Point-by-point response to the Referees

We would like to sincerely thank the anonymous reviewers and editor for their review of the manuscript. We have considered all the comments and provided detailed responses of how each comment has been addressed in the revised manuscript below.

---

1. Response to Anonymous Referee #1

Major Comments:

Introduction – The introduction is quite short and I don't think gives the reader a thorough overview of previous literature on this topic and where this research sits within the field. There have been lots of other studies that have focused on the impacts of spatial and temporal resolution of rainfall on hydrological model output and you need to clearly explain how your research builds on these previous studies. I found it difficult to identify from the introduction what the research gap was and how this study addressed that research gap.

Response: Thank you for the comments. We have rewritten the literature review for the manuscript. The revised version contains an updated introduction, referring to the ongoing progress of the study for the sensitivity analysis of rainfall data to model performance both on temporal and spatial scales. We described in more details about the attempts for improving model performance and the monition of our study. We compared and discussed our idea with previous work on impacts of input variables in hydrological models.

Study area and hydrometeorological datasets – The rationale for your choice of catchments needs to be outlined. Why were these four catchments chosen? Do they have different climatological characteristics that make them interestingly different? A lot of the following analysis focuses on differences between these mesoscale catchments so it is important that the reader understands what these key differences are. Table 1 contained some interesting catchment characteristics but then these were not further explained.

Response: Thank you for the constructive comments. According to available flow records, four upstream catchments which are minimally impacted by human influences were considered in this study. These four catchments ranging in size from 417 km^2 to about 1300 km^2, along with a large difference in elevation and annual precipitation. Meanwhile, the map for rain gauge locations also shows different observation density for them. We added more describe the study catchments in the revised paper.

Performance criteria – The choice of performance criteria needs to be better justified as this has a large impact on the sensitivity of your results.

Response: We agree that model performance depends strongly on the performance criteria used in calibration. In our previous study, we compared the lumped HBV model performance for difference objective functions in a number of catchments on daily scale. Three criteria: (1) the Nash-Sutcliffe (*NS*), (2) Kling-Gupta efficiency (*GK*) that accounts for the water balances and the correlation of observed and simulated discharge series(Gupta et al., 2009), (3) the combination of *NS* and the *NS* of logarithm of the discharge (*NS+LNS*), were used to calibrate the HBV for 15 catchments(Bárdossy et al., 2016). The model parameters calibrated for every catchment were used to simulate the remaining 14 catchments for testing the transferability of parameters. As shown in the figure below, results for different performance criteria differ considerably. The difference of model performance for the performance measures can be explained by different focuses: *NS* is mainly focusing on high flows as it represents the squared difference between the observed and discharge series, *GK* focuses on water balances and good timing, and *NS+LNS* criterion is strongly influenced by low flow events. Model behavior is dependent on how one evaluates the performance of the model. From the matrix we could find that the model performance for different criteria shows similar treads. In this study, we hope to investigate the sensitivity of model to the input variables and sequentially find effective way for increasing accuracy of flood prediction. We pay for attention to high flows, therefore *NS* was selected as objective function to evaluate model performance. We added the discussion of the choice of performance measures in the manuscript.

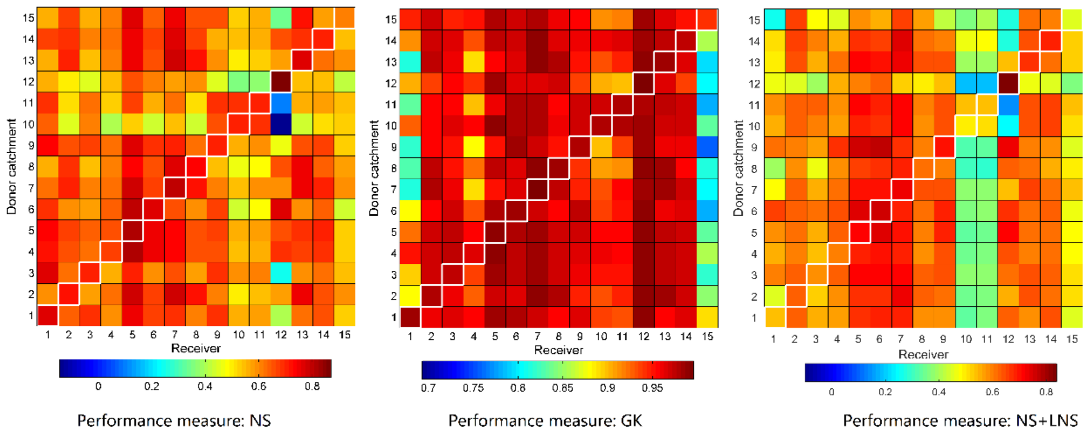


Figure. Color-coded matrices for the model performance of parameter transfer for 15 catchments using three difference performance criteria.

Minor Comments:

Abstract P1 L6 'Two different flavors of HBV' – this doesn't make sense to me. It would be better to just say two different formulations or types.

Response: Thanks for the suggestion. We replaced "Two different flavors of HBV" with "two different types of HBV".

P3 L20 'illustrates the frame of these four datasets' – again, this sentence doesn't make sense to me and needs rewriting.

Response: Revision made. We replaced this sentence with "Figure 3 shows the flow chart of the data collection and process".

Figure 6 As you are focusing on higher flows, I would also find it useful to have another plot (or combined with Figure 6) that focuses on the flow duration curve for flows higher than the 10th percentile of flow.

Response: We appreciate the referee's suggestion, and have added the flow duration curve for flows higher than the 10th percentile of flow (Figure 6(b)) in the revised manuscript.

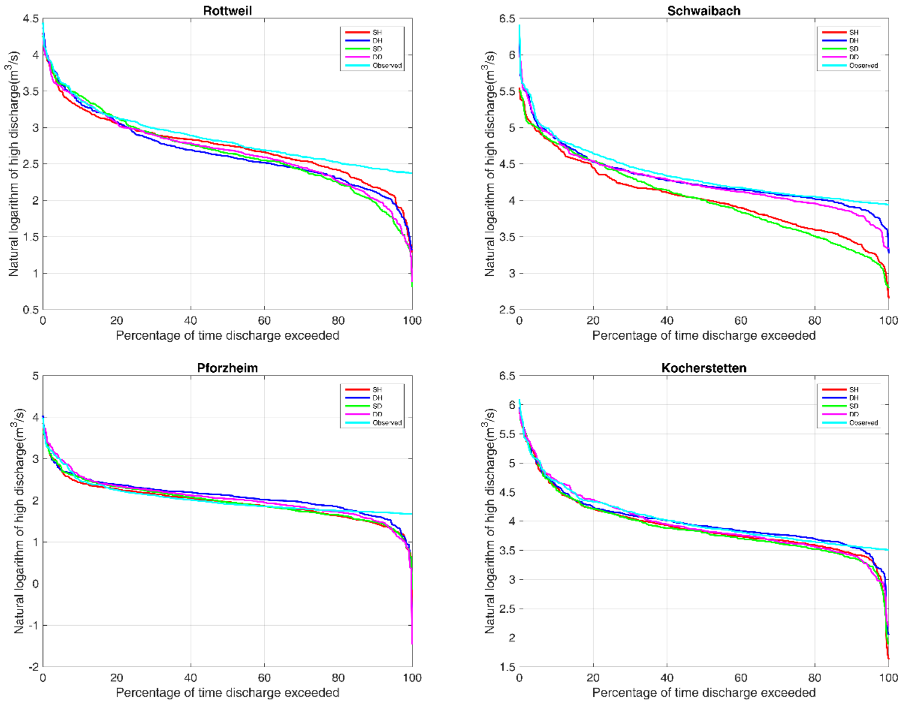


Figure 6(b). Comparison of the flow duration curve for flows higher than the 10th percentile of flow.

Figures 7 -10 need some improvement. The colour scheme needs to be changed in these plots so it is easier for the reader to distinguish between the different catchments. Currently it is difficult to pick out differences between catchments.

Response: Thanks for the suggestion. We have changed the colour of the plots for Figure 7-10 and Figure 12 to make a clear distinction between catchments.

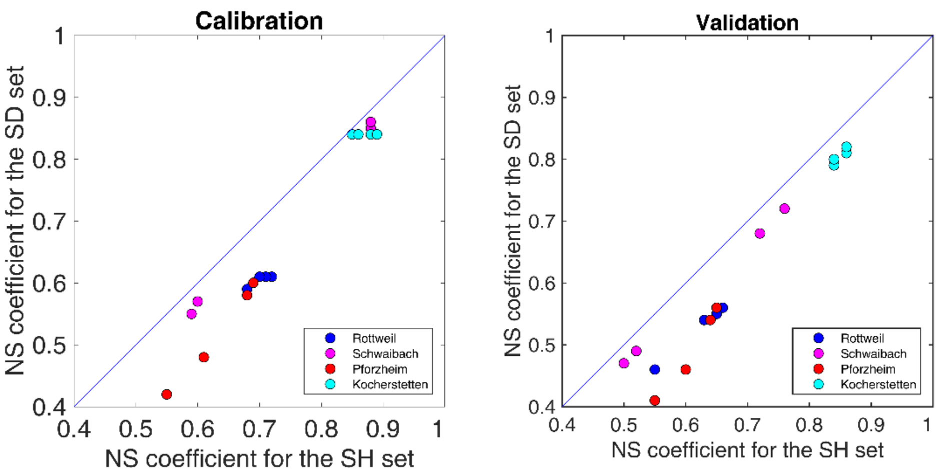


Figure 7. Comparison of NS model performance for using hourly and daily variables as model

input for the SH and SD sets.

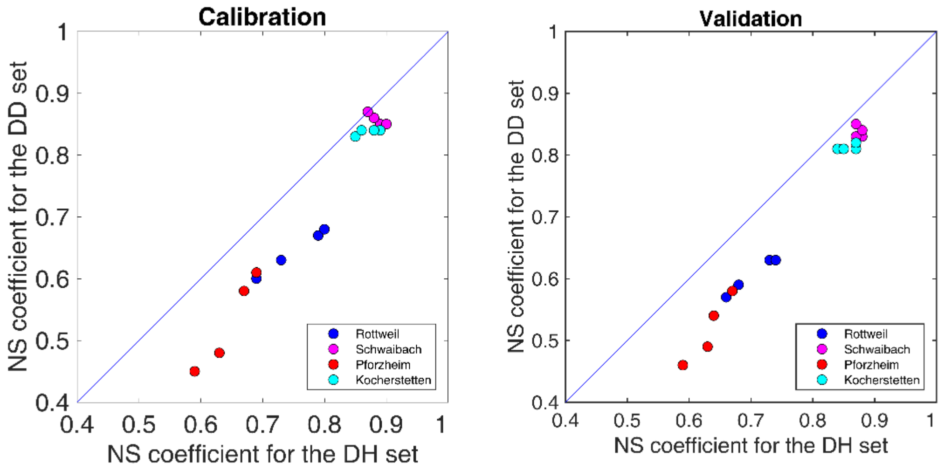


Figure 8. Comparison of NS model performance for using hourly and daily variables as model input for the DH and DD sets.

