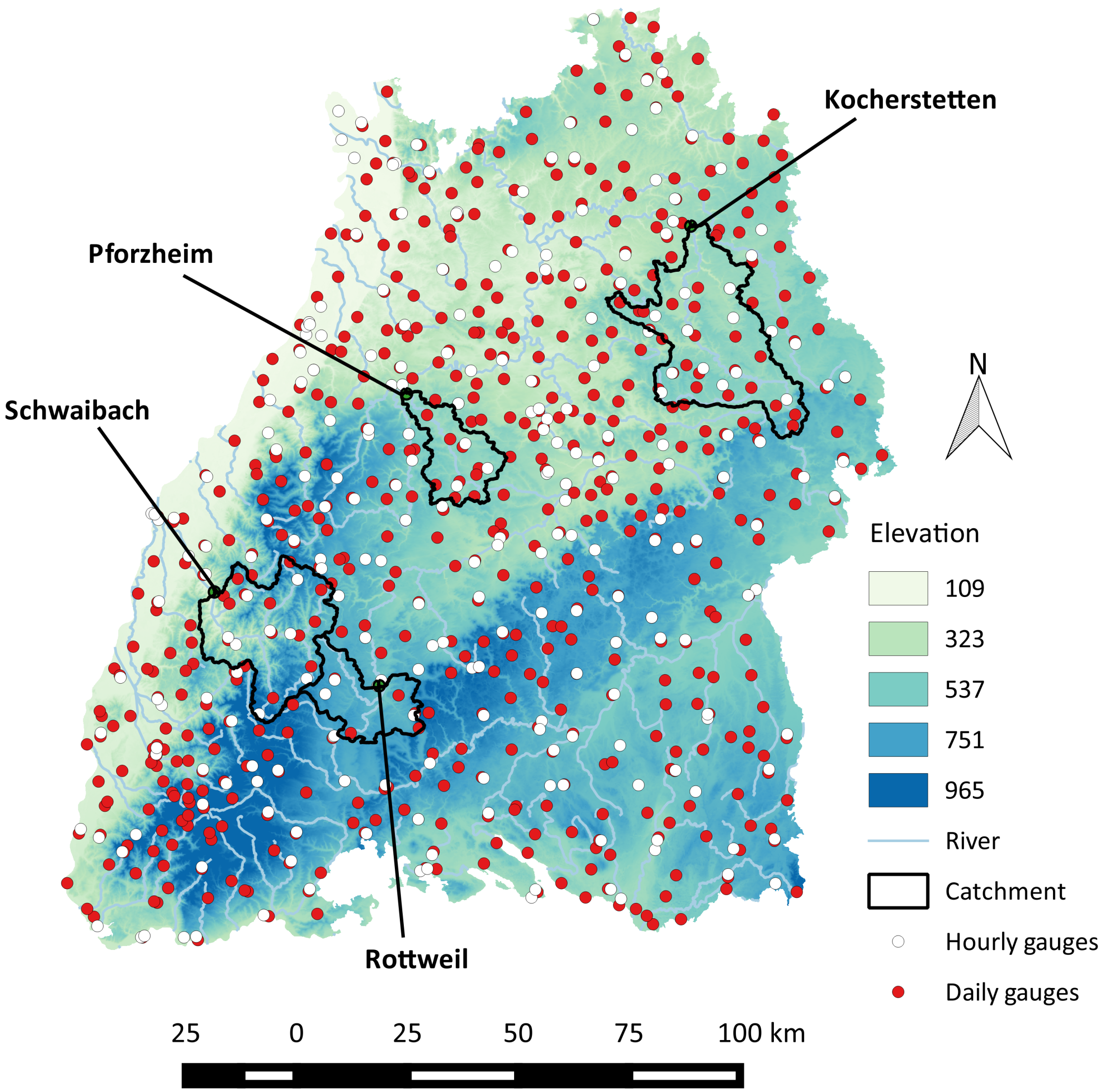

Figure 1. Locations of the ~~pluviometers(hourly)~~sub-daily and daily rain gauges in Baden-Württemberg and the four selected catchments.

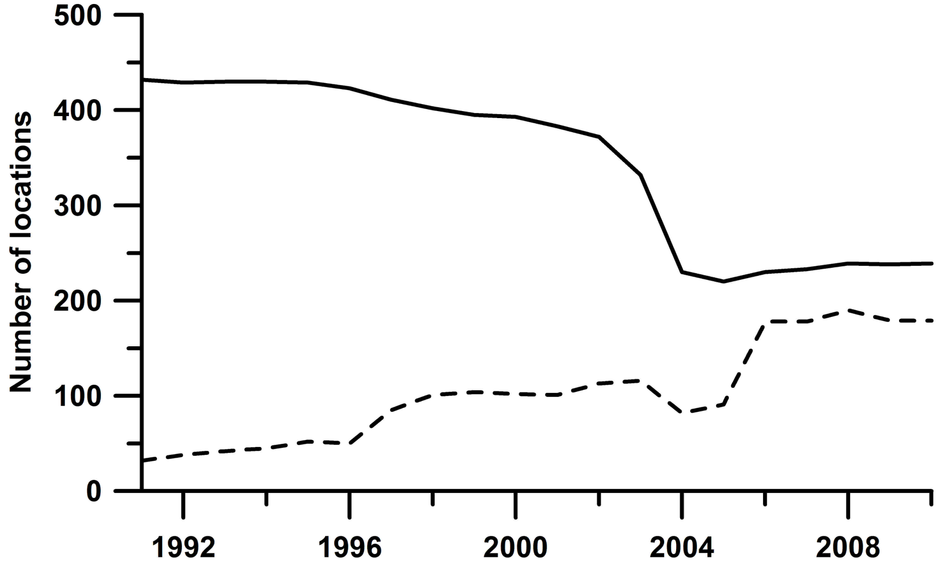

Figure 2. The number of available observation locations. Daily stations - solid line, Sub-daily stations - dashed line.

