First of all, Anna Sikorska and Nadav Peleg are gratefully acknowledged. Their suggestions and comments helped to improve manuscript significantly. Also Florian Pappenberger as handling editor is acknowledged. Please find below the detailed response to the comments of reviewer 1 (Anna Sikorska), reviewer 2 (Nadav Peleg) and the manuscript with tracked changes.

Response to Reviewer 1

This manuscript deals with the problem of disaggregating daily rainfall records into an hourly resolution and particularly assesses the added value of considering the spatial correlation between neighboring stations. The problem of rainfall disaggregation is particularly relevant for hydrological modelling when only daily data are available but high resolution simulations are still required as for flood forecasting or flood predictions in case of fast reacting catchments (with the concentration time smaller than a day). Thus, the manuscript is certainly of a broader interest. Generally, it is also well written. However there are several issues that need more explanations and thus I recommend a major revision.

The authors thank the reviewer for her comments and suggestions. All replies are indicated with page and lines, based on the track-change document.

Major comments:

1. One of my major comments is related to the hypothesis tested by the authors. The authors assume that differences apparent between differently disaggregated rainfall series (V1, V2, V3) will also be present in runoff simulations with a hydrological model. For testing this hypothesis, they use a bucket-type model (HBV). From their results, the authors did not observe any significant difference between these three different rainfall series when fed into the model. I think this is not surprising because this type of model, due to its rather simple structure, may smooth slight differences between different time series present at instant steps, as it reacts to the cumulative rainfall sums over the event rather than to its small variation in time. Despite that, I would expect that you could still see some differences if you analyze instantaneous flows (e.g. event peaks). Yet, if you look only at cumulative statistics of runoff such as monthly average discharges or flow duration curve, you most likely cannot see any differences because these statistics are derived from averaged runoff values. Consequently, these differences could indeed be minute. The only visible effect could be expected on summer and winter extremes. However, these extremes most likely occur in your catchments due to large (and most likely long lasting) rainfall events, for which an exact rainfall distribution within a day is less important. This may explain why you do not observe any differences in these statistics neither. I think these issues should be at least discussed in the manuscript, and particularly the choice of the runoff statistics for the method evaluation.

We thank reviewer 1 for raising her concerns about the choice of the model type. We try to discuss all concerns in the order of her comment. Reviewer 1 is worried about only slight differences between the different rainfall products V1, V2 and V3. However, as shown in Fig. 7 for areal rainfall of one subcatchment, the extreme values differ strongly (5-12 mm) between the different rainfall products for return periods from 2-50 years. Despite that, more frequent rainfall intensities also differ as indicated by Fig. 5 and Fig. 6. So there are differences between the rainfall products.

Also, the reviewer questions the type of model for the actual investigation. We have added a discussion on that in the manuscript (p13 15-21) and want to point out, that the catchments were divided into subcatchments (see also our reply to comment 4 of reviewer 1).

Nevertheless, we agree with reviewer 1, the runoff statistics FDC and Qmon on a daily basis are not suitable to show differences between the rainfall products. However, this was never the intention. Both runoff statistics have been applied only to achieve an overall plausible runoff behaviour for the continuous simulations. We have added a brief discussion on p15 5-9:

“FDC and Q-mon are used to represent the more frequent discharge values. Q-mon accounts for the temporal dependency on the inter-annual variation of the discharge. As mentioned before, the analyzes of FDC and Q-mon allows no direct validation of the rainfall products, but enables an overall plausible simulation of rainfall-runoff processes."

We thank reviewer 1 for the useful suggestion with the possibility of long-lasting events as main reasons for extreme runoff values. Therefore, we have distinguished between summer and winter extremes, so that e.g. the convective events in the fast responding catchment Pionierbrücke and long-lasting events in the winter period are taken into account. We mention the season-dependent genesis and its influence on runoff extremes on p3 34-p4 1:

"to take into account e.g. summer and winter floods with their different genesis and resulting runoff behaviour."

2. The authors use throughout the manuscript terms: recording and non-recording stations. It is however never explained what they mean with that and I assume this is not a generally used term and thus should be explained as it is significant for this manuscript. It appears that by recording stations they mean stations with hourly records and by non-recording - stations with daily records only. To my understanding, daily stations could also be assumed as recording. Consider using different terms or provide an explanation to the terms used.

We thank the reviewer for this hint. We have rephrased both terms to hourly (former: recording) and daily (former: non-recording) stations and believe, these terms are more intuitive for the reader.

3. Not all important details regarding the calibration of the hydrological model is given. Particularly, the fact that the model is calibrated independently with three different disaggregated datasets appears only in the discussion. These independent calibrations obviously lead to different parameter sets which are then used for simulating runoff. As these calibrated parameter sets compensate for possible errors in the model structure and in rainfall data, these errors are propagated on the simulated runoff (and computed statistics). This makes a direct comparison of runoff simulated with these three different time series difficult. Although the authors are aware of that, in my opinion, it would make more sense to use the model with the same set up. In this way, you could focus only on the effect of different rainfall time series and minimize the possible effect of parameter and model errors. Indeed, it could be worth a try to use only parameter sets derived from one calibration, e.g., with the V1 data set, and use it for both other disaggregated sets, i.e., V2 and V3 and in this way assess the gained effect of introducing the spatial consistence between stations into the set V1.

Again, we want to reply to all points/suggestions in the order of appearance. We added the information about the separate calibration for each rainfall product in the method section (p14 18). Also, we are aware of the possible compensation of rainfall product differences by the model parameters. This is the reason for investigations "c1) HBV-simulation results without calibration using three rain gauges as input" and "c2) WaSiM- simulation results without calibration using three rain gauges as input". For both, c1 and c2, no calibration was carried out. A neutral parameter set was applied to avoid (dis-) advantages for one of the rainfall products/biased results. Nevertheless, no differences with HBV and only slight differences with WaSiM could be identified.

4. The figure 1 and the Table 1 suggest that each of three studied catchments is divided into smaller sub-catchments. Do you actually use these sub-catchments for hydrological modelling or do you model the catchment as one unit? I expect that the spatial representation of rainfall may play a role when using sub-catchments instead of the entire catchment.

We agree with the argument of the reviewer, that subcatchments are important for the investigation of the spatial consistence. The discretization shown in Fig. 1 in several subcatchments (with approx. 20 km²) is applied for the investigation.

5. The disaggregation scheme (p. 8): how exactly do you decide which time step is considered to be wet and which as dry? Also, is the same disaggregation scheme used for all three catchments (i.e., the same scheme of distributing daily totals into hourly intervals) or is that adapted for each catchment independently? In addition, the authors write in lines 16-17 p. 8 that parameters of the disaggregation (which exactly?) are extracted directly from the observed high resolution data, which data do you mean exactly (the most recent hourly data)?

The brief explanation of the cascade model leads to open questions by reviewer 1. However, in accordance with our reply to reviewer 2 we have even shortened the description in the manuscript and strongly refer the interested reader to the original manuscript by Müller and Haberlandt (2015). However, we want to answer the questions also in our reply. The decision about the wetness-state of each time step is made randomly, based on probabilities estimated from observed time series. Hence, a second run of the disaggregation leads to a different time series. This is the reason why 10 different realizations have been used as input for the rainfall-runoff simulations. So the scheme in Fig. 3 shows only one possible realization for a single day. The parameters are estimated from the nearest hourly stations with a minimum record length of 7 years. This could be the most recently hourly time series or from a station installed from e.g. 1980-2002 For a description of the cascade model parameters the reviewer is kindly referred to Müller and Haberlandt (2015).

6. The disaggregation V3 uses the station with the highest values per day for deciding on an exact disaggregation way. Can you somehow verify that, i.e. how good it works for other stations? or could you justify this choice?
Reviewer 1 is confused by V3. V3 is not a new disaggregation method, only an alternative to V2 to implement spatial consistence into the disaggregated time series of V1. We have tried to clarify this by adding an additional sentence (12 27): " and is also based on the already disaggregated time series of V1."