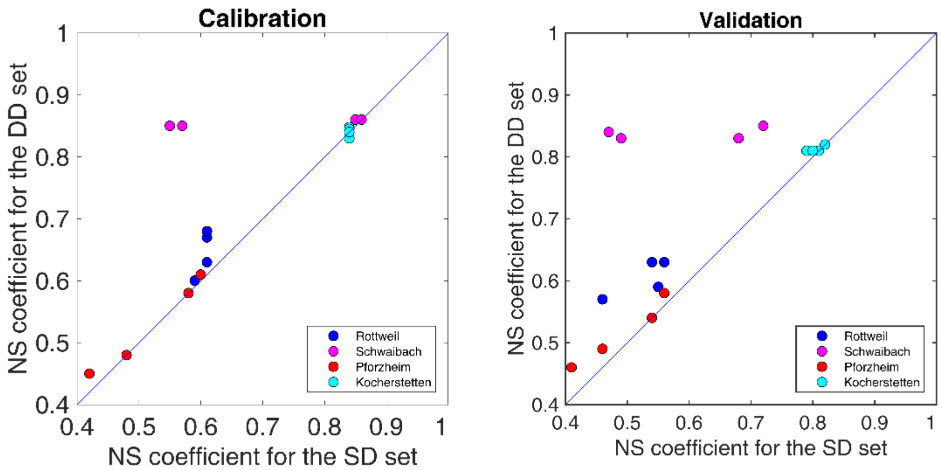


Figure 9. Comparison of model performance for different density of rainfall observation network, models were simulated based on daily time step.

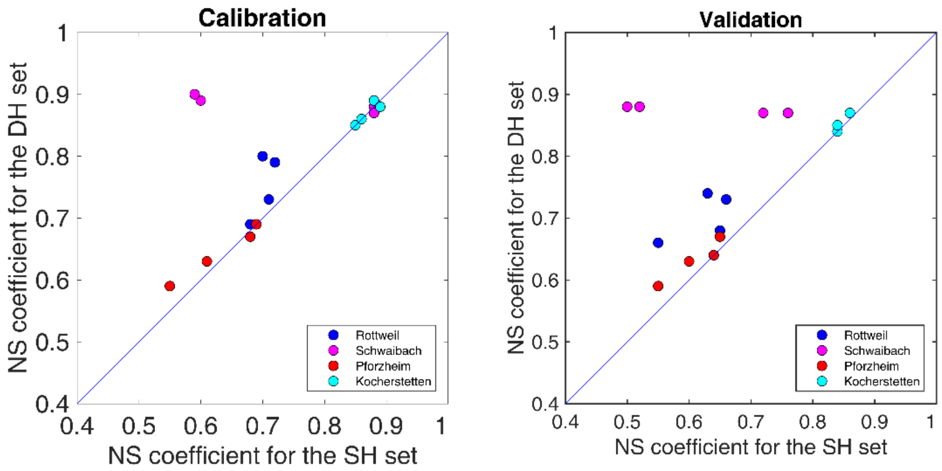


Figure 10. Comparison of model performance for different density of rainfall observation network, models were simulated based on hourly time step.

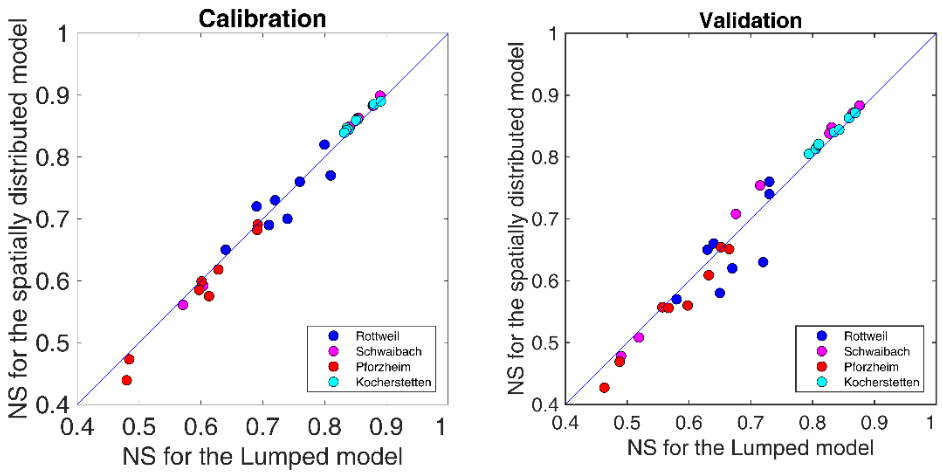


Figure 12. Comparison of model performance for different spatial resolution of model structure.

---

2. Response to Anonymous Referee #2.

- Summary

Overall, this is a very interesting paper that approaches the issues of the combined impact of temporal and spatial resolutions on the efficiency of a hydrological model,

by using both distributed and lumped versions of the same hydrological model, and different densities and time resolutions of precipitation. I find it however unnecessarily complex, the authors should not try to show us everything they have done, they should try to simplify it into a coherent ensemble. I suggest removing the part on the different rainfall densities, and only keeping the densest network (high density daily disaggregated into hourly). This will allow the authors to focus on the spatial and temporal resolution issues. Also, I suggest to widen the scope of the analysis, which only focuses on high flows presently (because of the chosen criterion).

Response: There are two main reasons that we presented the results based on different spatial resolutions in the manuscript. Firstly, results indicate the insensitivity of model performance to different spatial resolutions of rainfall for the study catchments, the increase of spatial resolution improved the simulation insubstantially. Secondly, compared to the idea of increasing spatial resolution of model inputs, which causes the complexity of model structure and parameters, using higher temporal resolution of rainfall by disaggregation method could be an easier and much lower cost way to improve model performance. The authors hope to keep the results based on different spatial resolutions of rainfall to emphasize the effects of disaggregation method in model improvement. In the revised manuscript, we have extended the discussion with the sensitivity analysis of model simulation to the choice of performance criteria.

- Literature issues

I would say that your literature review is quite superficial. Of course, given the considerable increase of published literature, it has become obviously impossible to read everything that is published on a given topic. However, when you aim to publish a paper in a given journal: : : you should perhaps try to look at what has been published there in more detail. It is a little annoying that you seem to ignore a paper that is precisely on the topic you address in your paper:

Lobligeois, F., V. Andréassian, C. Perrin, P. Tabary, & C. Loumagne. 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation on 3620 flood events. Hydrology and Earth System Sciences, 18: 575-594

And this is a pity because when you write that "the increase of spatial resolution improved the performance of the model insubstantially or only marginally for most of the study catchments", this is precisely what Lobligeois et al. find…

Response: We thank the referee for the comments and apology for the ignorance of the references. We rewrote the literature review part with an updated introduction, referring to the ongoing progress of the researches for the sensitivity analysis of model inputs both on temporal and spatial scales. In the revised version, we described in more details about the attempts for improving model performance and the motivation of our paper. We also compared and discussed our idea with previous work on impacts of input variables in hydrological models.

- Vocabulary issues

I understand that you use "pluviometer" for "recording pluviometer / raingage" and "daily station" for "non-recording pluviometer / raingage". This makes your paper difficult to follow.

Response: We replaced "pluviometer" with "sub-daily station" in the revised manuscript.

- Redaction issues

Your conclusion (especially the last paragraph) is difficult to understand. Try to be more explicit.

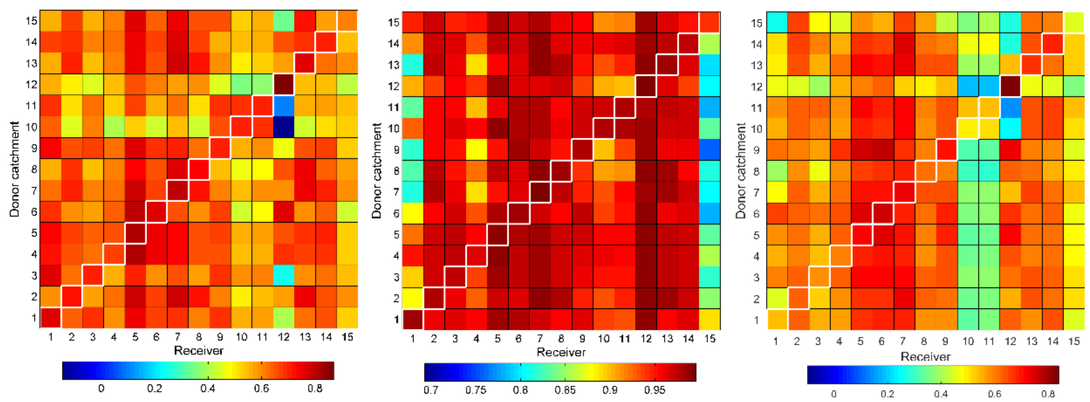
Response: We reorganized our conclusion part in the revised manuscript to make it more understandable.

- Performance criteria

By using the Nash and Sutcliffe criterion on non-transformed flows (instead of, for

example the NS on the square-root or the log or the inverse of flows) you make an explicit choice to focus on high flows only. Why? Could you extend your study by using another transformation in addition?

Response: The aim of this work is to investigate the sensitivity of model to rainfall data and sequentially find effective way for increasing accuracy of flood prediction. We pay for attention to high flows so the Nash-Sutcliffe efficiency was selected as objective function to evaluate model performance. In our previous study, we have compared the lumped HBV model performance for difference objective functions in a number of catchments on daily scale. Three criteria: (1) the Nash-Sutcliffe (*NS*), (2) Kling-Gupta efficiency (*GK*) that accounts for the water balances and the correlation of observed and simulated discharge series(Gupta et al., 2009), (3) the combination of NS and the NS of logarithm of the discharge (*NS+LNS*), were used to evaluate HBV for 15 catchments(Bárdossy et al., 2016). In addition, the model parameters calibrated for every catchment were used to simulate the remaining 14 catchments for testing the transferability of parameters. As shown in the figure below, results for different performance criteria differ considerably. The difference of model performance for the performance measures can be explained by different focuses: *NS* is mainly focusing on high flows as it represents the squared difference between the observed and discharge series, *GK* focuses on water balances and good timing, and N*S+LNS* criterion is strongly influenced by low flow events. Model behavior is dependent on how one evaluates the performance of the model. From the matrix we could also find that the model performance for different criteria shows similar treads. In this study, each calibration process requires 90000 running of HBV model to obtain 10000 best parameter sets. Due to the heavy computation, it is a little bit difficult to extend the study by using some other performance criteria within a short time. We added the discussion of the choice of performance measures in the revised paper.