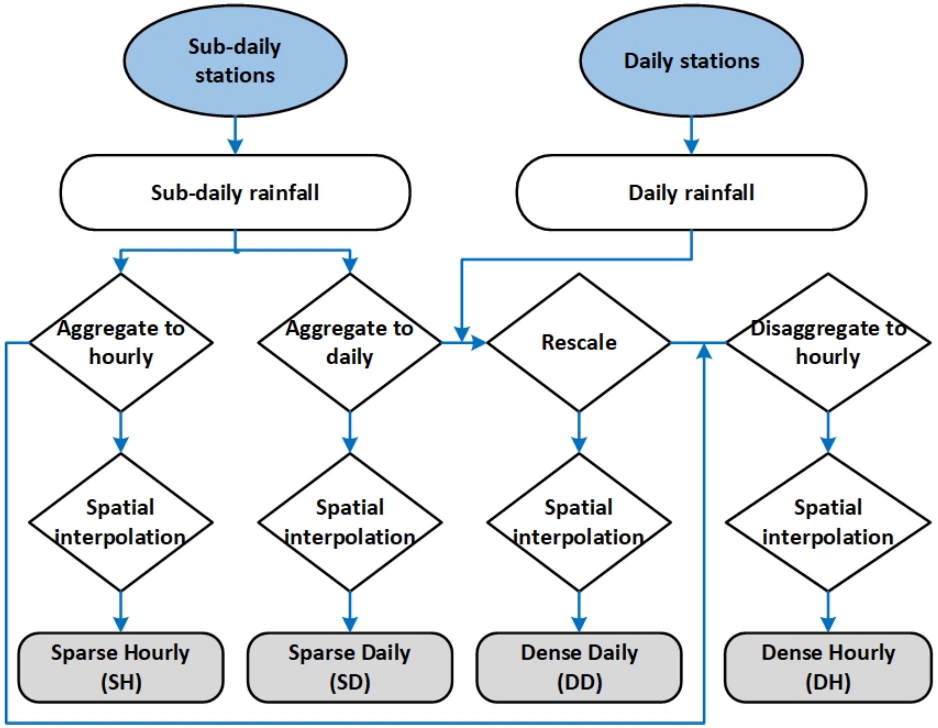

Figure 3. Schematic representation of four different ~~precipitation data sets~~rainfall datasets.

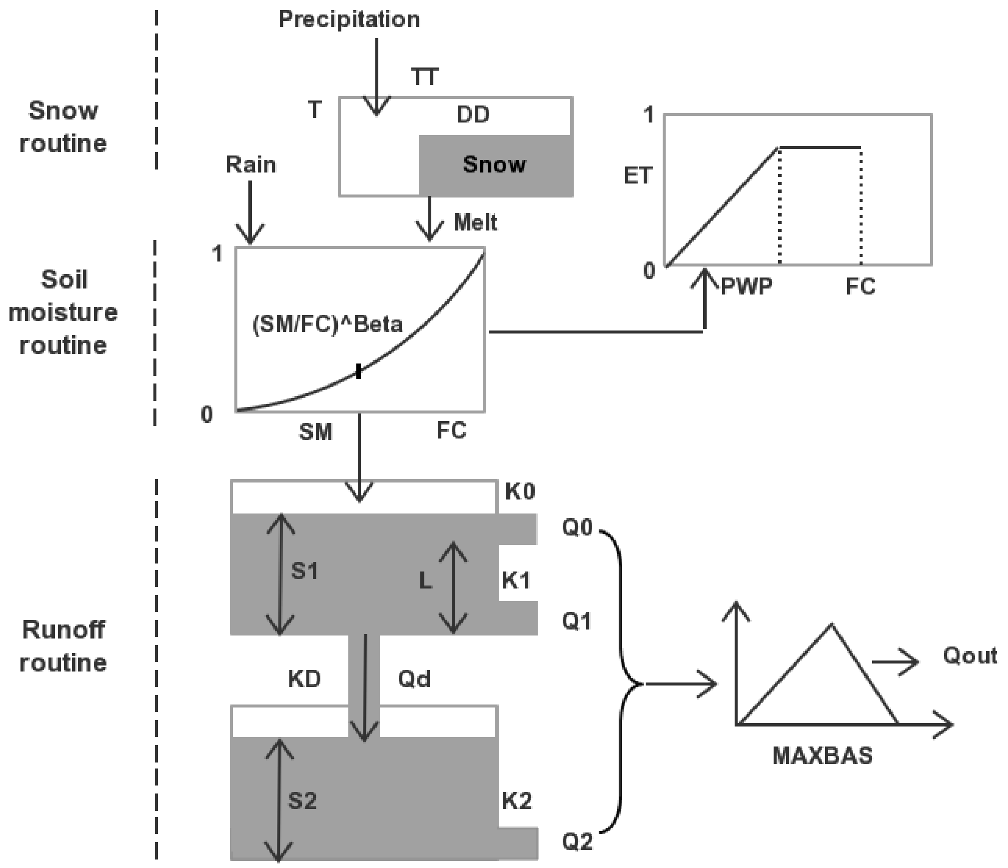

Figure 4. Schematic representation of HBV model.