Minor comments:
1. P. 5 l. 12-13: change the sentence into: An overview of rain gauges used in this study is given in Fig. 1 while their measuring periods in Tab. 2.
Thanks for the suggestion, we have rephrased the sentence.
2. Table 2; it could be a good idea to add the intervals of rainfall recording.
We thank the reviewer for the hint. We have changed the table and now the temporal resolution of the stations is included.
3. Use "rainfall-runoff model" instead of "rainfall-runoff-model" throughout the manuscript.
Thanks for the hint, we have rephrased it throughout the manuscript.
4. P. 6 l. 12-13: could you give a reference for the finding regarding the nonsensitivity to potential evapotranspiration?
The analyzes was carried out in a pre-study by a student (Herzog, 2013) and is hence not citable in a scientific journal. However, the HBV-IWW model was analyzed regarding its sensitivity to input time series of rainfall, temperature and potential evapotranspiration. The model was not sensitive to the latter one.
5. P. 6. L. 7-9: is temperature data corrected for the elevation and if yes how exactly?
The temperature data has not been corrected. Based on station data, an interpolation was carried out for the study area using External Drift Kriging. The additional information used is elevation.
6. p. 7 l. 9-10: do you mean here "hourly" observed time series? From the table 3 it appears that some records are available from much longer period.
We thank reviewer 1 for pointing out the missing information. Indeed, it should be "observed hourly time series". We have rephrased it.
7. p. 7. L. 13-14: it is not clear which data sources were used to extract the maxima over half of year (hourly, daily, monthly)?
We thank reviewer 1 for pointing this out. We added the data sources to avoid misinterpretations (p8 23-224).
8. Use terms "section" and "subsection" instead of "chapter" and "subchapter".
Thanks for the hint, we have rephrased it throughout the manuscript.
9. The paragraph in lines 18-21 on p. 7 could also be removed.
We thank reviewer 1 for the suggestion, but due to the length of section 3 we think that a brief overview at the beginning enables a better understanding of the investigation.
10. p. 10, l. 15: how many different realizations of the disaggregated time series did you use for these simulations?

We are thankful for pointing out the missing information. We have used 10 realizations for each rainfall product and added the information (p14 18-19).

11. P. 10. L. 26, R1 is not explained before.

We have removed this paragraph.

12. Table 5. The values for the HBV parameters: k1, k1, k2 and kperc are given in days. If the model is run at an hourly time interval, should not these parameters be expressed in hours?

We thank the reviewer for his concern about the units. As it can be identified for k0, not only integer values are possible (minimum for k0=0.25 d = 6 h). The unit [d] was only chosen here for a better understanding of the value, e.g. k2=500 d is more suitable/easier to follow in comparison to k2= 12000 h.

13. Fig, 5 and next: one realization from how many?

10 realizations for each rainfall product, please see our answer to minor comment 10.

References:

Müller, H. und Haberlandt, U. (2015). „Temporal Rainfall Disaggregation with a Cascade Model: From Single-Station Disaggregation to Spatial Rainfall“. *Journal of Hydrologic Engineering* 20 (11), p. 04015026.

Herzog, Y. (2013): “Sensitivität eines hydrologischen Modells auf Veränderungen in den Klimavariablen Niederschlag, Temperatur und potentielle Verdunstung”. *Student thesis* at the Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Faculty of Civil Engineering and Geodesy, Leibniz Universität Hannover (in German).

Response to Reviewer 2

In their paper, the authors explored runoff response to hourly rainfall series with different degrees of spatial consistence. Daily rainfall series were disaggregated using a multiplicative random cascade method to generate 3 rainfall products – one without spatial consistency, and two others with a different level of spatial consistency. The question of the need for spatial consistence of rainfall disaggregation for hydrological modeling is interesting and relevant for the readers of HESS. My concerns are mostly minor, but I do have one major concern: I am not convinced that the HBV model is the right model to use for this experiment. Firstly, the model parameterization can overcome the differences in distributed rainfall products (as is also mentioned by the authors). Secondly, the area of the sub-catchments is very large (>20 km^2), so the rainfall spatial variability is essentially not introduced into the model. What was the reasoning in choosing HBV model? The studied catchments are rather small and distributed hydrological models (as WaSiM) could be easily applied. Other than that, the text requires some further editing. There is a disproportion between the length of the text and the number of figures and tables. I suggest reducing the length of the manuscript (the text is very repetitive) and have some of the figures/tables as supplementary information. Moreover, the terminology is inaccurate in some places. My recommendations for the text editing, along with some minor comments, are listed below. Overall I think the numerical experiment suggested by the authors is sound, and that the hydrology community will benefit from the paper. If my discussion below seems critical, it is only because I want to improve the final manuscript.

The authors thank the reviewer for his comments and suggestions. All replies are indicated with page and lines, based on the track-change document.

[Page Lines]
[introduction] I am missing some discussion about the importance of rainfall spatial variability to the runoff in general. The focus is mainly on the number of rain-gauges needed, but the readers will benefit from the understanding that it is important to capture (by dense rain-gauge networks, remote sensing or modeling) the rainfall spatial pattern right, as using a single rain-gauge or a single time series of areal rainfall the simulated runoff is likely to be over-(under)estimated. I can think of several papers that discuss this point: Gires et al. (2012, JoH); Gires et al. (2013, UWJ); Paschalis et al. (2014; JoH); Ochoa-Rodriguez et al. (2015, JoH); Peleg et al. (2017, HESS).
We thank the reviewer for the useful references and have implemented a selection of the references in the manuscript (p2 10-13).
[2 11] "non-recording stations" and "recording stations" – I would adopt a simpler terminology, e.g. "hourly stations" and "daily stations" (or hourly-recording stations).
Indeed, a simpler terminology improves the understanding. We have changed the names to "hourly stations" and "daily stations".
[2 12] "time series from non-recording stations can then be disaggregated" – same here, I would simplify the terminology used. Consider revising to "daily time series can be disaggregated to hourly…"
In accordance with the former comment we have rephrased this part.
[2 15] "over other rainfall generators" – I suggest to replace this with "over other disaggregation methods" or similar. Rainfall generators preserve the statistics of a rainfall series but often not used to disaggregate a given time series while preserving the rainfall amount at the coarse scale.
We thank the reviewer for pointing out the misleading sentence. The advantage of the cascade model mentioned here is the temporal reference to observations, and hence, the facilitation of their statistics down to finer temporal resolutions. On the contrary, rainfall generators create an

hourly time series "out of nothing", which is a more difficult task (although they have other advantages). We have tried to clarify this point by rephrasing this sentence (p2 18-23).
[2 17] "higher" – finer.
Thanks for the hint, we have rephrased it.
[2 37] "three bivariate rainfall characteristics". Why 'bivariate'? You have a single variable, if only rainfall is explored. I would change the terminology to "three spatial rainfall characteristics" or "three spatial rainfall indices".
The reviewer is concerned about a wrong terminology. The rainfall characteristics are "bivariate", since they result from the comparison of exactly two time series. Each comparison leads to one point of the e.g. observation cloud in Fig. 5. In a multivariate analysis, all time series are taken into account with only one resulting, final point/value. Hereby, the distance-dependency gets lost. The term "bivariate" is common for these characteristics from the point of the authors view (Wilks (1998), Haberlandt et al. (2008), Müller und Haberlandt (2015, 2018)) and is hence kept in the manuscript.
[3 8] "investigations" replace with "studies".
Thanks for the hint, we have rephrased it.
[3 12] "rainfall data sets" – they all emerge from the same data set, consider replacing with "rainfall products" or similar to distinguish from the original time series.
Thanks for the hint, we have replaced it by "rainfall products" throughout the whole manuscript.
[3 20] "by amongst others" – change to "by others, as".
Thanks for the hint, we have rephrased it.
[3 22] "Runoff statistics have no connection to time" – please revise this sentence, runoff statistics are time dependent, e.g. statistics of runoff diurnal cycle.
The reviewer is confused by the terminology, "no connection to time". Indeed, runoff statistics have always a connection to time, e.g. the period or season of runoff time series used for analyzes, or, as the reviewer points out, single hours of a day to achieve a representative diurnal cycle. However, the all these statistics have no connection to a concrete time step/day/month – so it is impossible conclude from a statistic value (e.g. HQ_{100} or average discharge in the 5^{th} hour of a day) to its day of occurrence (it it has ever occurred!). We have tried to solve the issue be rephrasing the sentence to "Runoff statistics are time-independent" (p3 31).
[3 25] "to take into account different genesis" – not a clear sentence.
Thanks for the hint, we have rephrased it to "to take into account summer and winter floods with their different genesis and resulting runoff behavior" (p3 34-p4 2).
[3 27] "investigation area" – I believe "study area" is a more common phrase to use.
Thanks for the hint, we have rephrased it.
[3 27] replace "chapter" with "section".
Thanks for the hint, we have rephrased it throughout the whole manuscript.
[Fig. 1] What are "p-stations" and "gauges" stand for? I guess p-stations are rain-gauges and gauges stand for discharge-gauges. Please correct the legend accordingly.
We have changed it according to the reviewer's suggestion.
[Table 2] The names of the gauges are not important for the readers (and are not labeled in Fig. 1 or 2). They can be removed to shorten the length of the table.
We thank the reviewer for the useful suggestion and have changed the table accordingly.
[6 12] "has been shown not to be that sensitive as model input" – a reference is needed here.
The analyzes was carried out in a pre-study by a student (Herzog, 2013) and is hence not citable in a scientific journal. However, the HBV-IWW model was analyzed regarding its sensitivity to input time series of rainfall, temperature and potential evapotranspiration. The model was not sensitive to the latter one.
[7 6] "as per" – replace with "as in".
Thanks for the hint, we have rephrased it.