Performance measure: NS　Performance measure: GK　Performance measure: NS+LNS

Figure. Color-coded matrices for the model performance of parameter transfer for 15 catchments using three difference performance criteria.

- Interception

I would like to know how the interception process is accounted for in your version of HBV? This is important for your comparison, because the simple solutions that work well at the daily time step (i.e. neutralization of daily rainfall by daily pot. Evaporation) may not work as well at the hourly time step, which may require an interception store.

Response: In our model, the interception process is consisted in evapotranspiration. The approach of Penman equation (Penman, 1948) is used to estimate the daily potential evapotranspiration according to the long-term monthly mean air temperature and long-term monthly average potential evapotranspiration using observed daily average temperature. Due to the limitation of observed hourly temperature, air temperature and potential evapotranspiration were assumed to be constant over the whole day in our study. The actual evapotranspiration is calculated based on the available water in soil and permanent wilting point based on the

- Typos

There are a few typos in the paper. Please make a careful check.

Response: We sincerely apologize and have carefully reviewed our manuscript.

References:

Bárdossy, A., Huang, Y., and Wagener, T.: Simultaneous calibration of hydrological models in geographical space, Hydrology & Earth System Sciences Discussions, 12, 11223-11268, 2016.

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.

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.

Sensitivity of hydrological model to the temporal and spatial resolutions of rainfall input

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 the most important input for rainfall-runoff models, precipitation is usually observed 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 a higher temporal and spatial resolution of precipitation could provide an improvement in model performance. Four different gridded hourly and daily precipitation datasets, with a spatial resolution of 1×1 km^2 for the Baden-Württemberg state of Germany, were constructed using a combination of data from a dense network of daily rainfall stations and a less dense network of ~~pluviometers with high temporal-resolution rainfall observations~~sub-daily stations. Two different ~~flavors~~types of HBV models with different model structures, lumped and spatially distributed, were used to ~~test~~investigate the sensitivity of model performance on the spatial ~~resolution~~variability of precipitation. For four selected mesoscale catchments, ~~located at the upstream region of Baden-Württemberg,~~ these four precipitation datasets were used to simulate the daily discharges using both lumped and semi-distributed HBV models. Different possibilities of improving the accuracy of daily streamflow prediction were investigated. Three main results were obtained from this study: (1) a higher temporal resolution of precipitation improved the model performance if the observation density was high; (2) a combination of observed high temporal-resolution observations with ~~disaggregated~~disaggregate daily precipitation leads to a further improvement in the model performance; (3) for the present research, the increase of spatial resolution improved the performance of the model insubstantially or only marginally for most of the study catchments.

1 Introduction

~~Conceptual hydrological models have been developed to represent dynamic response of a particular catchment resulting from meteorological driving forces (Hundecha et al., 2008). Among meteorological variables, precipitation, which is traditionally measured using rain gauges, has a direct and crucial impact on dynamic response of a catchment (Obled et al., 1994; Ly et al., 2013). However, uncertainty in capturing the variability of precipitation by the rain gauges or wireless telemetering constitutes a significant source of uncertainty for hydrological modeling (Berne et al., 2004). Previous studies have shown that hydrological models are sensitive to the observation network density and data quality (Singh, 1997; Kobold and Brilly, 2006; Bardossy and Das, 2008; Xu et al.,~~

2013). Therefore, the precipitation input should be as accurate as possible to achieve better rainfall-runoff simulation and model parameter estimation (Cole and Moore, 2008; Ficchi et al., 2016).

Rainfall is one of the most important driving forces in hydrological modeling and produces a direct impact on catchment runoff response (Obled et al., 1994; Ly et al., 2013). In general, rainfall is measured by standard rain gauges or wireless telemetering pluviometers over a set period of time (e.g. daily, sub-daily). The instrument measuring error 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 is one of the main sources of uncertainty in model simulation and flood forecasting(Beven, 1998; Berne et al., 2004). 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.

Many research efforts have been carried out in the recent years for interpolating spatially distributed rainfall datasets (Goovaerts, 2000; Jeffrey et al., 2001; Hofierka et al., 2002; Haylock et al., 2008; Ly et al., 2013), as well as for the sub-daily disaggregation of daily rainfall (Parkes et al., 2013; Bardossy and Pegram, 2016). These approaches can potentially improve the quality and resolution of the precipitation data that are used as input for rainfall-runoff models, thereby reducing the uncertainty of hydrological models. By design, most of the hydrological models are flexible and can be easily adjusted to different time steps of input datasets. Hydrological models are normally classified as lumped or distributed, depending on the degree of spatial discretization when describing the catchment (Ly et al., 2013). Bruneau et al. (1995) indicated the temporal and spatial resolutions used for the inputs of the hydrological model have an important influence on the model performance. Kobold and Brilly (2006) suggested that calibrating hydrological models with sub-daily time steps can significant improve flood forecasting. Das et al. (2008) used different model structures to simulate daily runoff in the region of central Europe and showed that semi-distributed model structure could outperform lumped model structure.

In recent decades, a great amount of efforts have been put on investigating the influence of rainfall spatial variability in hydrological models. Different interpolation methods have been used to obtain the spatial distribution structure 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 has a significant influence on the timing and shape of hydrograph, while the temporal variability shows a certain impact to the peak of flood wave. Kobold and Brilly (2006) used a different number of rain gauge stations to derive areal rainfall and quantitatively assessed the uncertainty of rainfall inputs using HBV model in hourly time step. They found that the error in precipitation may lead to even greater error in the peak of flood. Bardossy and Das (2008) also invested the impact of spatial variability of rainfall by varying the distribution of the rain gauge network. They found that the transferability of model parameters calibrated based on spare and density precipitation information is 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 that the lack of spatial information is responsible for the low efficiency of distributed model. Xu et al. (2013) indicated that the increase of raingauge network density gradually improves the model performance up to some threshold, but no apparent improvement was observed when the number of rain gauges exceeded the 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. Emmanuel et al. (2015) proposed rainfall variability indexes to carefully evaluate the influence of rainfall spatial variability and implemented this approach in the model simulation for Cevennes catchment in France (Emmanuel et al., 2017). They found that higher spatial resolution of rainfall could achieve better model performance. However, the increasing of spatial resolution in model simulation leads to considerable complexity of model structure and requires for much more data than lumped version.

Simultaneously, the rainfall-runoff response of a catchment are strongly impacted by the temporal variability of rainfall (Bárdossy and Pegram, 2016). The high temporal resolution rainfall data is typically measured by pluviometer stations (wireless instruments recording at sub-daily intervals, be called sub-daily data in the following), which faces the problem of poor data quality caused by equipment malfunction or misreading. Compared with sub-daily data, the daily rainfall data are more reliable and plentiful. Disaggregating daily into sub-daily values 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 the resampling method is the best way for rainfall disaggregating. Bárdossy and Pegram (2016) used Gaussian Copula-based model for disaggregating daily data to infill the gap of pluviometer data, and they found that this conditional disaggregation of precipitation is reliable and applicable in various regions. Breinl and Di Baldassarre (2019) applied a spatial method of fragments to disaggregate daily precipitation into hourly values. Although a 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 significant improve the accuracy of flood forecasting.

Furthermore, certain studies focus on both the spatial and temporal resolution of rainfall. Bruneau et al. (1995) indicated the temporal and spatial resolutions used for the inputs of the hydrological model possess a considerable influence on model efficiency and parameters. Booij (2002) assuredly found that influence of model spatial resolution is indeed greater than rainfall temporal variability on the simulation of extreme flow. Meselhe et al. (2009) indicated that the physically based model is more sensitive to the spatial and temporal resolution of rainfall than the conceptual model. Zhu et al. (2018) found that the spatial variability of rainfall is much more sensitive for catchments larger than 2000km^2 under dry soil condition; while flood in the small catchments is controlled by the temporal variability of rainfall. Since a vast number of efforts had been made to improve the spatial or temporal resolution of rainfall, it is important to focus on a quantitative analysis 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.

The ultimate aim of this study is to undoubtedly gain more firsthand knowledge on the dependency of hydrological model performance on the precipitation data. 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 differentdistinctive model structures. The possibility of improving model performance on a daily scale was properly discussed. The manuscript is organized as follows: the introduction, followed by section 2, which describes the study area and the precipitation datasets used in this research. In section 3, the hydrological model and the calibration framework used in this research are explained, while section 4 presents the results and discussion of this work. The conclusions and outlook are provided in section 5.

2 Study area and hydrometeorlogical datasets

This study was tested in a semi-humid region in the Baden-Württemberg state of Germany (Figure 1) that characterized by temperate monsoon climate. Elevations of this state range from 85 m to 1 493 m 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 is evenly distributed through the year. However, its seasonality shows a weak trend. The monthly rainfall reaches its peak in June, whereas the month of October shows the least precipitation The meteorological observations used in this study was provided by the German Weather Service (DWD). Daily air temperature required for the rainfall-runoff model was interpolated on a 1×1 km^2 grid from the observations using the algorithm of External Drift Kriging (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 data from a dense network of daily precipitation stations (62 km^2/station in 1991) and from a less dense network of pluviometerssub-daily stations (144 km^2/station in 1991) with high resolution precipitation observations were used for this study. All available data from the time period 1991-2010 was considered. The number of available daily stations and pluviometerssub-daily stations varies according to different time period. Figure 2 illustrates the number of available observation locations 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 only 30 pluviometerssub-daily stations. The total number of daily stations decreased dramatically to 250 around 2003 and remained constant for the subsequent years. The number of pluviometerssub-daily stations kept increasing throughout the whole period and experienced a sharp increase from 100 to 200 in the year 2005. The following different precipitation datasets were created according to the available observed data:

1. High resolution observed precipitation was aggregated to hourly time steps and interpolated subsequently to a 1×1 km^2 grid using the ordinary Kriging (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 referred as Sparse Hourly (SH) set.

2. Observed daily precipitation 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. The variogram was based on the cross-correlations of the daily time series. This set will be referred as Dense Daily (DD) set.

3. High resolution precipitation 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 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 as Dense Hourly (DH) set.

Figure 3 illustrates the frame of these four different datasets 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 disaggregateddisaggregate daily 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 of (SD, SH) and (DD, DH) provide information on the effect of temporal resolution. While comparisons between (SD, DD) and (SH, DH), provide information on the influence of the rainfall observation network density.

Four mesoscale catchments (Figure 1), namely Rottweil, Schwaibach, Pforzheim and Kocherstetten, were selected from the upstream region for testing the sensitivity of model performance to 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 in size from 417km^2 to about 1300km^2 , along with a large difference in elevation and annual precipitation. It can be seen clearly from the map that these four catchments have different rain gauge density, the Schwaibach catchment, which located in the mountain area with various elevations(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 in the 1970s at the Swedish Meteorological and Hydrological Institute (SMHI) (Bergström and Forsman, 1973). Due to its simplicity, low demand of inputs and few model parameters, HBV model has been a preferred model for rainfall-runoff simulation and flood forecasting. Figure 4 represents the structure diagram of HBV model (Singh, 2010). In general, three main modules are included in HBV model, namely snow routine, soil moisture routine and runoff routine (Hartmann, 2007; Singh, 2010).