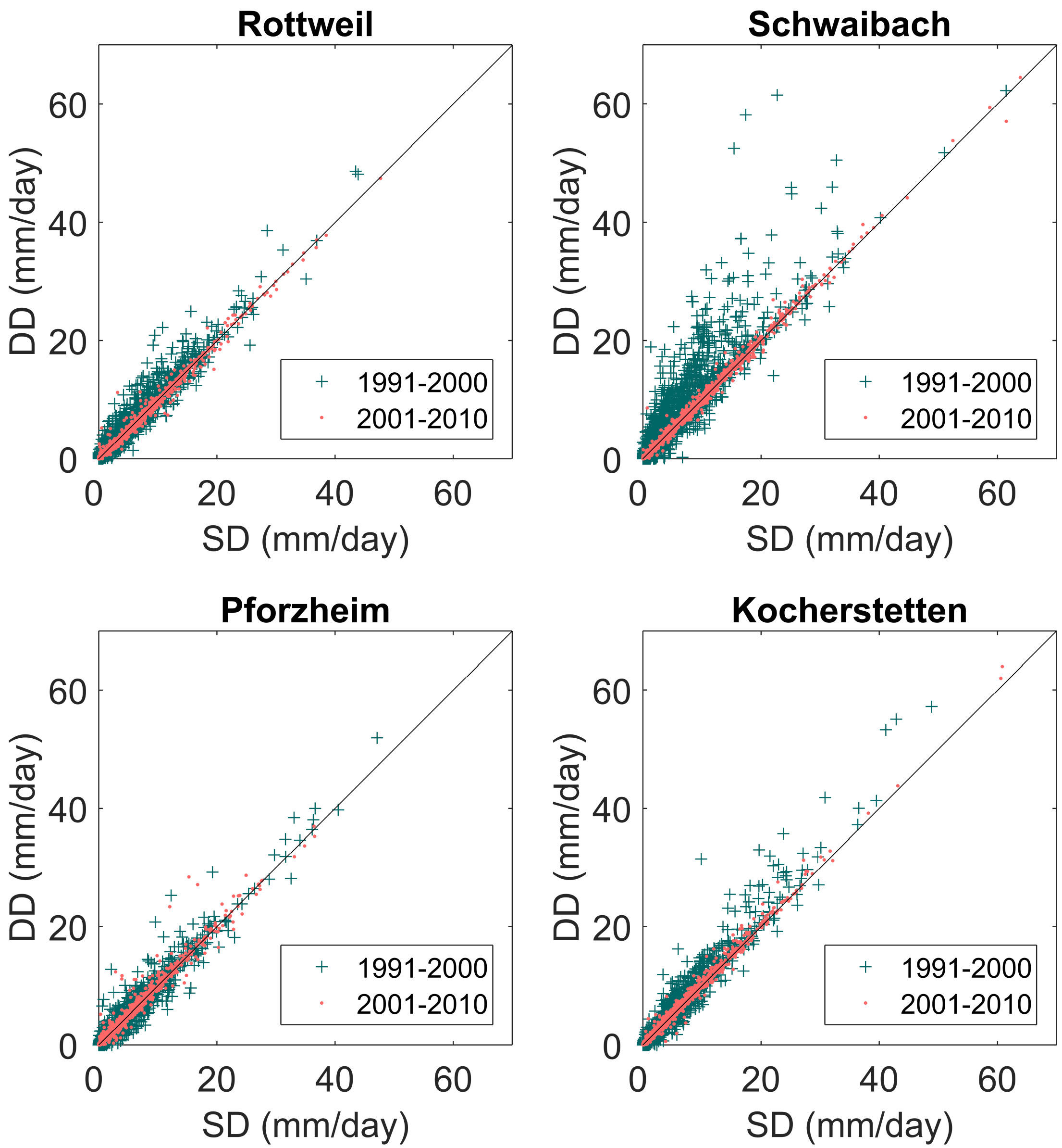

Figure 5. Comparison of the daily ~~precipitation~~rainfall data that interpolated using different ~~observation network density~~densities of rain gauge network.