[7 11] Please define “monthly extreme values”. Do you mean hourly extremes on monthly basis? If not, than I would expect monthly extreme values time series to start with the daily discharge time series.

Thanks for the hint, we have rephrased it to the reviewers suggestion “hourly extreme values on a monthly basis” (p8 20-21).

[7 25] “The most important input for rainfall-runoff models are long and high-resolution rainfall time series from a dense rain gauge network” – This sentence is more suitable to the introduction section and it needs to be supported with a reference. High-resolution – do you mean temporal and spatial? If so, also weather radars and, if catchments are large enough, satellite rainfall data can be used.

The reviewer is right with his argument, that this sentence fits better to the introduction part. Indeed, the information has already been given in section 1 (with references). Hence, we have deleted the sentences here.

[7 25] to [8 3] This paragraph is somehow a repetitive to what was already stated in the introduction. It can be removed from the text.

Please see our reply to the former comment.

[8 15] to [9 6] I think this part can be also removed as it is described (in details) in Müller and Haberlandt (2015). Unless there is some important information here that is later discussed. The disaggregation scheme is well illustrated (Fig. 3) and explained in the preceding paragraph.

We thank the reviewer for the useful suggestion and have deleted the paragraph.

[9 8] “b) Bivariate characteristics” – replace with “Rainfall spatial characteristics indices” or similar.

The reviewer is referred to our reply to comment [2 37].

[9 9] to [9 13] “The disaggregation of single time series is carried out without taking into account time series of surrounding stations. For each time series the cascade model distributes the wet time steps randomly during a wet day due to its disaggregation scheme. Hence, spatial consistence of rainfall is underestimated after the disaggregation. Spatial consistence is defined in this investigation by bivariate spatial rainfall characteristics, the namely probability of occurrence, Pearson’s coefficient of correlation, and the continuity ratio (Wilks, 1998)” – This all is a repetitive of what was already mention in the introduction. Be concise. I would replace the first paragraph of this sub-section with a single sentence, e.g. “The rainfall spatial characteristics are following the ones used by Haberlandt (2008) and are briefly described in the following”.

We thank the reviewer for his suggestion. We have replaced the introduction paragraph (p11 2-3).

[Eq. 1] “Z” – stands for rainfall intensity?

Yes, we have added the missing information.

[Eq. 2] The “x” in the denominator should be deleted. It reads like a variable.

We have replaced the misleading operation sign by a representative operator.

[Eq. 1 and 2] Consider removing them. I think that the text to describe Eq. 1 is sufficient for the readers to understand the rainfall occurrence score and Pearson’s coefficient of correlation (Eq. 2) is quite well known.

We are thankful for the suggestion of the reviewer. However, the third characteristic (Eg.3) is rather known and should remain in the manuscript. To keep it consistent, we keep both equations. They are also essential for the validation of the spatial consistence.

[9 29] “(see Fig. 4)” – I don’t see it.

We thank reviewer 2 for this hint. Indeed, the regression lines are in the former Fig. 5, now in Fig. 4.

[10 6] to [10 27] I found the description of V2 to be very long; I am not sure if the readers needs all this information about the method. I recommend to shorten this part (and removing Eq. 4).

The concept of V2 is already explained at the introduction. Are there any modification from what is presented in Müller and Haberlandt (2015)?
We thank reviewer 2 for the useful suggestion. We have shortened the paragraph.
[10 28] to [10 33] The part describing V3 is concise and well written, but again – there are many repeats to what was already written in the previous sections.
We thank reviewer 2 for the hint. We have shortened the text in the introduction section to avoid repeats.
[Fig. 4 and Table 5] Can be moved to the Supplementary Information (SI).
We are thankful for the reviewers suggestion and have moved both to the SI.
[11 23] to [11 26] A repetitive.
We agree and have shortened the paragraph.
[12 1] Please define the periods for summer and winter.
We have implemented the missing information (p14 14-15).
[12 10] FDC should be calculated on both hourly and daily scales! Important information can be obtained from exploring both scales. What is the point in disaggregating the rainfall to hourly scale and examining it on daily scale?
Reviewer 2 is right- an evaluation of the rainfall products on an hourly basis with FDC and Q-mon on a daily basis can be questioned. However, this is not the idea behind including both runoff criteria. Rainfall products are expected to influence the extreme values in summer and winter half year. FDC and Q-mon are taken into account additionally to represent the overall behavior of the rainfall-runoff processes to e.g. achieve realistic filling volumes for all storages before a severe rainfall event occurs.
We have rephrased the whole paragraph and added some explanations for clarification (p15 2-10).
[Eq. 5] Remove the “max”.
We have removed it.
[Eq. 6] Can be moved to the SI.
We thank reviewer 2 for the suggestion to shorten the text. However, a movement of Eq. 6 to the SI would result in an implementation of the applied quantiles in the text. A complete neglecting of the applied quantiles in the manuscript would complicate the understanding of the calibration procedure. Hence, we prefer to leave Eq. 6 in the manuscript.
[Eq. 7] Remove “min”. What is the logic behind the weights?
We have removed it.
[Table 5] Move to SI.
We are thankful for the reviewers suggestion and have moved it to the SI.
[Section 3.2a] I am missing the information about how the HBV model is “distributed” in space. Are the catchments represented as one unit or many? If many, I would like to see a figure with how the units are distributed in space (following the sub-catchments illustrated in Fig. 1 and 2?). This is a critical point as you later discuss the spatial representation of rainfall over the catchments, but it is not clear how the model relate to rainfall in space.
Reviewer 1 points out the missing information about the spatial discretization of the HBV-model. Indeed, the subcatchments shown in Fig. 1 and 2 represent the spatial discretization. We have implemented additional sentences pointing it out in the method section along with a paragraph regarding model selection, providing also some information with a focus on the choice of interpolation method and its reasons (p13 12-27).
[14 1] “The spatial resolution of WaSiM applications covers several scales ranging from tens of meters to a few kilometers” – but what is used here? For example, what was the spatial resolution of the modeled rainfall?

We thank reviewer 2 for pointing out the misplaced information. The spatial resolution of WaSiM is included in the discussion section (150 m x 150 m)). We have implemented it in the method section as well (p16 10-11).

[14 22] The calibration period is quite short, isn't it?

For the choice of calibration and validation period a classical split-sampling was applied. Depending on the total period available, the first half of the period was used for the calibration, the second one for the calibration. Indeed, for Tetendorf it is maybe critical with14 years in total. However, this is discussed in the results section (p28 12).

[15 2] to [15 13] and [Table 7] I do not see the need in repeating the results reported in Müller and Haberlandt (2015) here. It can be replaced with a one line sentence indicating the method advantages and limitations, but this should be anyhow done prior to the result section (i.e. in the methods section).

We agree with the reviewer's suggestion and have shortened the results summary and moved it to the methods section (p9 21-24).

[Fig. 5] Where are all the dots (other rain-gauges) coming from? Is there any reason to present the different scores for a distance of 250 km? I recommend limiting the distance to 25 or 50 km, to agree with the catchment size. Please give some information about the fitting that are presented. If the fits are not discussed, than the lines can be removed. It will be useful to have the same figure for the other catchments in the SI.

We thank the reviewer for pointing out the missing information. The black circles result from observations (details are explained in Müller and Haberlandt, 2015). We have added the information in the figure captions. The analyzes was carried out for the other two catchments as well and can be found in the SI. According the suggestion the x-axes have been limited to 40 km.

[16 15] "areal rainfall intensity" – please define how it was calculated. A simple arithmetic mean?

The description was included in the discussion section. We have moved it to the method section (p13 21-23).

[Fig. 6] Can also be moved to the SI, to reduce the number of figures in the paper.

We thank the reviewer for this suggestion, the figure has been moved to SI.

[17 14] "runoff".

We have corrected it.

[17 15] Why Weibull? Is the fit good for all rainfall products (V1 to V3)? What is the length of the rainfall series used to generate Fig. 7?

The reviewer is concerned about the goodness-of-fit of the Weibull-distribution function for all rainfall products. Here, only the Weibull-plotting position has been used, which needs no fitting. It combines only the rainfall intensity and its rank in a sorted population for comparative analyzes. The length of the time series is 53 years, we have added the information in the figure caption.