First of all, the snow accumulation and melt process is estimated by the relatively simple degree-day method (Rango and Martinec, 1995) using two parameters: degree day factor (*DD*) and threshold temperature for snow/rain (*TT*) (as shown in Equation 1). In this method, 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 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 the ratio of direct runoff to effective precipitation ($\frac{\Delta Q}{\Delta P}$) can be estimated by:

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

where *SM* denotes the actual soil moisture and $Beta$ determines the proportion of effective precipitation contributing to runoff at a 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 potential evapotranspiration (PE_M) (Penman, 1948):

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

Here 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 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 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 were applied: lumped HBV and spatially distributed HBV, respectively. In the lumped model, precipitation, temperature and potential evapotranspiration were supposed to be equally distributed among the catchment and all the processes were calculated for the whole catchment. Previous studies have indicated that the altitude is an important reason for the spatial differentiation of meteorological elements, such as temperature, precipitation, evapotranspiration and snow melt. Therefore, the spatially distributed HBV model was constructed to separated the whole catchment into several zones based on topographic elevation. The 1×1 km^2 grid based precipitation and temperature data were computed averagely according to elevation zone and used as inputs for model simulation. In the spatially distributed model, the snowmelt and soil moisture modules related parameters can be adjusted differently for each elevation zone. The parameters controlling the runoff response processes were estimated for the whole catchment similarly to the lumped model (Das et al., 2008).

There are 15 parameters describing the HBV model, where only 9 parameters were selected for calibration in this study. Table 2 lists the initial upper and lower limit of 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. 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 model performance.

3.2 Performance criteria

Previous study have shown that model performance strongly depends on the selection of performance criteria (Gupta et al., 2009). The simulated result and model parameters using different objective functions differ considerably as they have different focus (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, the Nash-Sutcliffe (NS) efficiency coefficient (Nash and Sutcliffe, 1970) was used in this study to assess the model performance based on observed discharge. NS efficiency is one of the most widely used performance criteria in model simulation. It focus on high flow as it evaluates the squared difference between simulated and measured streamflow. NS efficiency can be calculated using the following equation:

~~In this study, the Nash-Sutcliffe (NS) efficiency coefficient (Nash and Sutcliffe, 1997) between the observed and simulated streamflow was used to access the model performance:~~

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

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

Meanwhile, the Mean Square Error (MSE) of the flow for the time period that the observed discharge is greater than or equal to 90% high flow valuehigher than the 10th percentile of flow was calculated 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 $Q_o(i)$ and $Q_m(i)$ are the observed and modeled discharges when the observed discharge is greater than or equal to 90% high flow valuehigher than the 10th percentile of flow.

3.3 Model calibration experiments

A split sample calibration methodology has been applied in this study to separate the whole data series into two equal periods: 1991-2000 and 2001-2010. Model calibration was carried out for both time periods and a cross-validation analysis was performed as well. For each calibration run, the first water year data was used as 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 as input variables for model calibration and validation. We also assessed the effects of the temporal and spatial resolutions of the inputs on the model performance in terms of Nash-Sutcliffe efficiency and the mean square error 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 100 m was taken to divide the whole study catchment into several 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 if the combination of daily scale model and hourly scale model leads to a better prediction in streamflow. It is interesting to investigate the similarities of different temporal resolution. Therefore, the common calibration tragedy was proposed in this study to calibrate the daily scale model and hourly scale model simultaneously. This kind of approach is expected to identify robust model parameters for the application of model in different temporal resolutions. 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 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 quality of the rainfall products was assessed and compared for 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 sets SD and DD for both calibration decades (Figure 5). It can be seen clearly from the figures that the interpolated precipitation datasets display some difference for all study catchments. The asymmetry of the scatterplots is fairly obvious for the first decade (1991-2000). In general, the DD dataset leads to higher value than the SD dataset. The reason behind this is mainly because the low density of ~~pluviometers~~sub-daily observations during the period of 1991-2000 leads to big errors in the spatial interpolation of rainfall. This is especially the case for Schwaibach catchment which varies strongly in geographical elevation (from 190 m a.s.l. to 1028 m a.s.l.). For the period 2001-2010, the SD and DD sets become similar in magnitude along with the increasing of available sub-daily observations.

4.2 Calibration and validation model performance

As designed 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 for calibration and cross-validation. This leads to 16 calibration runs and 16 validation runs for every catchment. As mention before, each simulation could obtain10 000 parameter sets with similar model performance. To make it simple, we took the mean value of the corresponding 10 000 model performances to represent the model efficiency.

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 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 for different data periods are also different. For the Schwaibach and Pforzheim catchments, the calibrated model performance for the time period 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 ~~raingauge~~rain gauge density inside or nearby 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 for both calibration and validation; the loss in NS coefficient is about 0.3 when compared to the corresponding results of the sets DH and DD. This indicates that systematic interpolated precipitation errors have critical influence on model calibration.

The flexibility of model in flood prediction is analyzed with the behavior of high flow. Tables 5 and 6 list the mean square error of ~~90% high flow~~flows higher than the 10th percentile of flow for 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 performance corresponding to the temporal resolution

Firstly, the model performance of different temporal resolution was compared for all datasets and model structures. For the pairwise comparison, all the conditions are the same in the model except for the 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 that were interpolated using only high-resolution precipitation observations (SH, SD). Figure 9 compares the corresponding results for the rainfall datasets that incorporated observed daily value with high-resolution observations (DH, DD). The result shows that all the scatters are laying below the diagonal for the different level of observation density. For both calibration and validation periods, the simulations using hourly data as model input outperform the one that based on the daily resolution. For the dataset with low observation network density, the average NS value of set SH is about 0.73 for calibration period and 0.68 for 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 of set DD is 0.72 and 0.69, respectively. The hourly scale model performs better than the daily model indicating that the dynamic runoff of catchment could be better simulated with a higher temporal resolution of input variables. According to the distances from the diagonal to the scatter plots, we could find that the difference in model performance for different temporal resolution 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 performance corresponding to observation density

Results also indicate that rainfall data 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. It shows obviously from the location of points that the simulated model performance of DD set is generally better than the result of SD set for both calibration and validation periods. The average NS model performance of DD set over all simulations is about 0.71 while the value for SD set is 0.64. The model performance for the hourly based simulation shows similar trend as the model performance for the daily time step. As shown in Figure 11, the model calibration of DH set outperform the result of SH set. The results demonstrate that the high observation density could lead to considerable improvement of model performance for both daily and hourly time scales.

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

4.5 Comparison of the performance corresponding to the spatial resolution

The 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. 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 the lumped ones, especially for the catchments having high NS coefficient. Despite the increase in model performance being only marginal. However, for the catchments with relatively weak model performance, the lumped model could even lead to slightly better performance than the semi-distributed model structure, especially for the validation period that 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 number of parameters to be identified 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. The results support the findings of Das et al. (2008) that distributed model structures does not significantly improve model performance.

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

4.6 Common calibration of models with different temporal resolutions

As shown before, the combination of hourly observations and daily observations lead to the improvement of data quality as the sets DH and DD show better model performance than the sets SH and SD. Furthermore, common calibration of 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) should be converted according to the temporal resolution of model. The common calibration was performed for two decades separately, and the 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 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 that the common parameters outperform the individual ones for about half of the all simulations. It indicates that common calibrated parameters based on different time steps could be a feasible approach for increasing the temporal transferability of models. The reason for the robustness of common parameters might be that common calibration tragedy could provide more information for identifying model parameters.

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

5 Conclusions and outlook

~~This paper investigated the impact of the observed precipitation data in model simulation and parameter estimation. The sensitively of model performance to different temporal and spatial resolutions of input variables were also tested in this study. Two different model structure, lumped and spatially distributed HBV models, were used to simulate daily runoff using precipitation data sets with different time resolutions and interpolated using different observation network density. The models were applied to four upstream catchments using NS coefficient as objective function and the mean square error for the high flow was also assessed. The common calibration scenario was proposed to calibrate the model with different time scale simultaneously to provide robust model parameters.~~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 lead to any improvement of model performance. Both the lumped HBV and spatially distributed HBV models were applied to simulate the daily runoff for four mesoscale catchments driven by four different types of precipitation 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 were evaluated using Nash-Sutcliffe efficiency and the mean squared error of flows higher than the 10th percentile of flow. The sensitivity of model to the temporal and spatial resolutions of rainfall was compared. In additional, the common calibration approach was proposed to calibrate the models with different time steps simultaneously for seeking robust model parameters.

~~The calibration results indicate that rainfall data quality has a significant impact on model performance. Interpolation of hourly precipitation using disaggregated daily value as additional information could potentially enhances the quality of the data and reduces the uncertainty of the model inputs. The result shows that higher temporal resolution could significantly improved the model performance if the observation density was high. A combination of observed high-resolution observations with disaggregated daily precipitation leads to a further improvement. For the present study, the lumped and spatial distributed model structures perform very similar indicating that higher model resolution does not or only marginally improve the model performance.~~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 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 approach provides an effective way for increasing the temporal resolution of rainfall and the performance of model simulation. However, the lumped and spatially distributed HBV model perform very similarly, indicating that higher model resolution does not or only marginally improve the model performance for the study catchments. The result supports 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 relatively uniform precipitation.

A great amounts of efforts had been made to improve the performance of rainfall-runoff model in recent years. The results of this study suggested that higher temporal resolution of inputs always outperform the lower ones, an effective data disaggregation could lead to an improvement of the model performance. Results also indicated that higher spatial resolution of model, which cause the complexity of model structure and parameters, do not always enhance the model performance. Compared with the idea of increasing the spatial resolution by distributed model, increasing the temporal resolution of model inputs by disaggregation method could be an easier and much lower cost way to improve model performance.As discussed 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 that data disaggregation approach could 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 inputs with disaggregation method could be an easier and more efficient to improve model performance, compared with increasing the model spatial resolution that comes at a cost of increasing the complexity of model structure and parameters.

In this study, all the hourly model outputs were aggregated into daily and only the daily streamflow was involved in the evaluation of model performance. As the daily-based rainfall-runoff response of a catchment is mostly dominated by rainfall amount and actual evapotranspiration, we believe that the variability in precipitation have rather strong impacts on the smaller temporal scales, such as the hourly response of discharge. Meanwhile, in the spatially distributed model, the subcatchments were separated only based on the topographic elevation, which might be not enough to represent the full spatial variability. Moreover, the impacts of precipitation to the hourly response and the full distributed model structure could be considered as the next phase of work.This study focuses on high flows and uses only the Nash-Sutcliffe efficiency as the objective function to investigate the model sensitivity. As model performance highly depends on the selection of objective functions, the model sensitivity can be different if using different performance criteria. 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 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. Special thanks to 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.	Streamgauge name	Longitude (^{o}E)	Latitude (^{o}N)	Area (km^2)	Elevation (m)	Annual precipitation (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	Precipitation data set	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 data set	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 ~~90% high flow~~flow higher than the 10th percentile of flow for the lumped HBV model.