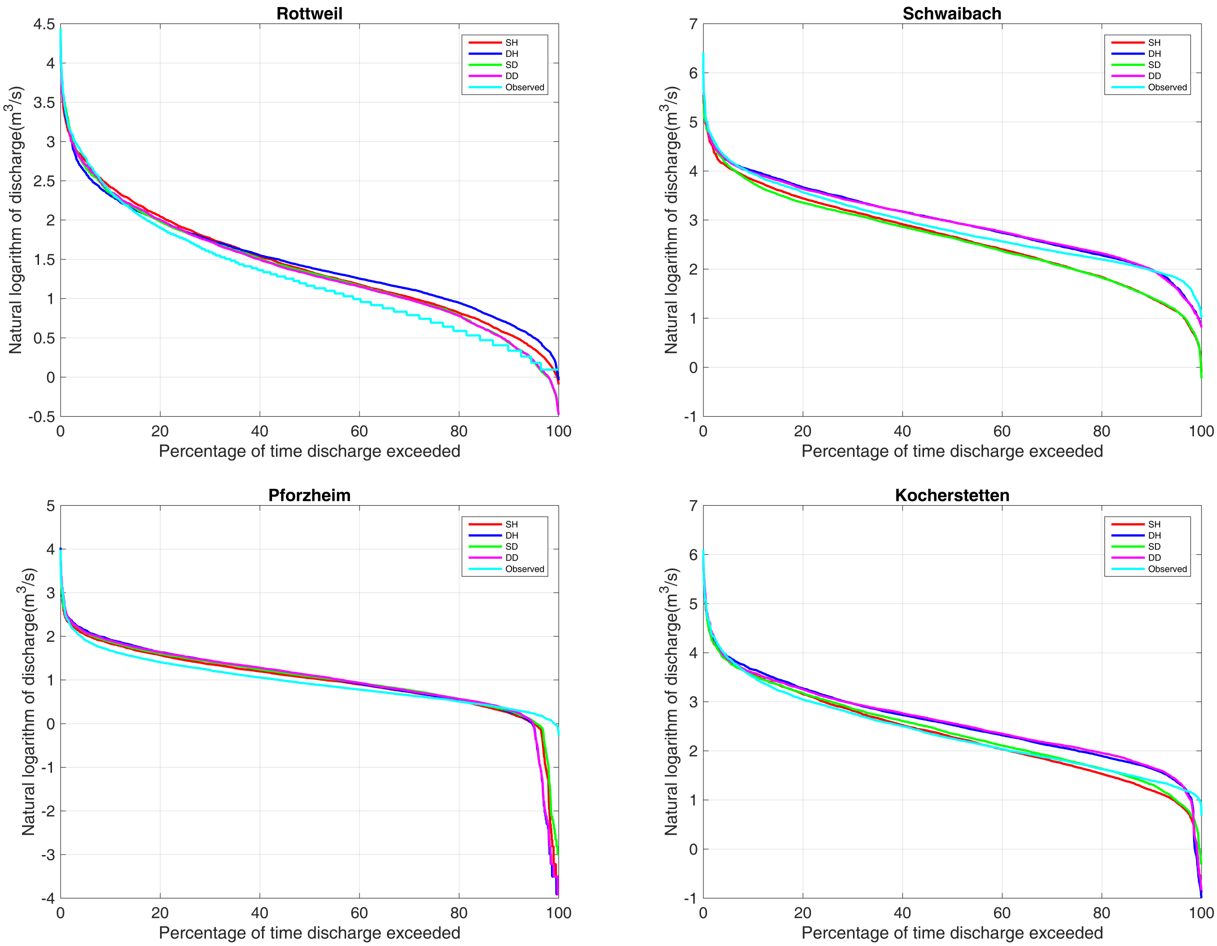

Figure 6. Comparison of the simulated flow duration curve.