[Fig. 7] What is the size of sub-catchment 2? From Figure 1 I would estimate around 20 km^2. If this is the case, I would argue that V2 and V3 are likely overestimating the extreme rainfall intensities. For example, ~38 mm h^{-1} for a 10-year return period over a ~20 km^2 sounds quite a lot for me. It can be reasonable for a measurement from a single rain-gauge, but as we are looking at areal rainfall I would expect much lower values – even in the range of V1 (as the extreme rainfall intensity is expected to be smaller for the same return period when shifting from a point scale to a larger areal scale, e.g. our recent study in JoH [Peleg et al., 2018] and many others). I would suggest to compare the resulted areal rainfall to an observed extreme rainfall from a single-gauge, even the comparison will be areal to point, just to get a sense of the differences between the scales.

Reviewer 2 points out several issues regarding differences between point and areal rainfall, which we try to answer in the following. First, the reviewer's estimation of the catchment size with approx. 20 km² is right (we have added the missing information regarding spatial distribution in the HBV-model description (p 13 14)). A comparison of the estimated areal rainfall with "observations" is not possible, as the reviewer points out himself, since only one hourly station is available for this catchment. A comparison, as suggested, with rainfall from a single-gauge is also not possible, since the maximum record length of available hourly data is shorter (and a comparison of extreme values resulting from different periods with different lengths would not be representative) and the station is located outside the catchment. Both not possible comparisons show the need for the rainfall disaggregation.

However, we totally agree with the reviewer's point of decreasing rainfall intensity with increasing area size. This is most important for high-resolution rainfall (as in Peleg et al., 2018), but decreases with increasing time step length due to cumulative character of a single time step (cumulating rainfall i) over time and due to e.g. cloud movements ii) over space). However, simulations are carried out continuously and hence, also underestimations occur in the actual setup (Müller and Haberlandt, 2018), if the rain gauge was not located in the center of the storm/event. Hence, a reduction by e.g. areal reduction factors is not included, since i) this would introduce another uncertainty and ii) is only common for load-case applications with using only extreme values as input. Also, the areal rainfall of the shown subcatchment results from 3 hourly time series and hence can be seen as representative for the area.

Nevertheless, we have implemented a discussion on that in the HBV-model description (p13 23-27).

[19 4] "flood quantiles are shown for a return period of 100 years" – It doesn't make sense as the observed period is much shorter. I would focus on 50 y return period to reduce the uncertainties.

The reviewer's suggestion is right, extrapolation is limited to the length of observations. We mention this one sentence later: "However, the extrapolation is limited by the length of the simulated runoff time series." (p24 2-4) The limitations are three times the runoff time series length and are indicated for each catchment by a dashed line in all figures showing extreme values. Results for higher return periods are not discussed in the manuscript. Nevertheless, for all three catchments extreme values are shown up to return periods of 100 years to enable comparisons between the figures (and for the catchment Pionierbrücke the extrapolation leads to return periods of 75 years.

[Fig. 12] For which catchment? Or is it for the entire region? Same comments as of Fig. 5 above.

The reviewer points out a missing information. The seasonal characteristics are estimated using all available stations for the entire regions with long time series (please see also the reply to the former comment). We have added the missing information in the manuscript in the same sentence (p28 3-5).

[25 5] to [25 10] More suitable to be in the Introduction section.

The sentences without results have been moved to the introduction.

[27 13] "It can be summarized, that the number of rain gauges has only a minor, but no systematic influence on runoff statistics for the catchments used in this investigation" – but likely not because of the number of rain-gauges in a catchment but because of the hydrological model that was used! I would like to see the same analysis using a fully distributed hydrological model that can account for spatial rainfall variability at the sub-catchment scales.

We see the reason for the reviewer's concern. We have discussed this in the model description (p13 14-20). For Pionierbrücke, additional simulations with WaSiM (150 m x 150 m spatial resolution) have also not shown systematic differences between the rainfall products (only slight differences for seasonal extreme values). Since Pionierbrücke was the catchments with our highest expectations to see differences resulting from the different rainfall products, we did not carry out further investigations for Tetendorf and Reckershausen.

[27 14] A repetitive.

We thank the reviewer for the hint and have deleted the sentence.
[Table 12 and 13] SI.
We thank the reviewer for the suggestion and have moved both tables to the SI.
[31 8] to [31 15] There are some repetitive here is well.
Thanks for the hint, we have removed the repetitive.
[32 9] "the IDW algorithm with an altitudinal rainfall adjustment, which was carried out by a linear regression model" – The IDW is likely to smooth the rainfall in space, thus reducing the spatial rainfall variability and the variability in flow.
We agree with reviewer 1, IDW leads to a smoothing of rainfall. However, the overall idea behind this investigation is to analyze the (non-, semi-) simultaneous occurrence of rainfall at different stations, which is assumed to cause the variability in space. A smoothing by IDW occurs for all rainfall products V1, V2 and V3 and the effect of the interpolation method should be minor in comparison to the effect of the spatial consistence.
[32 10] But FDC compare daily discharges, right? I guess that at hourly scale the differences are clearer.
The reviewer is concerned about the temporal resolution of the runoff time series used for the FDC. For the majority of the simulated periods, differences in the discharge during a day are small on an hourly time step. Hence, the daily resolution delivers satisfying results for FDC as a runoff statistic applied to represent the overall runoff behavior. For days with e.g. intense rainfall events and hence a high-dynamical runoff response the daily values are not representative. The representation of these instantaneous runoff values/peaks is very important and hence, these are represented by two of the four applied runoff statistics (summer and winter extremes) (please see also comment [12 10]). The authors do not think that a FDC based on hourly values would improve the calibration of the models and/or show greater differences between the different rainfall products. Also, the calibration and analyzes procedure based on a daily basis was applied before (e.g. Wallner and Haberlandt, 2015).
[38 6] to [38 11] Part of the reason is because the sub-catchments sets in SWMM model are often much finer than the 20-km set by HBV in this study. When exploring the hydrology response using small sub-catchments with SWMM the effect of the distributed rainfall in space are evident on the hydrological flows (see for example the study by Peleg et al. 2017, HESS). I have reasons to believe that if HBV model was set to have many more small sub-catchments for this study, the results of the differences between V1, V2 and V3 would look different. That rise the question of the suitability of HBV model with the current setting of a few large sub-catchment to explore the sensitivity of the hydrological response to different rainfall spatial characteristics.
We agree with the reviewer regarding the required spatial resolution for urban hydrology. Due to less storage and retention capabilities (no soil or vegetation the catchment reacts much faster. Hence, finer temporal and spatial resolution of rainfall is required. For catchments with a size of few hundred square kilometers, hourly resolution (Melsen et al., 2015) and thus a coarser spatial resolution is sufficient. The reviewer assumes higher differences between the rainfall products if the subcatchments would be smaller. However, this assumption was investigated by the application of WaSiM for catchment Pionierbrücke. Please see our reply to the comment [27 13].
[Conclusion section] The conclusion part is a mix of discussion, summary and conclusions. Consider revising it to make the outcome of the experiment clearer to the readers.
Indeed, the final section should also include a summary. Hence it was renamed to "Summary, conclusions and outlook". We follow the suggestion of the reviewer and have moved parts with discussions to the last but one section of the manuscript. However, we think the formulated open questions should remain in this section.

References:

Haberlandt, U., Eschenbach, A.-D. Ebner von und Buchwald, I. (2008). „A spacetime hybrid hourly rainfall model for derived flood frequency analysis“. *Hydrology and Earth System Science* 12 (6), p. 1353–1367.

Herzog, Y. (2013): "Sensitivität eines hydrologischen Modells auf Veränderungen in den Klimavariablen Niederschlag, Temperatur und potentielle Verdunstung". *Student thesis* at the Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Faculty of Civil Engineering and Geodesy, Leibniz Universität Hannover (in German).

Melsen, L. A., Teuling, A. J., Torfs, P. J. J. F., Uijlenhoet, R., Mizukami, N. und Clark, M. P.: „Hydrology and Earth System Science Opinions: The need for process-based evaluation of large-domain hyper-resolution models“. *Hydrology and Earth System Science* 20, 1069–1079, 2015.

Müller, H. und Haberlandt, U. (2015). „Temporal Rainfall Disaggregation with a Cascade Model: From Single-Station Disaggregation to Spatial Rainfall“. *Journal of Hydrologic Engineering* 20 (11), p. 04015026.

Müller, H. und Haberlandt, U. (2018). „Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology“. *Journal of Hydrology 556*, p. 847-864.

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., Burlando, P.: Spatial variability of extreme rainfall at radar subpixel scale. J. Hydrol.

Wallner, M. and Haberlandt, U.: Klimabedingte Änderung von Hochwasserabflüssen im Aller-Leine-Einzugsgebiet - Eine Fallstudie mit HBV-IWW. Hydrol. Wasserbewirtsch. 59 (4), 174–183, 2015.

Wilks, D. S. (1998). „Multisite generalization of a daily stochastic precipitation generation model“. *Journal of Hydrology* 210 (1-4), p. 178–191.