Catchment	Precipitation data set	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 ~~90% high flow~~flow higher than the 10th percentile of flow for the distributed HBV model.

Catchment	Precipitation data set	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

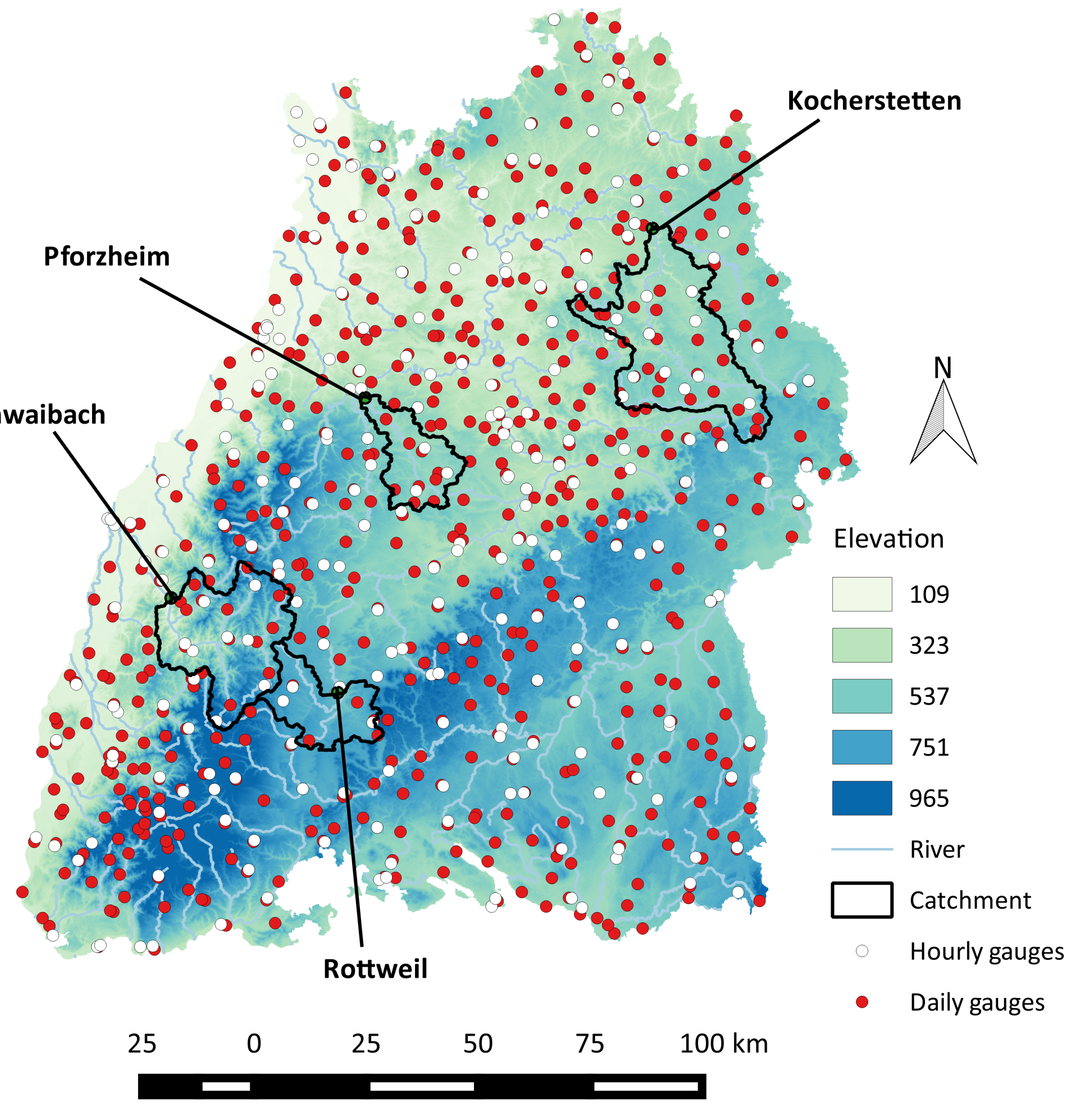


Figure 1. Locations of the pluviometers(hourly) and daily rain gauges in Baden-Württemberg and the four selected catchments.

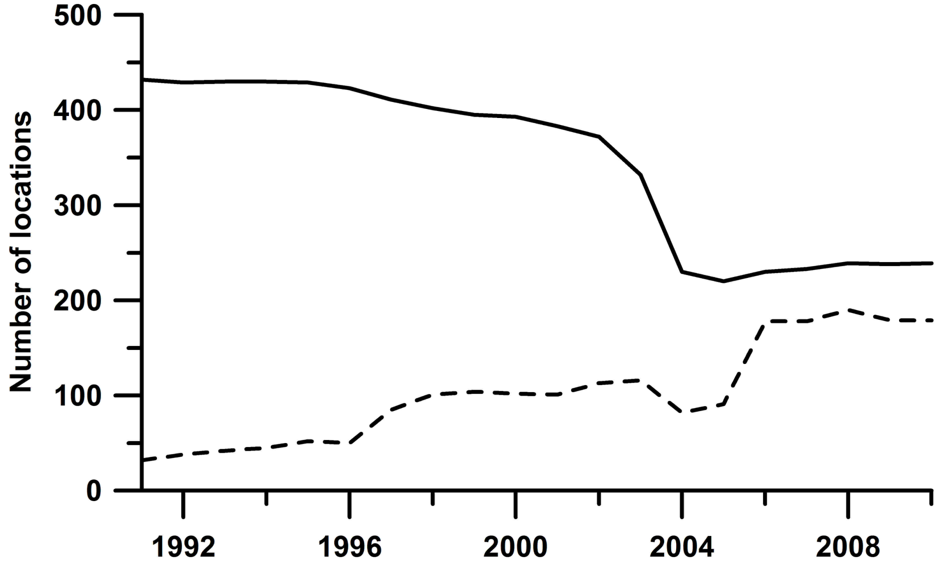


Figure 2. The number of available observation locations. Daily stations - solid line, ~~pluviometers~~Sub-daily stations - dashed line.

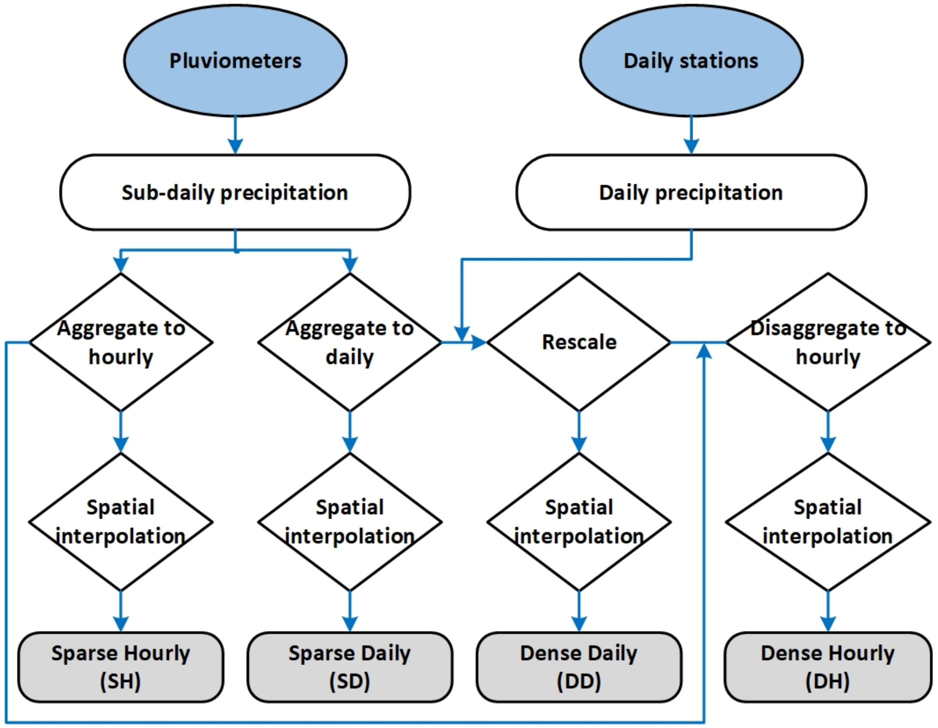


Figure 3. Schematic representation of four different precipitation data sets.

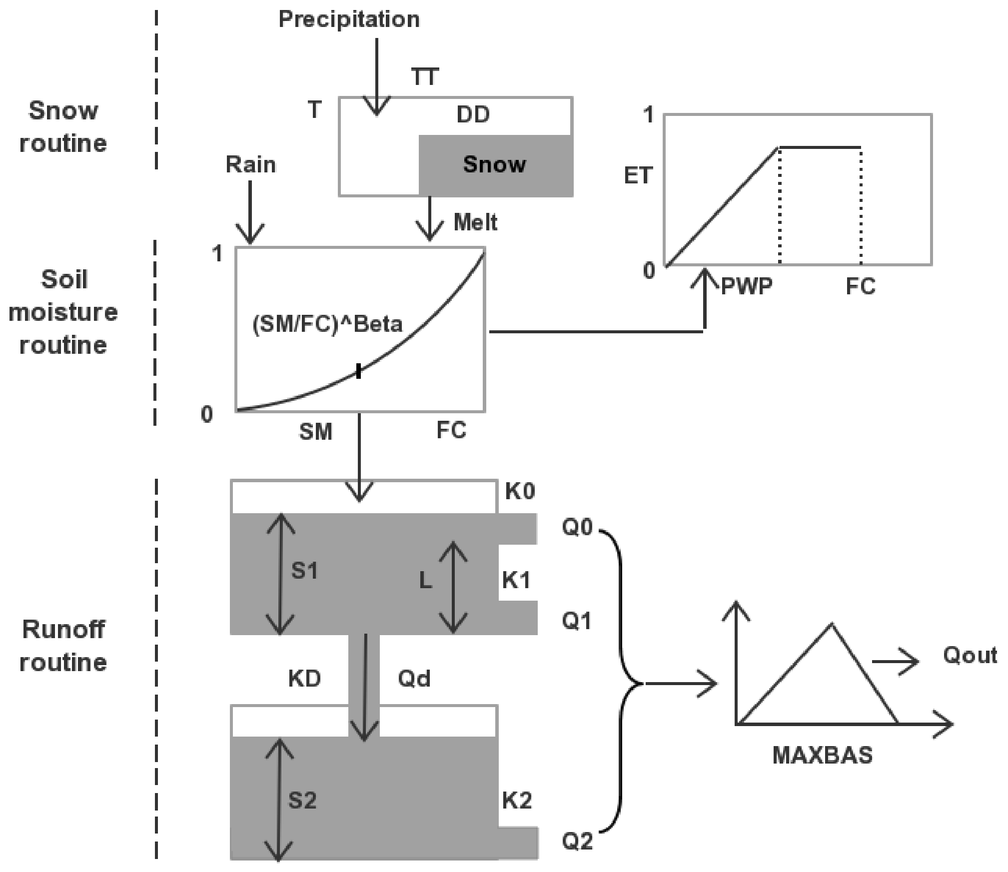


Figure 4. Schematic representation of HBV model.