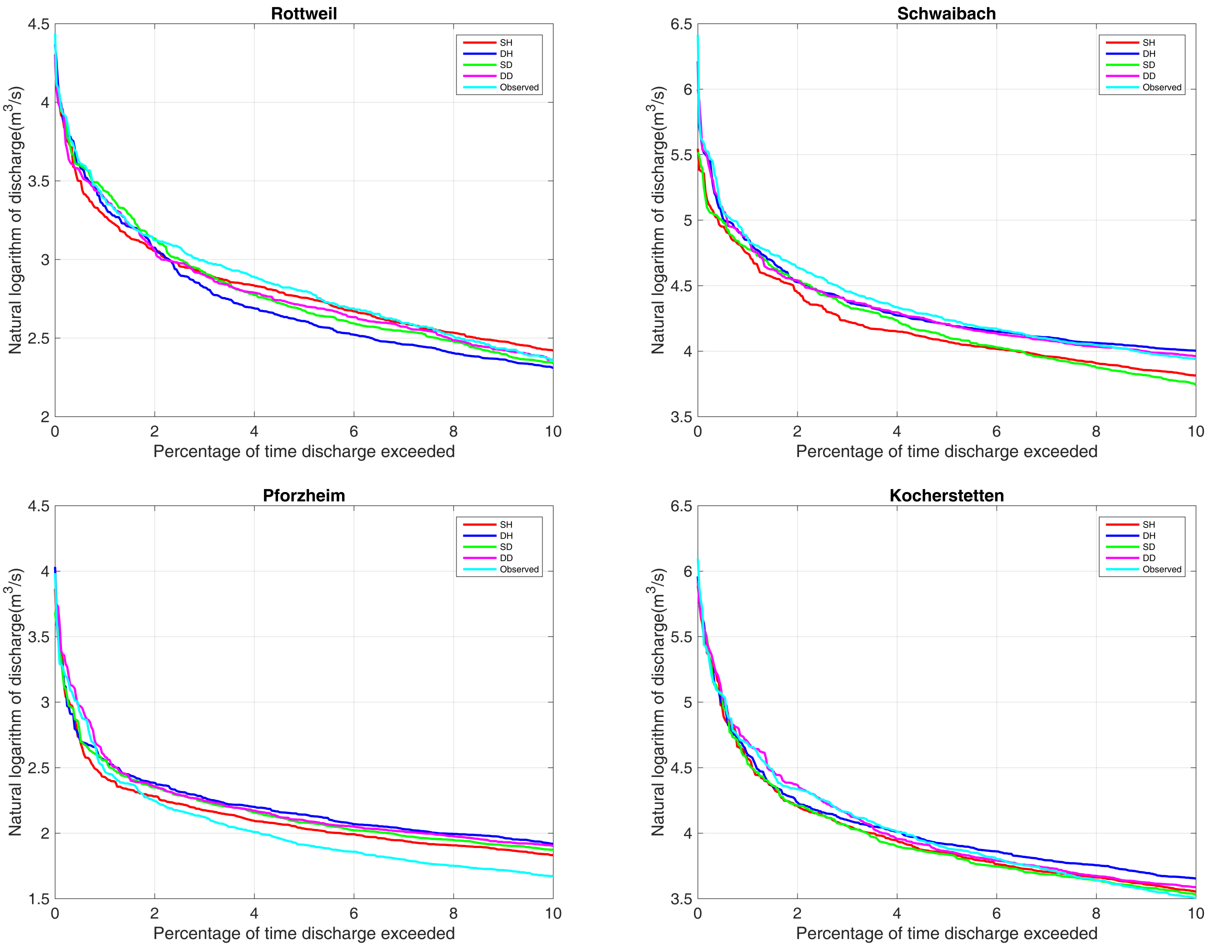

Figure 7. Comparison of the simulated flow duration curve for flows higher than the 10th percentile of flow.

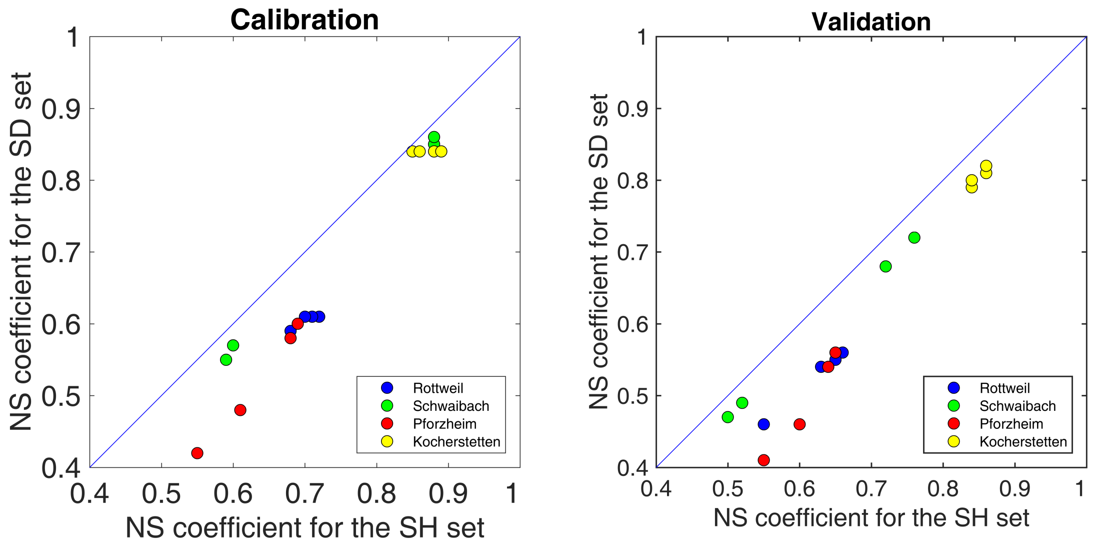

Figure 8. Comparison of NS ~~model performance~~coefficient for using hourly and daily ~~variables~~rainfall as model input for the SH and SD sets.

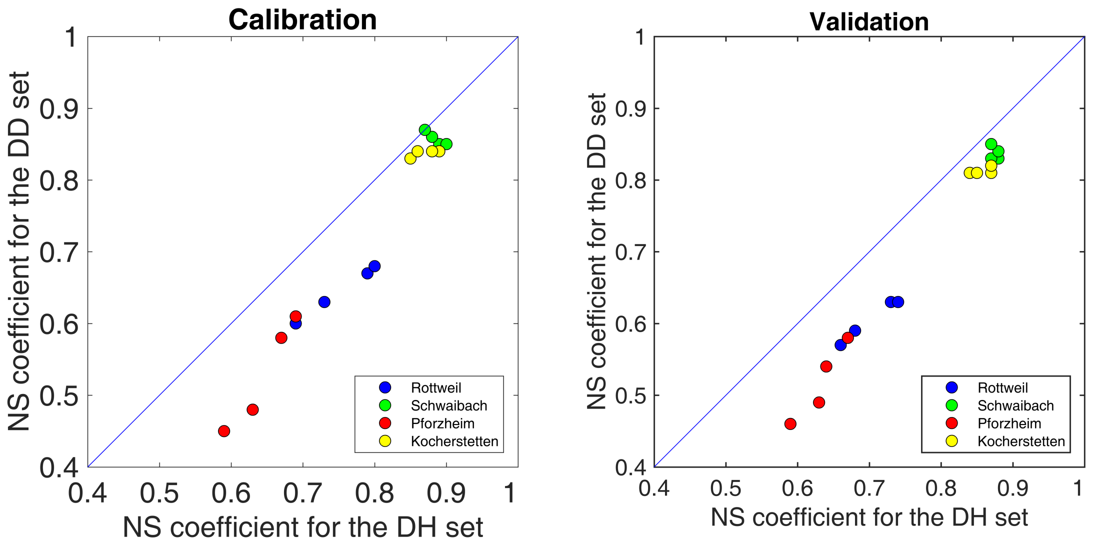

Figure 9. Comparison of NS ~~model performance~~coefficient for using hourly and daily ~~variables~~rainfall as model input for the DH and DD sets.