Rainfall disaggregation for hydrological modeling: Is there a need for spatial consistence?

Hannes Müller[1], Markus Wallner[2], Kristian Förster[1,3,4]

[1]Leibniz Universität Hannover, Institute of Hydrology and Water Resources Management, Leibniz Universität Hannover, Hanover, 30175, Germany
[2]bpi Hannover - Beratende Ingenieure, Hanover, 30177, Germany
[3]Institute of Geography, University of Innsbruck, Innsbruck, 6020, Austria
[4]alpS - Centre for Climate Change Adaptation, Innsbruck, 6020, Austria

Correspondence to: Hannes Müller (mueller@iww.uni-hannover.de)

Abstract.

In this investigation, the influence of disaggregated rainfall ~~data sets~~ products with different degrees of spatial consistence on rainfall runoff modeling results is analyzed for three meso-scale catchments in Lower Saxony, Germany. For the disaggregation of daily rainfall time series into hourly values a multiplicative random cascade model is applied. The disaggregation is applied on a per station basis without consideration of surrounding stations, hence subsequent steps are then required to implement spatial consistence. Spatial consistence is here represented by three bivariate spatial rainfall characteristics, complementing each other. A resampling algorithm and a parallelization approach are evaluated against the disaggregated time series without any subsequent steps. With respect to rainfall, clear differences between these three approaches can be identified regarding bivariate spatial rainfall characteristics, areal rainfall intensities and extreme values. The resampled time series lead to the best agreement with the observed ones. Using these different rainfall ~~data sets~~products as input to hydrological modeling, we hypothesize that derived runoff statistics are subject to similar differences as well. However, an impact on the runoff statistics summer and winter peak flows, monthly average discharge and flow duration curve of the simulated runoff time series cannot be detected. Several modifications of the investigation using rainfall runoff models with and without parameter calibration or using different rain gauge densities lead to similar results in runoff statistics. Only if the spatially highly resolved rainfall-runoff WaSiM-model is applied instead of the semi-distributed HBV-IWW-model, slight differences regarding the seasonal peak flows can be identified. Hence, the hypothesis formulated before is rejected in this case study. These findings suggest that (i) simple model structures might compensate for deficiencies in spatial representativeness through parameterization and (ii) highly resolved hydrological models benefit from improved spatial modeling of rainfall.

1. Introduction

Flood quantiles are important information for the creation of flood hazard maps, the construction of riverfront buildings and landscape development plans, for example. For ungauged catchments and catchments with short discharge observation periods, rainfall-runoff modeling is a possibility to obtain long, simulated discharge time series which can then be used for derived flood frequency analysis.

The most important data input for rainfall-runoff modeling are rainfall time series (Beven, 2001). Melsen et al. (2015) gave an overview of typical processes for different catchment sizes and corresponding temporal resolutions. For catchments with areas of a few hundred square kilometers, time series with hourly resolutions are required for the simulation of instantaneous flood peaks. In most of these cases, observed rainfall time series of that kind are i) too short or ii) the network density is too low. This is anBoth are issues-, because i)it limits the length of the simulation period and hence the derivable flood frequencies and ii) effects the representation.Investigations to the influence of spatial rainfall patterns can be found in (Krajewski et al. (1991), Ogden and Julien (1993), Obled et al. (1994) and Nicotina et al. (2008)) and hence the areal rainfall used as input for the rainfall-runoff simulations.

Usually, time series of non-recordingdaily stations have much longer observation periods and a higher network density. Daily time series can be disaggregated to hourly time series by usingUsing information from the observed, hourly time series of recording stations, time series from non-recording stations can then be disaggregated. One possible method for the disaggregation of rainfall is the multiplicative random cascade model (e.g. Olsson, 1998), which was originally introduced within the field of turbulence theory (Mandelbrot, 1974). The use of observed daily time series as input is a strong advantage of the cascade model, over other rainfall generators (e.g. Poisson-cluster models (Rodriguez-Iturbe et al., 1987, Onof et al., 2000)), since starting with "true" rainfall amounts and intermittency facilitates their conservation to higher finer temporal resolutions, overwhile other rainfall generators (e.g. Poisson-cluster models (Rodriguez-Iturbe et al., 1987, Onof et al., 2000)), try to generate time series with a certain temporal resolution and target statistics without any temporal reference to observations.

With the micro-canonical cascade model, the rainfall amount of a coarse time step (e.g. a day) is conserved exactly through the disaggregation process, so that an aggregation of the disaggregated time series would result exactly to the original observed time series. Starting from a daily resolution, an hourly temporal resolution is achieved, which is a convenient input resolution for many rainfall-runoff models. However, this disaggregation method is a univariate process, carried out for single time series only which are independent from the time series of surrounding stations. Through the systematically random distribution of the rainfall amount within a day, unrealistic patterns of rainfall are generated and the spatial consistence of rainfall is missing. If an unrealistic spatial distribution of rainfall is used within a rainfall-runoff simulation, it can be assumed that this affects the simulated runoff. However, a realistic spatial representation of rainfall is essential if the time series serve as input for rainfall-runoff modeling (e.g. Gires et al., 2013, Paschalis et al., 2014, Ochoa Rodriguez et al., 2015, Peleg et al., 2017).)

Müller and Haberlandt (2015) have introduced a resampling scheme as a subsequent step after the disaggregation process, which can be used for the implementation of spatial consistence within disaggregated time series. Spatial consistence is hereby defined by three bivariate rainfall characteristics: the probability of occurrence, the coefficient of correlation after Pearson, and the continuity ratio (Wilks, 1998). The implementation of spatial consistence for hourly time series was proven by the above mentioned bivariate characteristics in addition to areal rainfall intensities resulting from the disaggregated time series. Without resampling, areal rainfall intensities were underestimated. The resampling algorithm was additionally tested for time series of 5 min resolution by Müller and Haberlandt (2016). Bivariate rainfall characteristics as well as the simulated runoff from an artificial sewage system were positively validated against observed rainfall time series and its resulting simulated runoff.

Haberlandt and Radtke (2014) overcame the lack of spatial consistence using a parallelization approach. For each day they chose the relative diurnal cycle from the disaggregation time series of the station with the highest daily rainfall amount and then adopted this diurnal pattern for all other time series. This procedure leads to simultaneous rainfall events for all stations with registered rainfall for that day. This homogenous rainfall could , which leads to an overestimation of simulated floods, but is acceptable preferred in comparison to a possible underestimation. by the before mentioned unrealistic spatial rainfall patterns. However, Ding et al. (2016) also used disaggregated time series for their rainfall-runoff analyzes with a focus on instantaneous peak flows, but without any subsequent changes to the disaggregated time series. Neither a systematic over- or underestimation of simulated discharge and flood peaks can be found in both investigations.

It can be questioned, why the simulation results from both investigationsstudies, based both upon unrealistic spatial rainfall behavior, leads to an acceptable representation of observed discharge characteristics. The hypothesis of this study is that rainfall data sets products with different degrees of spatial consistence will result in different areal rainfall intensities and hence influence runoff statistics derived from simulated runoff time series. Therefore, three different rainfall data sets products are used as input for rainfall runoff modelling: disaggregated time series with (Müller and Haberlandt, 2015) and without (Ding et al., 2016) implemented spatial consistence, and thirdly time series with an “overestimated spatial consistence” by parallelization (Haberlandt and Radtke, 2014). A systematic comparison is carried out including rainfall-runoff simulations with and without calibration, differing station densities and different rainfall-runoff models.

In general, calibration and validation of rainfall-runoff model parameters are carried out by a quantitative comparison of simulated and observed time series. This strategy is not applicable by using disaggregated rainfall time series as input, since the daily rainfall amount is distributed randomly in time during a day. Hence, the temporal connection between rainfall and runoff is missing. An alternative strategy is the calibration on runoff statistics and has been applied before by others, asamongst others Yu and Yang (2000), Westerberg et al. (2011), Haberlandt and Radtke (2014), Wallner and Haberlandt (2015) and Ding et al. (2015). Runoff statistics are time-independenthave no connection to time, but contain useful information about the hydrograph and hence about the hydrological regime and its characteristics. It is assumed, that by a simultaneous consideration of different complimentary runoff statistics, the runoff behavior can be represented sufficiently. Possible runoff statistics are: runoff extremes for different seasons of a year (to take into account e.g. different summer and winter floods with their different genesis and resulting runoff behavior), flow duration curves (to describe the overall behavior), and average monthly values (to describe the inter-annual variability).