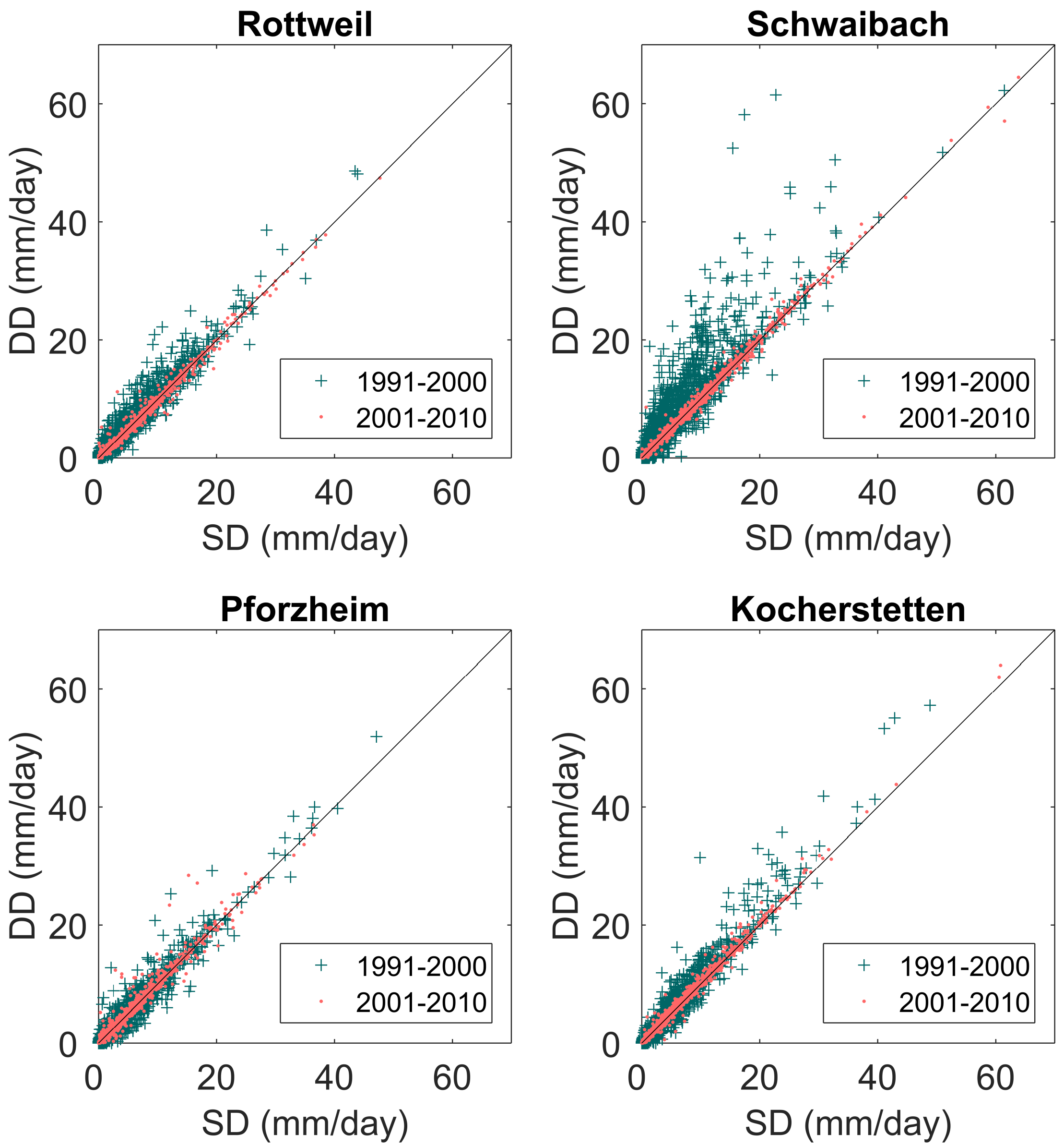


Figure 5. Comparison of the daily precipitation data that interpolated using different observation network density.

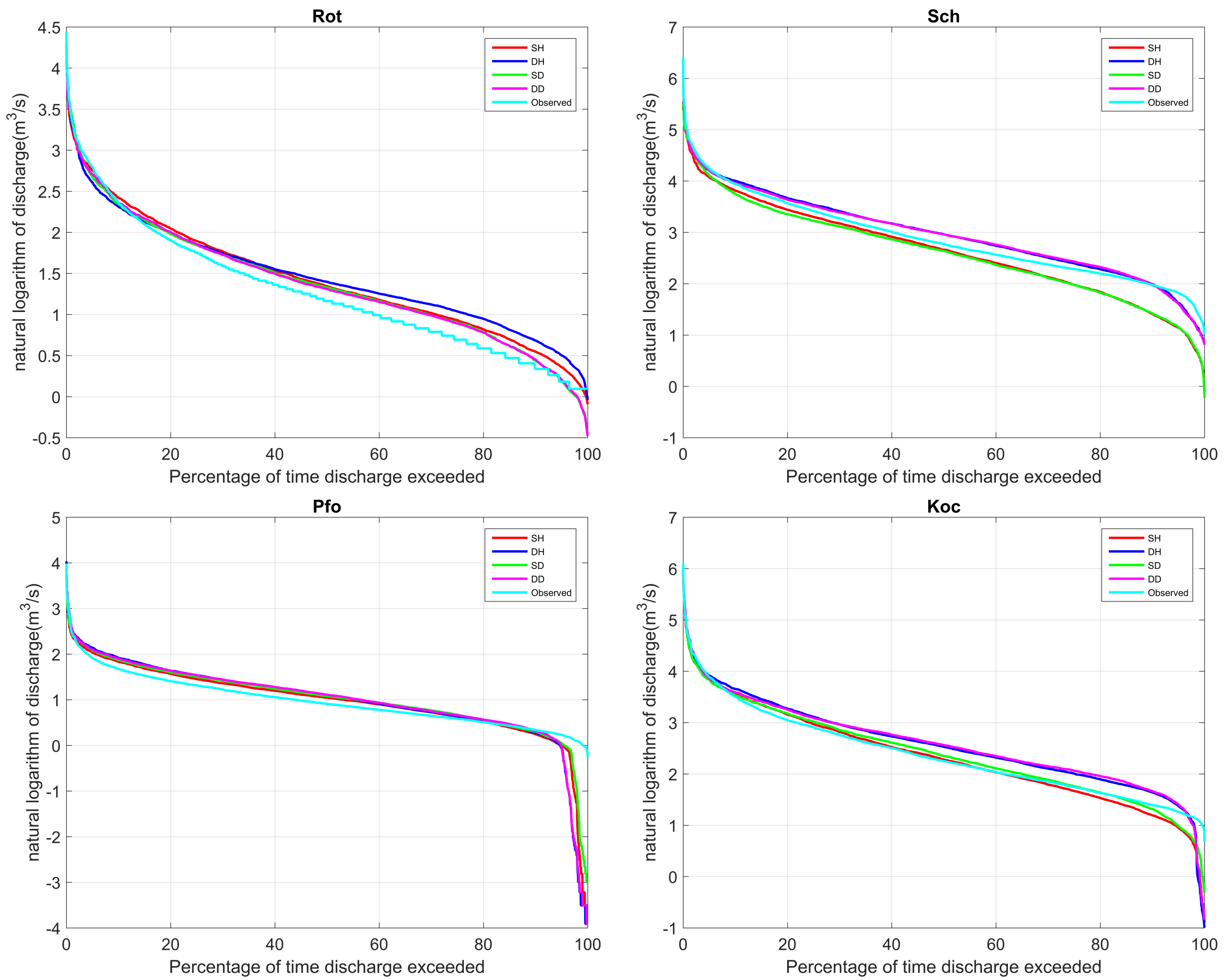


Figure 6. Comparison of the simulated flow duration curve.

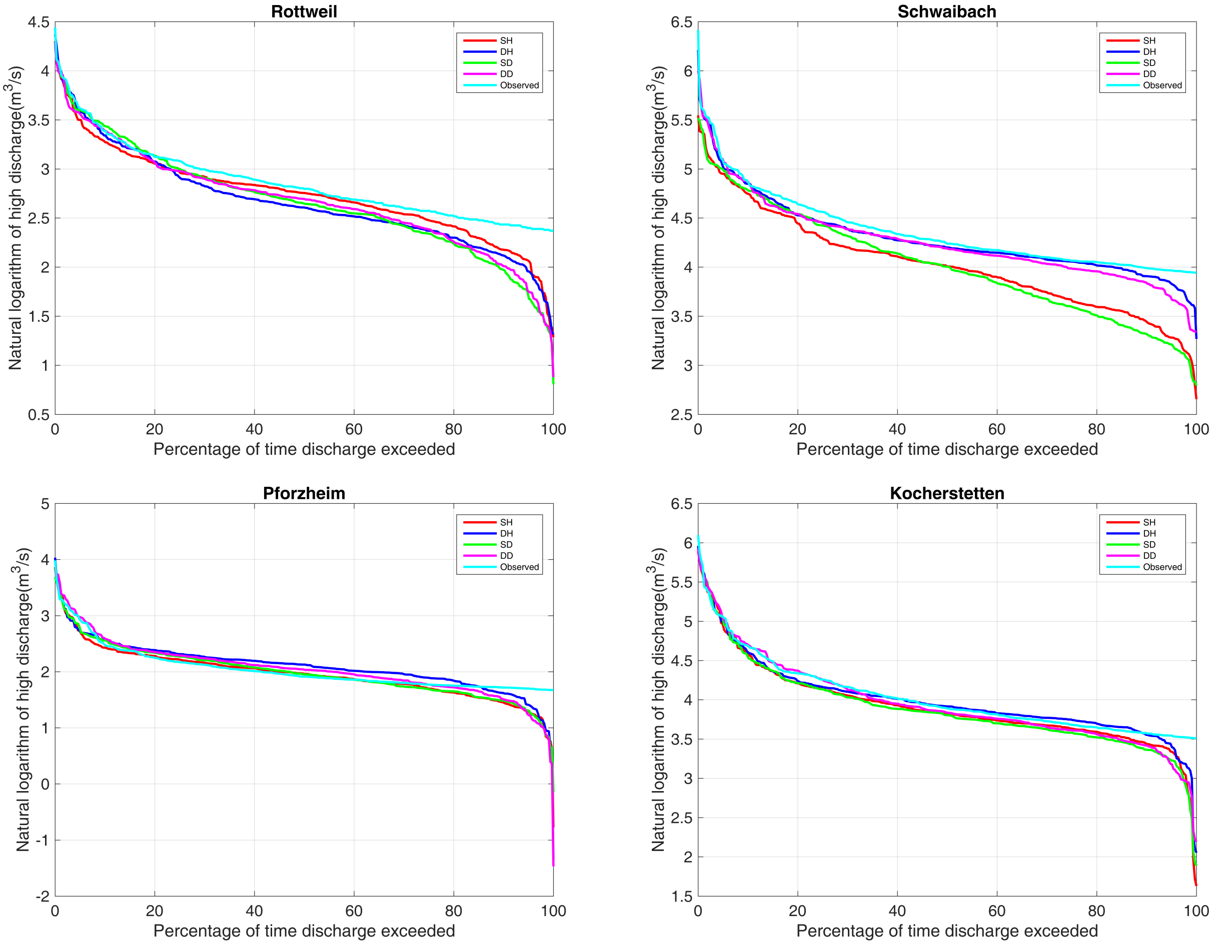


Figure 7. Comparison of the simulated flow duration curve for flows higher than the 10th percentile of flow.