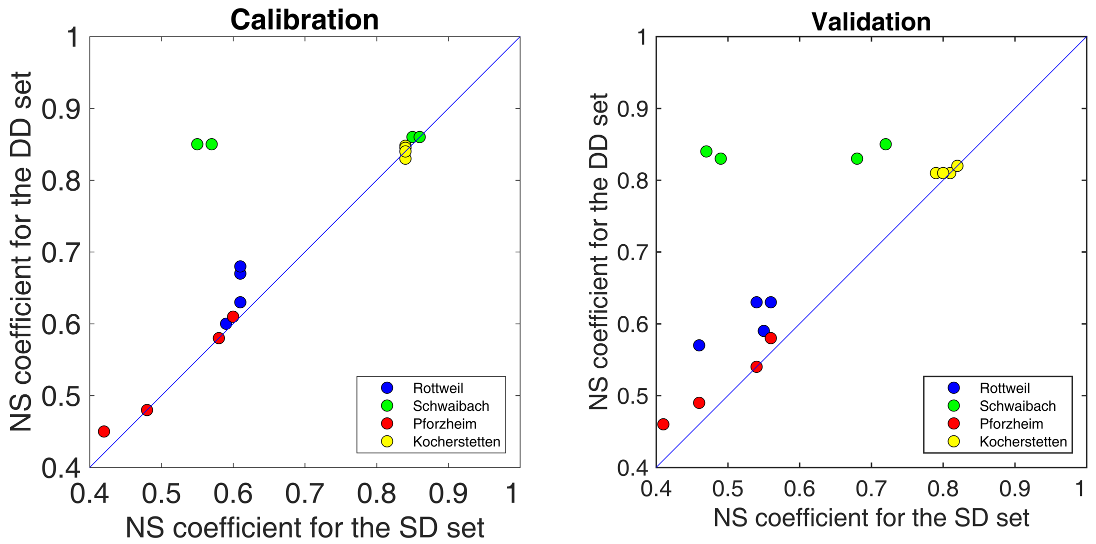

Figure 10. Comparison of ~~model performance~~NS coefficient for different ~~density~~densities of ~~rainfall observation~~rain gauge network, models were simulated based on daily time step.

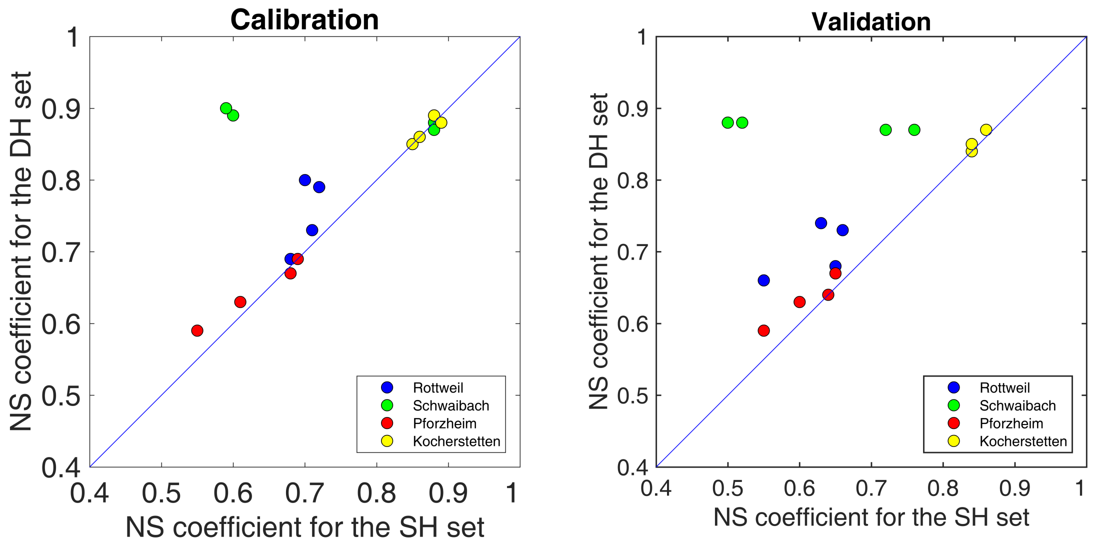

Figure 11. Comparison of ~~model performance~~NS coefficient for different ~~density~~densities of ~~rainfall observation~~rain gauge network, models were simulated based on hourly time step.

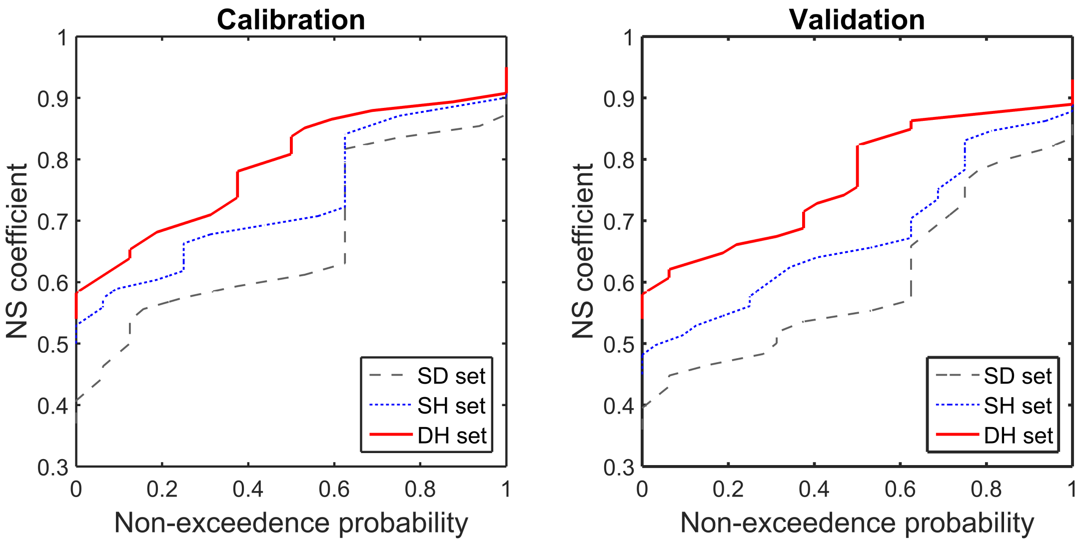

Figure 12. Cumulative distribution of NS coefficient for model calibration using different ~~precipitation~~rainfall datasets .