The manuscript is organized as follows: after a brief description of the ~~investigation~~ study area and the data in ~~chapter~~ section 2, the rainfall generation including the implementation of spatial consistence and the applied rainfall-runoff models including the calibration technique are explained in section ~~chapter~~ 3. Section ~~Chapter~~ 4 includes the results for both the rainfall generation and rainfall-runoff-modeling. A summary of the rainfall-runoff model results is provided in section ~~chapter~~ 5 and general conclusions and a brief outlook are provided in section ~~chapter~~ 6.

2. Data & Investigation area

2.1 Catchments

The investigation is carried out for three catchments in the Aller-Leine river basin, namely Reckershausen, Pionierbrücke and Tetendorf (see Fig. 1). The river basin is situated in Lower Saxony, Northern Germany. Based on the Köppen-Geiger climate classification, the river basin can be divided into a temperate oceanic climate in the north and a temperate continental climate in the south (Peel et al. 2007). For Reckershausen an additional investigation regarding rain gauge network density is carried out. All ~~recording~~hourly and ~~non-recording~~ daily stations for Reckershausen are shown in Fig. 2.

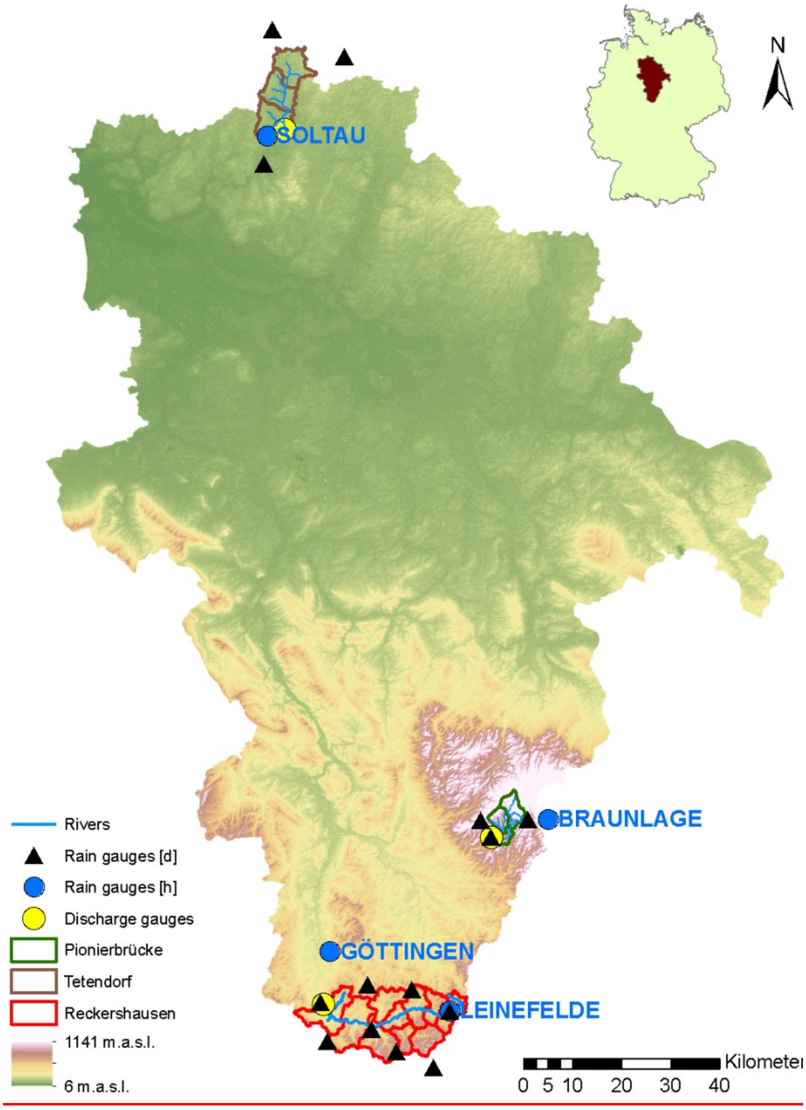

Fig. 1. Location of all three catchments in the Aller-Leine-river basin and its location in Germany.

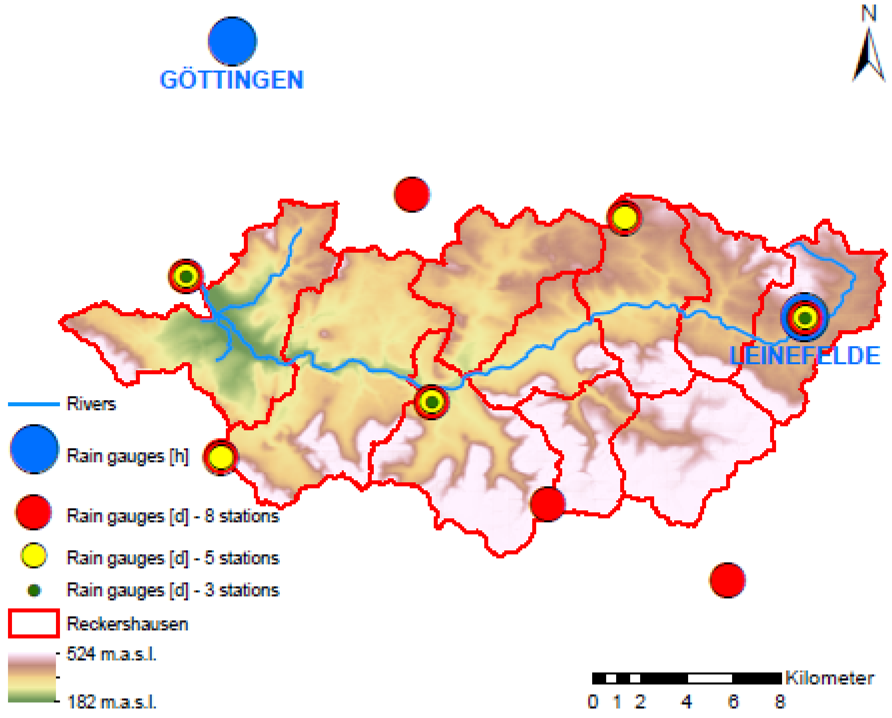

Fig. 2. Catchment Reckershausen including sets of 3, 5 and 8 ~~non-recording~~ daily stations used for network density analysis.

The catchments differ concerning area and elevation as well as land use and soil conditions. A brief description can be found in Table 1. The soil information is extracted from the soil map BÜK1000 of the Federal Republic of Germany with a scale of 1:1,000,000 (Hartwich et al., 1998). Information regarding the land use is extracted from the CORINE database (Federal Environment Agency, 2009). The time of concentration has been estimated as per Kirpich (1940).

Table 1. Brief description of the investigated catchments with fraction of dominant soil type and land use

Catchment	River	Area [km²]	Sub-catchments	Time of concentration [h]	Dominant soil type	Dominant land use
Pionierbrücke	Sieber	44	2	1.8	Spodic Cambisols (77 %)	Coniferous forest (81 %)
Tetendorf	Böhme	110	3	7.2	Haplic Podzols / Dystric Regosols (40 %)	Non-irrigated arable land (39 %)
Reckers-hausen	Leine	321	10	7.4	Dystric Cambisols (37 %)	Non-irrigated arable land (59 %)

2.2 Climate data

For the rainfall disaggregation, time series of ~~recording~~ hourly and ~~non-recording~~daily stations are required. Time series of the ~~recording~~ hourly stations are used for the parameter estimation of the cascade model (described in section ~~chapter~~ 3.1 a), which is in turn used for the disaggregation of the time series of the ~~non-recording~~daily stations. An overview of rain gauges used in this study is given in Fig. 1 while their measuring periods in Tab. 2~~An overview of rain gauges and measuring periods used in this study is given in Fig. 1 and Table 2.~~ For the ~~non-recording~~daily stations, the chosen period is the longest available period with data for all stations in a catchment. From Table 2 it can be seen, that time series lengths have a longer duration for ~~non-recording~~daily stations in comparison to those from ~~recording~~ hourly stations for all catchments (up to 2.7 times for Pionierbrücke). Additionally, the number of ~~non-recording~~daily-stations is higher.