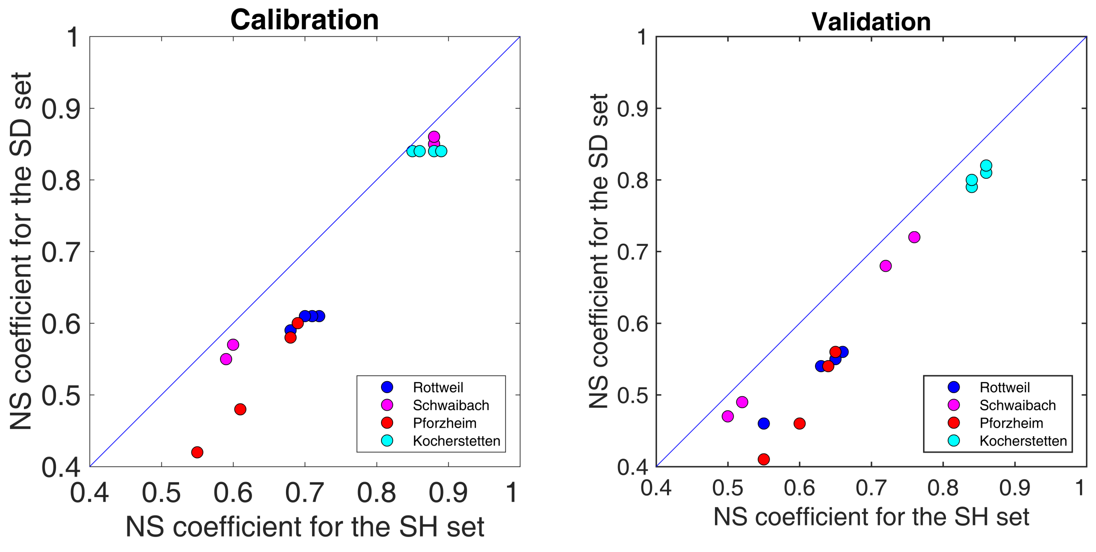


Figure 8. Comparison of NS model performance for using hourly and daily variables as model input for the SH and SD sets.

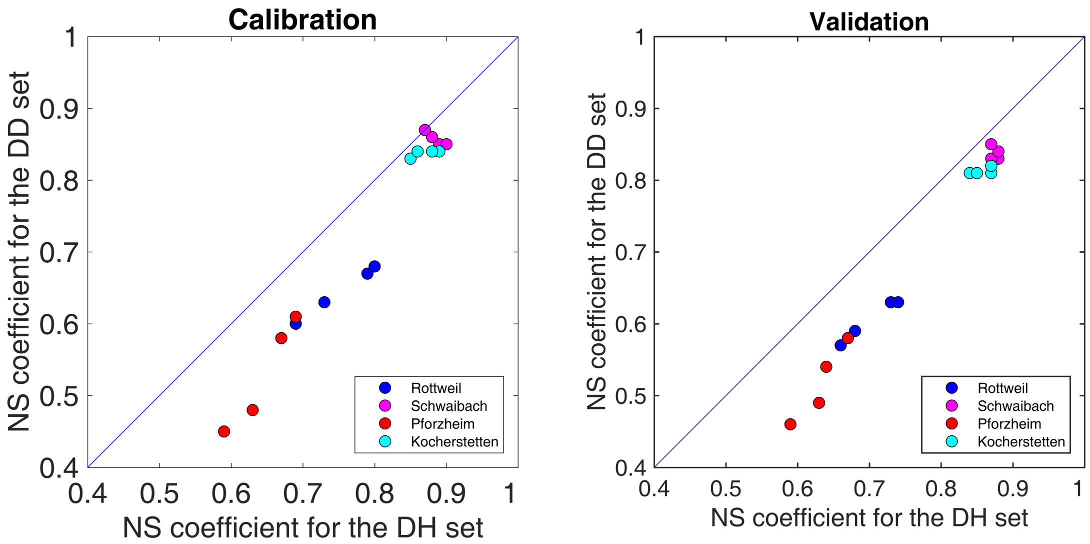


Figure 9. Comparison of NS model performance for using hourly and daily variables as model input for the DH and DD sets.

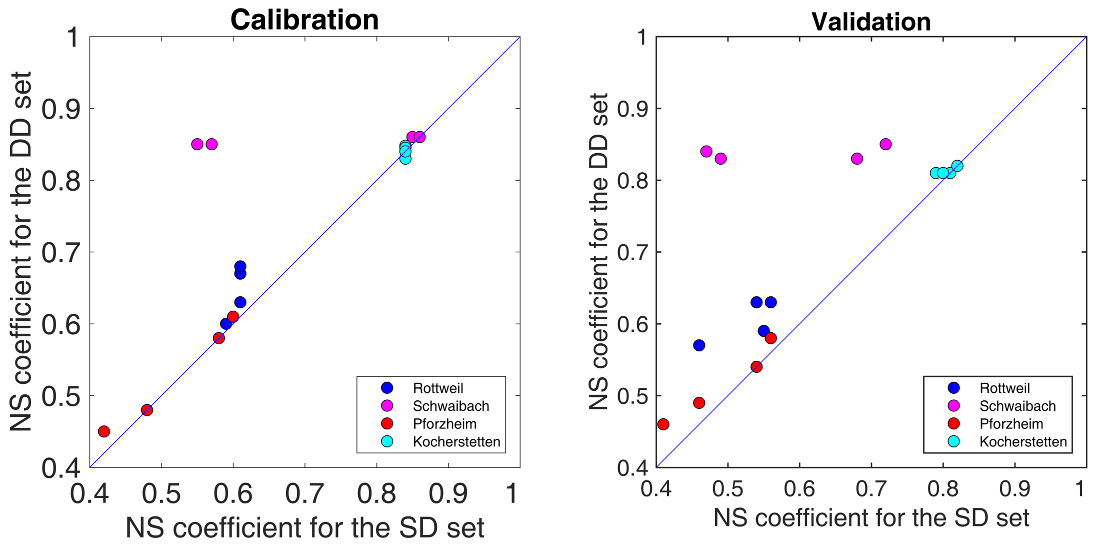


Figure 10. Comparison of model performance for different density of rainfall observation network, models were simulated based on daily time step.

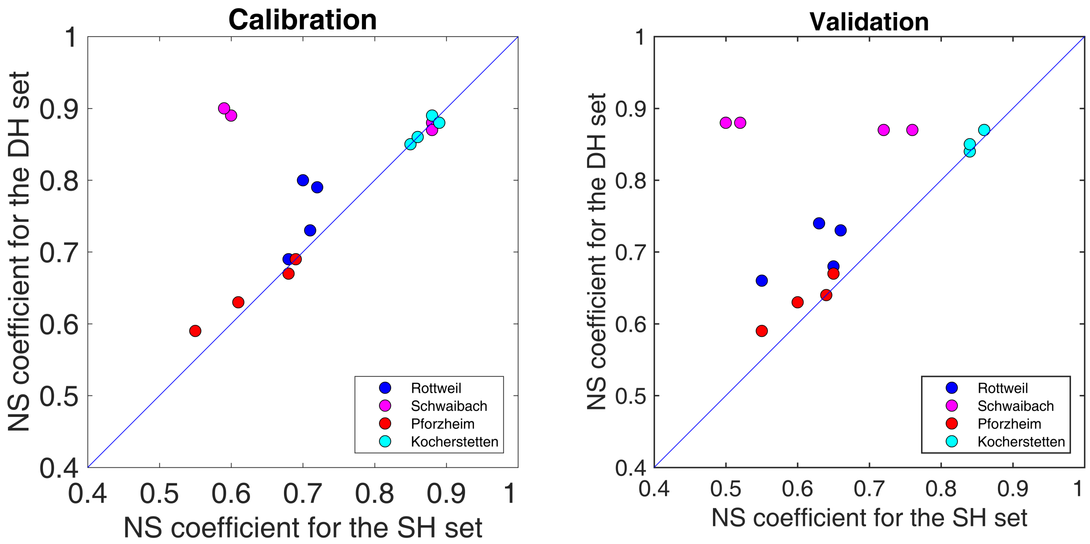


Figure 11. Comparison of model performance for different density of rainfall observation network, models were simulated based on hourly time step.

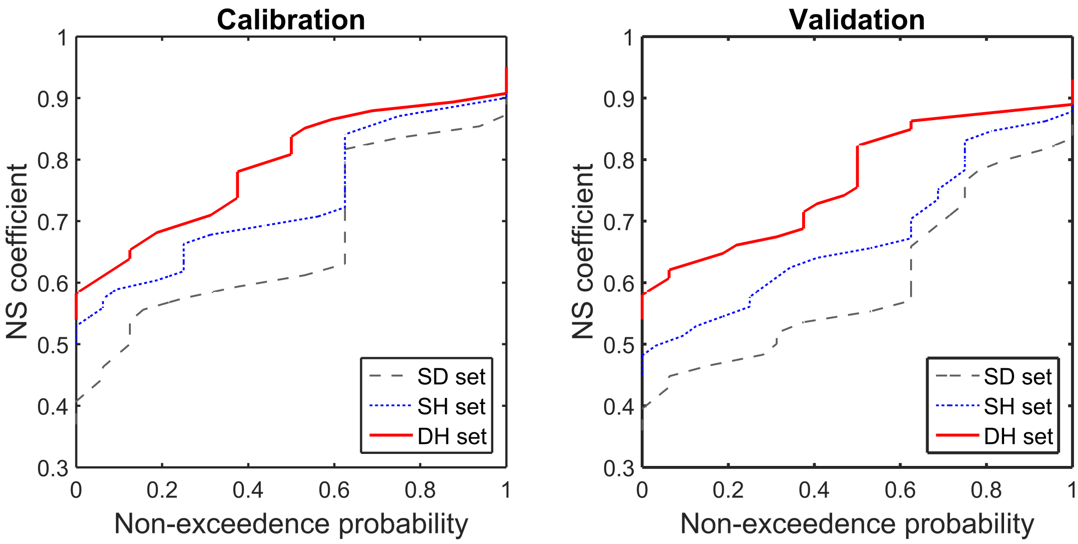


Figure 12. Cumulative distribution of NS coefficient for model calibration using different precipitation datasets .