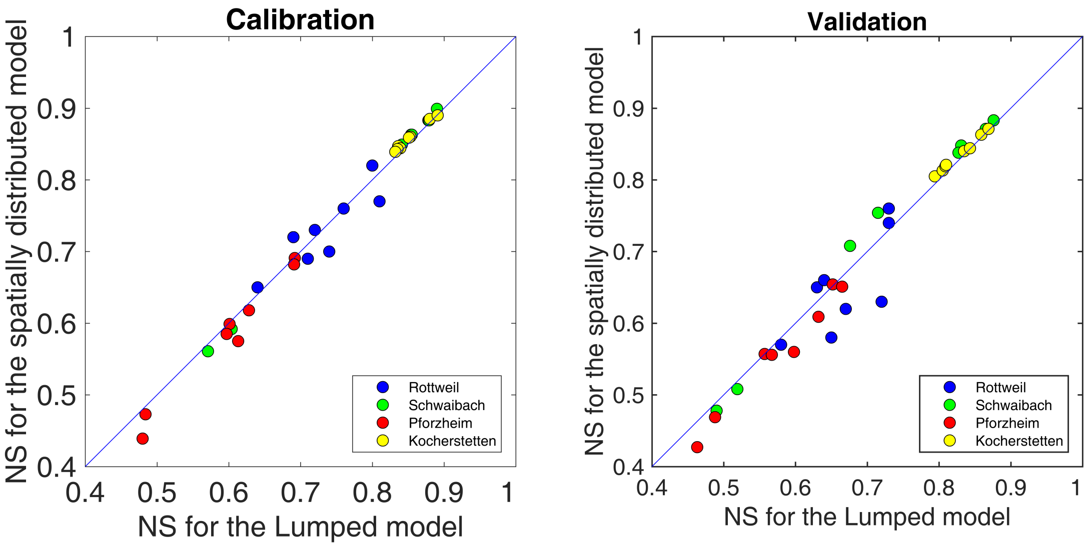

Figure 13. Comparison of ~~model performance~~NS coefficient for different spatial resolution of model structure.

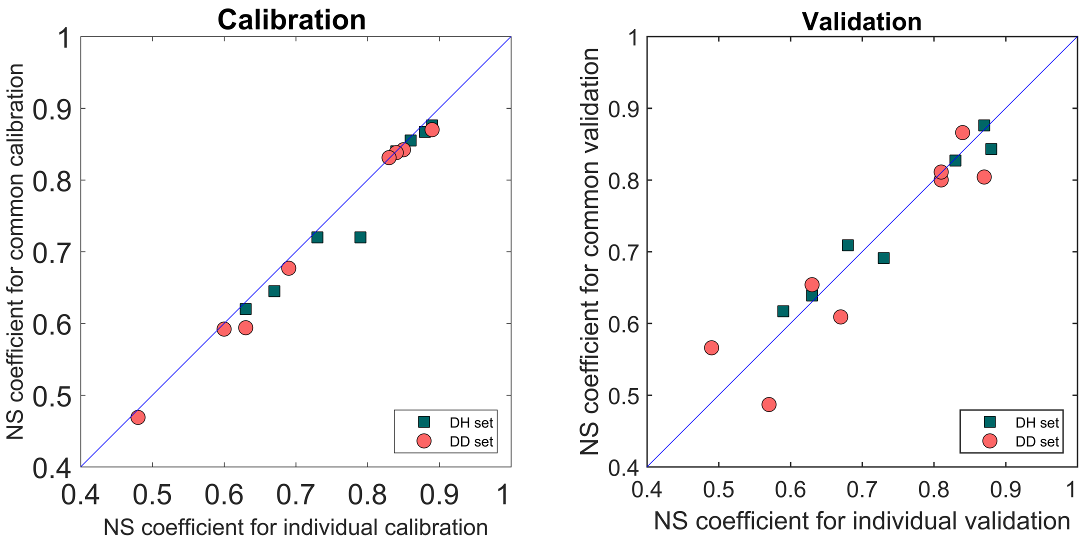

Figure 14. Comparison of ~~model performance~~NS coefficient for individual calibration and common calibration ~~for~~using datasets with different temporal ~~resolution datasets~~resolutions.