Table 2. Rain gauges and time series lengths used for each catchment

Catchment	Type	Rain gauges	Start	End
Pionierbrücke	Daily	3	1950	2004
	Hourly	1	1993	2013
Tetendorf	Daily	3	1984	2006
	Hourly	1	1993	2013
Reckershausen	Daily	8	1972	2006
	Hourly	2	1993	2013

~~Catchment~~	~~Rain gauges~~	~~Type~~	~~Start~~	~~End~~
~~Pionierbrücke~~	~~Rehberger Grabenhaus~~	~~Non-recording~~	~~1950~~	~~2004~~
	~~Sieber~~			
	~~Hauskuehnenburg~~			
	~~Braunlage~~	~~Recording~~	~~1993~~	~~2013~~
~~Tetendorf~~	~~Bispingen-Huetzel~~	~~Non-recording~~	~~1984~~	~~2006~~
	~~Fallingbostel~~			
	~~Schneverdingen~~			
	~~Soltau~~	~~Recording~~	~~1993~~	~~2013~~
~~Reckershausen~~	~~Wachstedt~~	~~Non-recording~~	~~1972~~	~~2006~~
	~~Leinefelde~~			

~~Heiligenstadt-Kalteneber~~			
~~Reinholterode~~			
~~Uder~~			
~~Bornhagen~~			
~~Friedland (Lower Saxony)~~			
~~Gleichen-Elbickerode~~			
~~Göttingen~~ ~~Leinefelde~~	~~Recording~~	~~1993~~	~~2013~~

For the rainfall-runoff-model HBV (see section 3.2), time series of precipitation, temperature and potential evaporation are needed. The following description of data processing of temperature and potential evaporation is based on Wallner et al. (2013) and was carried out for the whole Aller-Leine basin. The temperature time series were derived through an interpolation using External Drift Kriging of 38 ~~recording~~ hourly stations with hourly resolution, whereby the additional information is elevation.

The calculation of the potential evaporation is carried out using the Turc-Wendling method on a daily basis (DVWK, 1996). The required sunshine duration per day was derived through Ordinary Kriging using 29 stations. To achieve an hourly resolution, daily values have been divided by 24, since the inter-daily distribution of potential evaporation has been shown not to be that sensitive as model input. Different land use types have been taken into account by using an average land use parameter (DVWK, 2002) similar to the crop coefficient. All input data were interpolated and subsequently aggregated to subcatchment scale.

For the WaSiM model, which is applied only for the catchment Pionierbrücke, climate time series are needed as point or gridded information on an hourly basis. From the climate station Braunlage, time series of temperature, relative air humidity, and wind speed are available with an hourly resolution. Global radiation was only available on a daily basis, but has been disaggregated to hourly values using an approach as ~~per~~ in Förster et al. (2016).

2.3 Runoff data

The available discharge data of the three catchments is listed in Table 3. While observed hourly time series are only available since 2000 (Pionierbrücke) and 2004 (Tetendorf and Reckershausen), observed extreme values exist for much longer periods. Daily discharge time series exist for at least as long as the period of the ~~monthly~~ hourly extreme values on a monthly basis.

For the calibration, a special focus is given to the extreme values of the summer (01.05.-31.10.) and winter period (01.11.-30.04.). Therefore, the maximum observed value of each half year were extracted from both data sources, observed hourly time series and monthly extreme values, to generate periods as long as possible.

Table 3. Available periods of runoff data types

Catchment	Hourly discharge time series	Daily discharge time series	Monthly extreme values
Pionierbrücke	2000-2013	1929-2006	1952-2005
Tetendorf	2004-2013	1986-2000	1986-2000
Reckershausen	2004-2009	1964-2006	1974-2005

3. Methods

The method section chapter consists of two subsectionschapters. In the sub sectionchapter 3.1, the multiplicative cascade model for the disaggregation of rainfall time series is explained. Additionally, two methods for the implementation of spatial consistence in the disaggregated time series are presented. The descriptions of the two rainfall-runoff models HBV and WaSim and the calibration procedure for HBV can be found in subsectionchapter 3.2.

3.1 Rainfall generation

a) Rainfall disaggregation

The most important input for rainfall-runoff models are long and high-resolution rainfall time series from a dense rain gauge network. Data of that kind is not available in most cases. However, time series from non-recording stations (e.g. daily resolution) are often available for long periods and with a high spatial density. The objective of the disaggregation is to generate high-resolution time series from the time series of the non-recording stations with information from the observed, high-resolution data.

The multiplicative random cascade model (Müller and Haberlandt, 2015) is applied for the disaggregation of time series of the daily stations. A general scheme of this model is shown in Fig. 3. One coarse time step is divided into *b* finer time steps of equal length. The branching number b determines the number of finer time steps and is in the first disaggregation time step *b=3* and in all following disaggregation steps down to 1 h resolution *b=2*. The cascade-model is micro-canonical, so the rainfall amount of each time step is conserved exactly. A re-aggregation of the disaggregated time series yield the observed time series used for the disaggregation. Since the focus of this investigation is not on the disaggregation itself, only a brief description of the model is included below. For a more detailed explanation, the interested reader is referred to Müller and Haberlandt (2015) for a more detailed explanation,. However, the main results are a slight underestimation of dry spell duration (relative error of -6 %), fraction of dry intervals (-3 %), wet spell duration (-12 %) and amount (-9 %), while average intensity is slightly overestimated (4 %). While the autocorrelation function also shows underestimations, the extreme values are well represented.

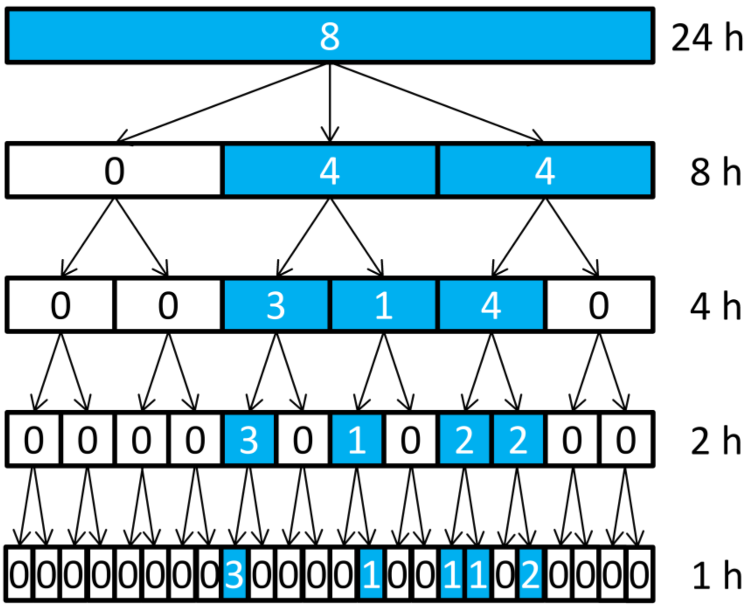

Fig. 3. General disaggregation scheme of the applied multiplicative cascade model (values inside the boxes represent rainfall amount, blue or white box color indicates wet or dry time steps, respectively)

~~All parameters can be extracted directly from the observed, high-resolution time series by a reverse application of the cascade model (Carsteanu and Foufoula-Georgiou, 1996). For the parameter estimation, the position and volume class of the wet time step to disaggregate is taken into account. The position describes the wetness state of the preceding and succeeding time steps, so that four states can be distinguished: starting, enclosed, ending, and isolated position. For each position class, two volume classes are distinguished, whereby the arithmetic mean of all rainfall intensities for each position is chosen as the threshold between the lower and upper volume class. The structure of different position and volume classes in the observed time series and its conservation during the disaggregation process was found before by e.g. Buishand (1977), Olsson (1998), Rupp et al. (2009) or Thober et al. (2014). Only for the first disaggregation step (from 24 h to 8 h) is no position taken into account (Müller and Haberlandt, 2015). While a uniform distribution is used for the first disaggregation step for the splitting of the rainfall amount on the number of as wet identified time steps, for *b=2* empirical distribution functions are used. For the sake of completeness it should be mentioned here that an unbounded cascade model has been used (Marshak et al, 1994). For all disaggregation steps with *b=2*, the same parameter set is applied due to the findings of scale invariance over these temporal resolutions (see Veneziano et al. (2006) and references therein).~~

b) Bivariate characteristics

For the definition of spatial consistence applied in this study the bivariate rainfall characteristics are following the ones used by Haberlandt (2008) and are briefly described in the following:

~~The disaggregation of single time series is carried out without taking into account time series of surrounding stations. For each time series the cascade model distributes the wet time steps randomly during a wet day due to its disaggregation scheme. Hence, spatial consistence of rainfall is underestimated after the disaggregation. Spatial consistence is defined in this investigation by bivariate spatial rainfall characteristics, the namely probability of occurrence, Pearson's coefficient of~~

~~correlation, and the continuity ratio (Wilks, 1998). These characteristics have been used before by e.g. Haberlandt (2008) and will be briefly described here:~~

The probability of occurrence $P_{k,l}$ describes the probability of rainfall occurrence at the same time at two stations *k* and *l*:

$$P_{k,l}\left(z_k > 0 \mid z_l > 0\right) \approx \frac{n_{11}}{n} \quad , \qquad (1)$$

where *n* is the total number of non-missing observation hours at both stations, z_i is the rainfall intensity and the number of simultaneous rainfall occurrence at both stations is represented by n_{11}.