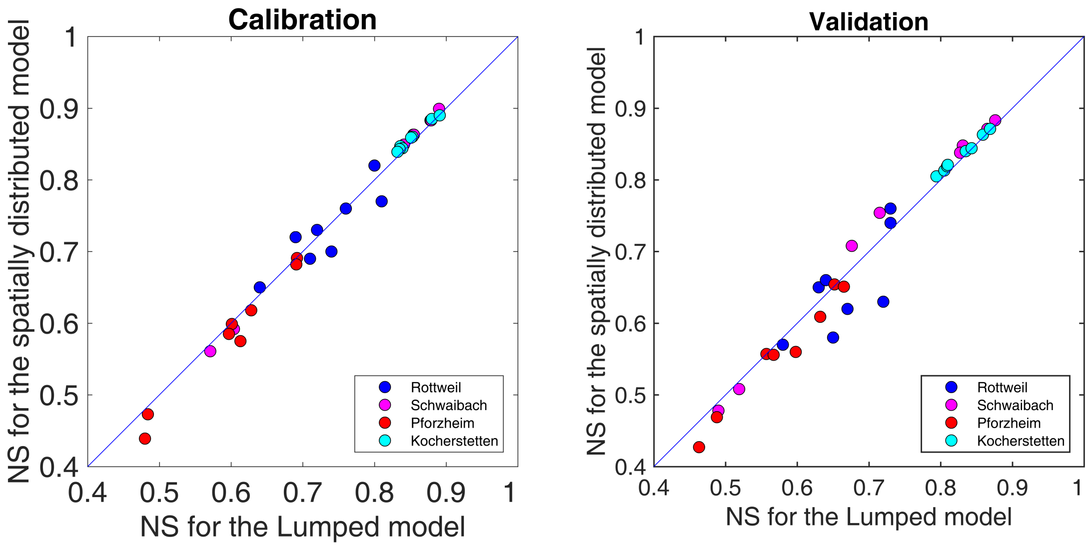


Figure 13. Comparison of model performance for different spatial resolution of model structure.

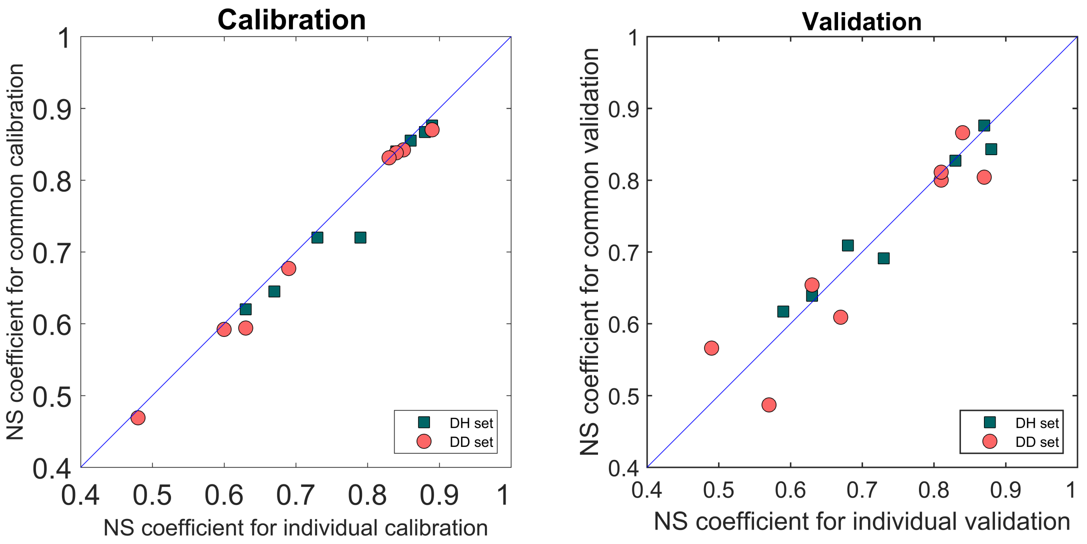


Figure 14. Comparison of model performance for individual calibration and common calibration for different temporal resolution datasets.

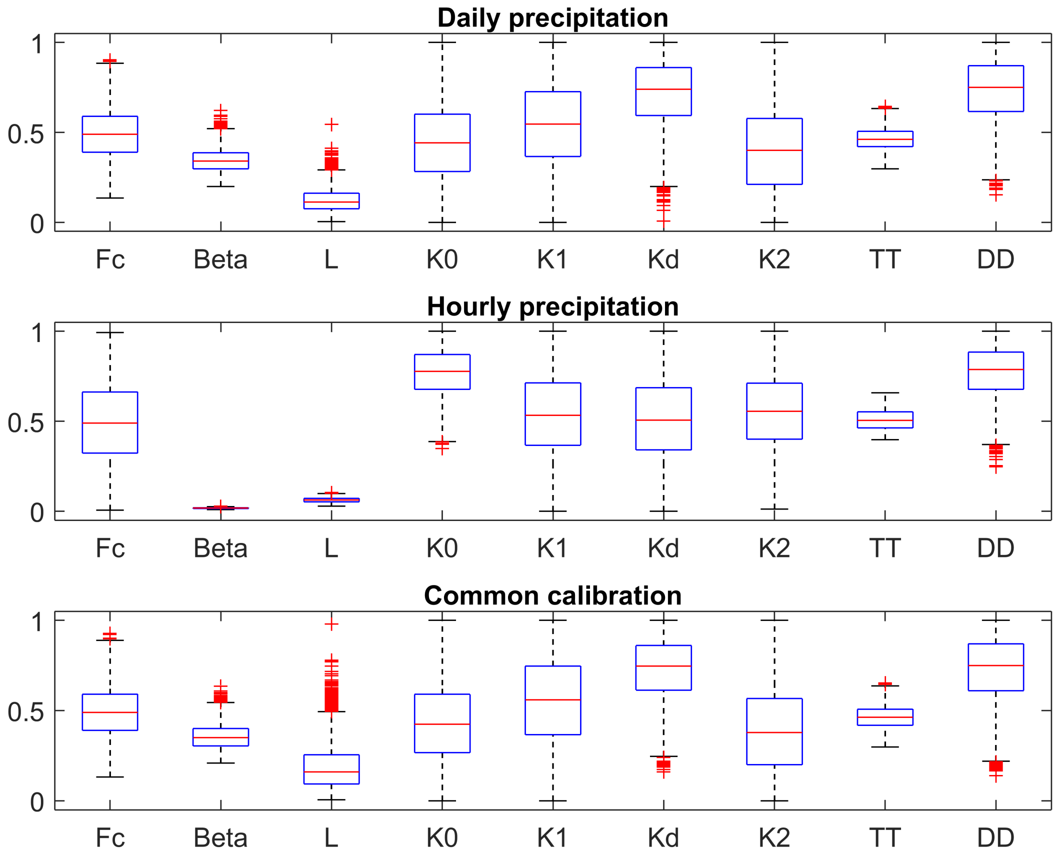


Figure 15. Comparison of model parameters for different temporal resolution for Rottweil catchment.

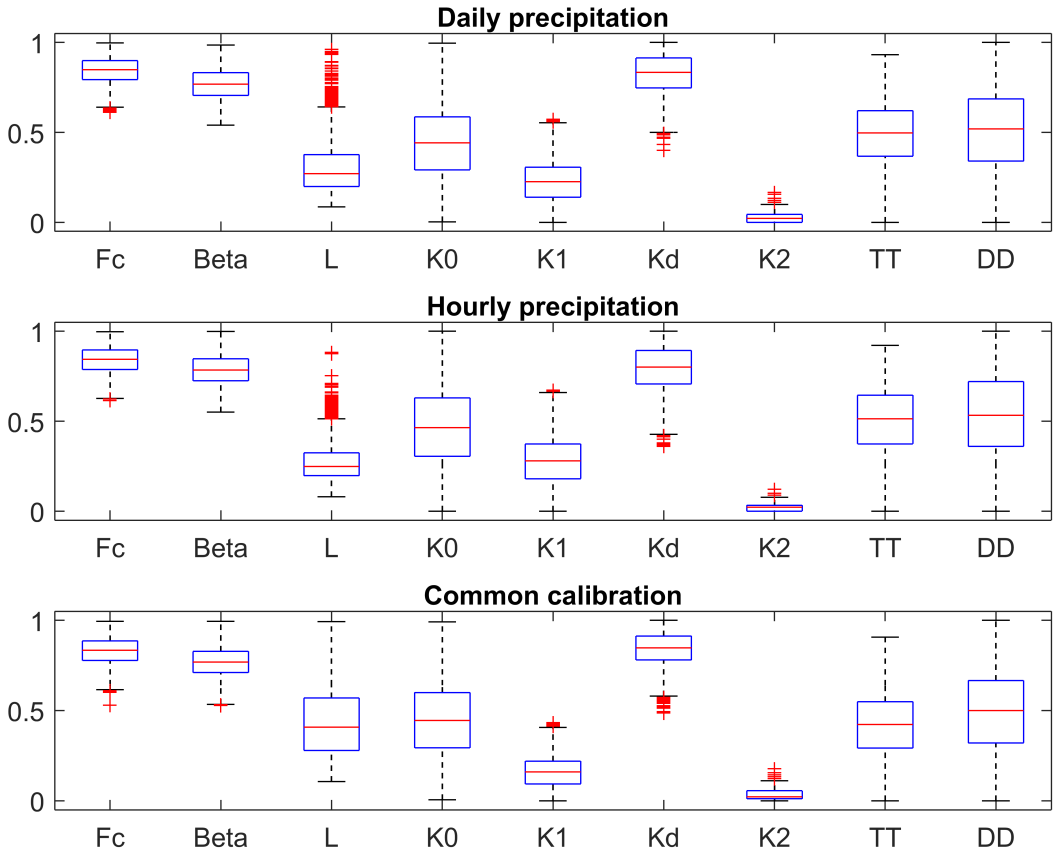


Figure 16. Comparison of model parameters for different temporal resolution for Pforzheim catchment.