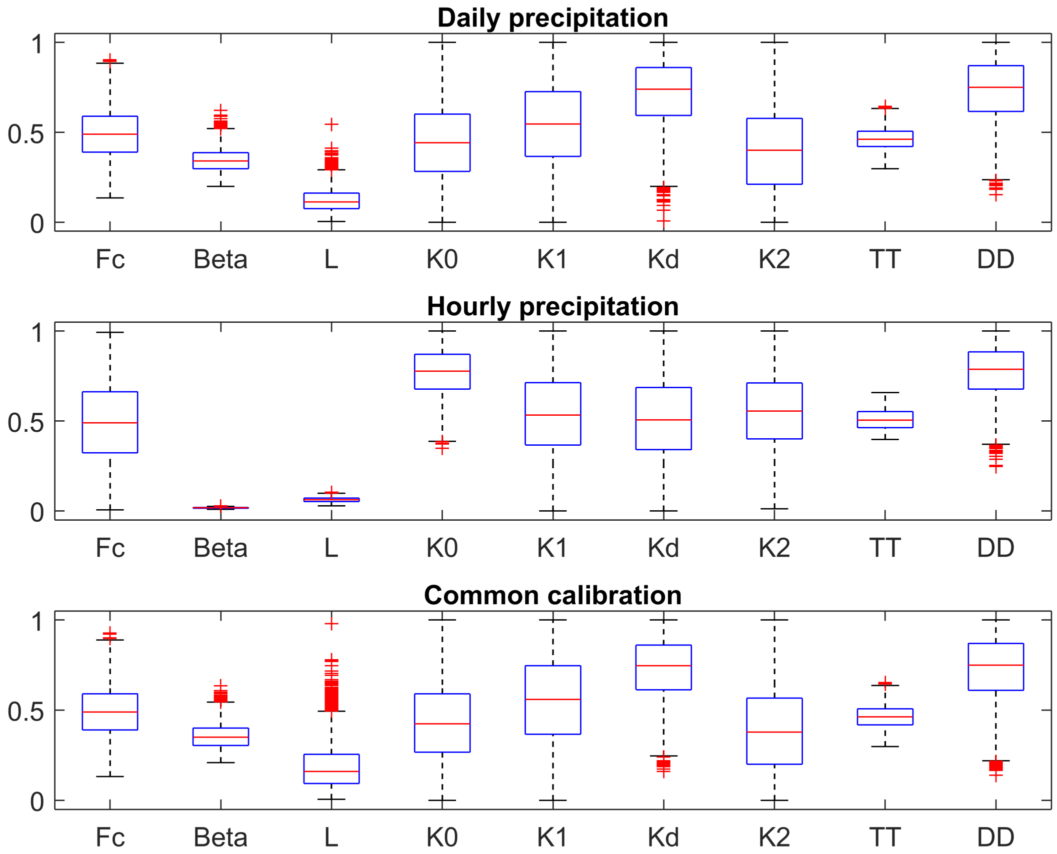

Figure 15. Comparison of model parameters for different temporal ~~resolution~~resolutions for Rottweil catchment.

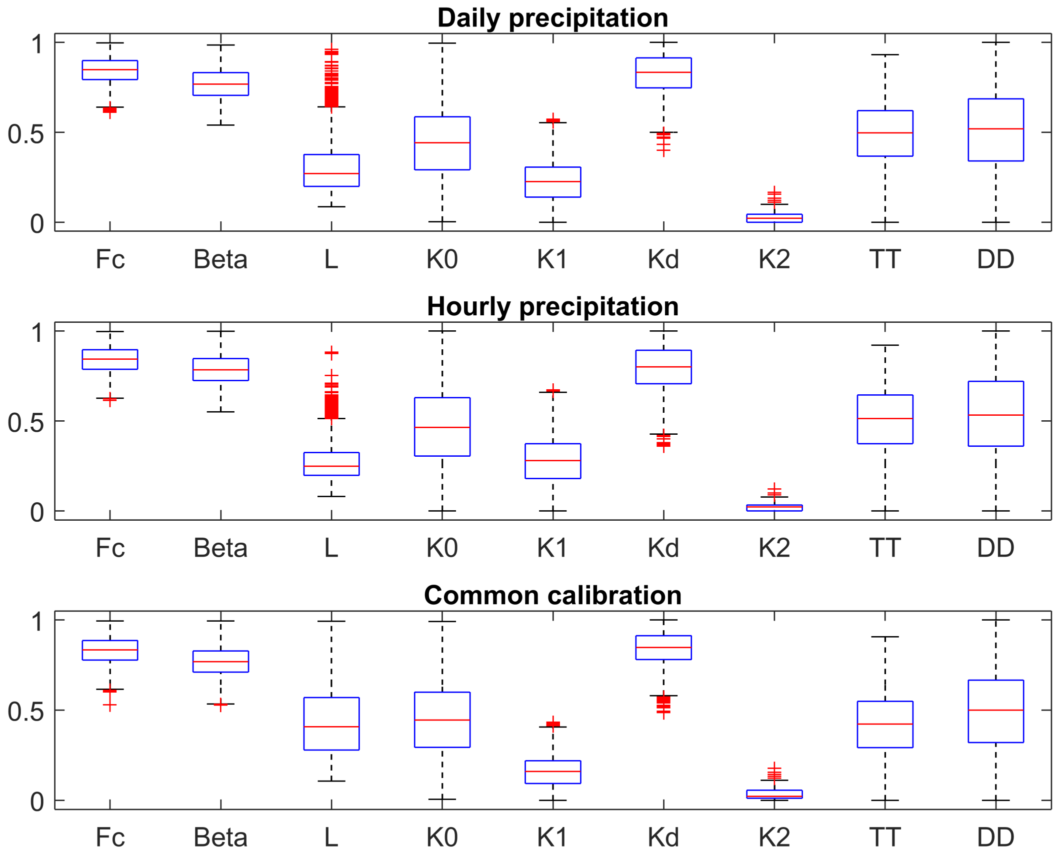

Figure 16. Comparison of model parameters for different temporal ~~resolution~~resolutions for Pforzheim catchment.