The Pearson's coefficient of correlation ρ describes the relationship between simultaneously occurring rainfall at two stations *k* and *l* as a measure of the linear relation between both rainfall time series (Eq. (2)). Breinl et al. (2014) used this coefficient before for multisite rainfall generation:

$$\rho_{k,l} = \frac{cov(z_k,z_l)}{\sqrt{var(z_k) \times var(z_l)}} \text{, } z_k > 0, z_l > 0 \quad . \qquad (2)$$

Müller and Haberlandt (2015) found an intensity-dependency for Pearson's coefficient of correlation and distinguished between ρ(k≤4 mm) and ρ(k>4 mm), which is adopted here.

The continuity ratio $C_{k,l}$ compares the expected rainfall amount at one station for times with and without rain at the neighboring station (*E.* is the expectation operator):

$$C_{k,l} = \frac{E(z_k|z_k>0,z_l=0)}{E(z_k|z_k>0,z_l>0)} \qquad (3)$$

These characteristics are distance-dependent and prescribed values can be estimated as functions of the separation distance between two stations from observed data (see regression lines in Fig. 44 for each characteristic).

c) Implementation of spatial consistence

As mentioned before, the disaggregation of single time series is a point-process with no surrounding stations taken into account. Input ~~data sets~~ rainfall products for the rainfall-runoff models consisting of just the disaggregated time series without subsequent steps to implement spatial consistence are referred to as V1 (no implementation of spatial consistence). Two methods for the implementation of spatial consistence and resulting in the rainfall products~~data sets~~ V2 and V3, are applied in this investigation.

The first method, resulting in V2, is based on simulated annealing (Aarts and Korst, 1965, Kirkpatrick et al., 1983), a non-linear optimization method from the group of resampling algorithms. The aim of simulated annealing is to modify the disaggregated time series and in doing so minimizing ~~the~~ an objective function including the deviations between the observed bivariate rainfall characteristics and those from the disaggregated time series. ~~(Eq. 4)~~ Relative diurnal cycles are swapped without changing the structure of the time series or the absolute daily totals of rainfall amounts.~~, since exact~~

conservation is an advantage of the micro-canonical cascade model. The elements, which are swapped during the resampling process, are relative diurnal cycles. This enables the exact conservation of rainfall amount for each day. Swapping is only allowed between the same position and volume class (see chapter 3.1), which enables the structural conservation of the generated time series.

The interested reader is referred to Müller and Haberlandt (2015) for further details.The following description of the resampling algorithm is based on Müller and Haberlandt (2015) . Only a brief explanation is given here and the interested reader is referred to their investigation for further details.

One time series is chosen randomly from all disaggregated time series and used as the reference time series *R1* in the resampling process. A second time series *M1* is chosen randomly from the leftover disaggregated time series and will be modified during the resampling process. Relative diurnal cycles of *M1* are swapped with the aim to minimize the objective function:

$$O_{k,l} = w_1 \times (P_{k,l} - P^*_{k,l} *) + w_2 \times (\rho_{k,l,\le 4} - \rho^*_{k,l,\le 4}) + w_3 \times (\rho_{k,l,>4} - \rho^*_{k,l,>4}) + w_4 \times (C_{k,l} - C^*_{k,l}) \quad (4),$$

that summarizes the differences between prescribed values (indicated by *) and the actual values. The simulated annealing algorithm has the potential to leave local optima of the objective function by also accepting "bad swaps" (worsening of the objective function). Bad swaps are accepted with a certain probability based on the annealing temperature T_a, which is reduced during the resampling process. The fraction of bad swaps acceptance is decreasing during the resampling process. If the objective function does not improve after a certain number of iterations, the time series *M1* becomes an additional reference time series *R2*. In the next run, a new randomly chosen time series *M2* from the leftover disaggregated time series will be fitted to the two reference time series *R1* and *R2*. For the minimization of the objective function, the average value of $O_{k,l}$ of all station pairs is used. This procedure repeats, as long as there are leftover disaggregated, non-modified time series.

The second method, resulting in rainfall data setproduct V3, is a more pragmatic solution. It and was introduced by Haberlandt and Radtke (2014) and is also based on the already disaggregated time series of V1. . For each day, the station with the highest rainfall amount is identified. The relative diurnal cycle of this station is transferred to all other stations for this day. This parallelization is carried out for all days of the disaggregated time series. The varying diurnal distributions of rainfall at each station without spatial patterns, leading to an underestimation of spatial consistence, is instead transformed to a simultaneous occurrence of rainfall at all stations with an overestimation of spatial consistence.

Both methods are compared against using the disaggregated time series without any subsequent steps. For analyzes and discussion of the impacts of these methods, the designations listed in the summarizing Table 4 are used.

Table 4. Short characterization of the three rainfall productsdata sets

Starting point	Subsequent step	Rainfall occurence at different stations	Designation
Disaggregated	none	Random	V1

time series	Resampling	Intersecting	V2
	Parallelization	Simultaneous	V3

3.2 Hydrological Models

For analyzing the impact of rainfall ~~data sets~~ products with different spatial consistencies, two models, HBV-IWW (Wallner et al., 2013) and WaSiM (Schulla, 1997, 2015), are used. All simulations are carried out continuously. This enables the derivation of flood frequency analyzes and avoids uncertainties from unknown initial conditions resulting from event-based modeling (Pathiraja et al., 2016). Additionally, an initial phase of one year is used as a spin-up period to achieve plausible initial conditions for all storages.

a) HBV-IWW including calibration procedure

The HBV-IWW model is based on the HBV model that was originally developed at the Swedish Meteorological and Hydrological Institute (SMHI) in early 1970s (SMHI, 2008) and was modified by Wallner et al. (2013). HBV-IWW, for simplification titled HBV, is a conceptual model, where runoff generation and runoff transformation are represented by simple relationships between storage and effective precipitation, respectively runoff (see ~~Fig~~flow chart of the model in the supplementary material S1~~. 4~~). For the spatial discretization of the study areas subcatchments (see Fig. 2) with an approx. area of 20 km² are applied. It could be questioned, if a rainfall-runoff model with subcatchments is useful for the validation of the spatial consistence of rainfall. A daily station covers an area of 65 km² on average in Germany (Müller, 2016). This spatial resolution is not increased by the cascade model in this study, since only a temporal disaggregation is applied. Also, no additional information is gained by a model with higher spatial resolution. So the only disadvantage could be sort of numerical diffusion due to the spatial resolution. However, since subcatchments of this size are used throughout a number of studies, the HBV with this spatial resolution represents the state-of-the-art and is applied for the actual investigation.

For the estimation of the areal rainfall of each subcatchment, a two-step approach was chosen. First, rainfall is interpolated with a Nearest-Neighborhood approach on a raster basis with cell widths of 1 km. In the second step, areal rainfall for each subcatchment is calculated by an arithmetic mean of all raster cells within the subcatchment. If the areal rainfall of a subcatchment is dominated by one station, it could be questioned if areal rainfall intensities should be reduced (by e.g. areal reduction factors (Sivapalan and Blöschl, 1998, Veneziano and Langousis, 2005, Wright et al., 2013)) to avoid an overestimation (e.g. Peleg et al., 2018). Since also underestimations occur in the continuous simulation if this station was not in the center of the storm, no areal reduction was carried out.

Snow accumulation and melt is based by a threshold temperature and the degree day method. After the snow storage, all precipitation and snow melt is entering the soil storage where actual evaporation is considered. Depending on the state of the soil storage, water is released to the upper groundwater layer from where surface runoff and interflow can occur. Both are controlled by a storage coefficient. Water from the upper groundwater layer can also percolate to the lower groundwater layer. The outflow from the latter is representing the baseflow component. Surface runoff, interflow and baseflow are finally summarized and transformed via a triangular unit hydrograph. River routing is carried out via the Muskingum method. Further details about the model parameters (Table 5) can be found in Wallner et al. (2013) and in the supplementary material S2.

The calibration of rainfall-runoff models is traditionally carried out by modifying the model parameters so that the simulated hydrograph best represents the observed hydrograph (e.g. Beven, 2001). However this is not possible for disaggregated time series, since the rainfall is distributed randomly inside a day. For example, an observed peak flow can be reproduced perfectly in its magnitude in the simulated runoff time series, but will occur at a different time point. Hence, only runoff statistics can be used for the calibration. For this investigationFor the calibration, the following runoff statistics are used: quantiles of the distribution functions fitted to the extreme values of i) summer (Extr-Su, May to October) and ii) winter (Extr-Wi, November to April), iii) quantiles of the flow duration curve (FDC), and iv) monthly averages (Q-mon) are used for calibration. The calibration is carried out for each rainfall product separately, but for all 10 realizations.at the same time (resulting in one parameter set for 10 realizations) The calibration procedure is also illustrated in Fig. 4 the supplementary material S1..

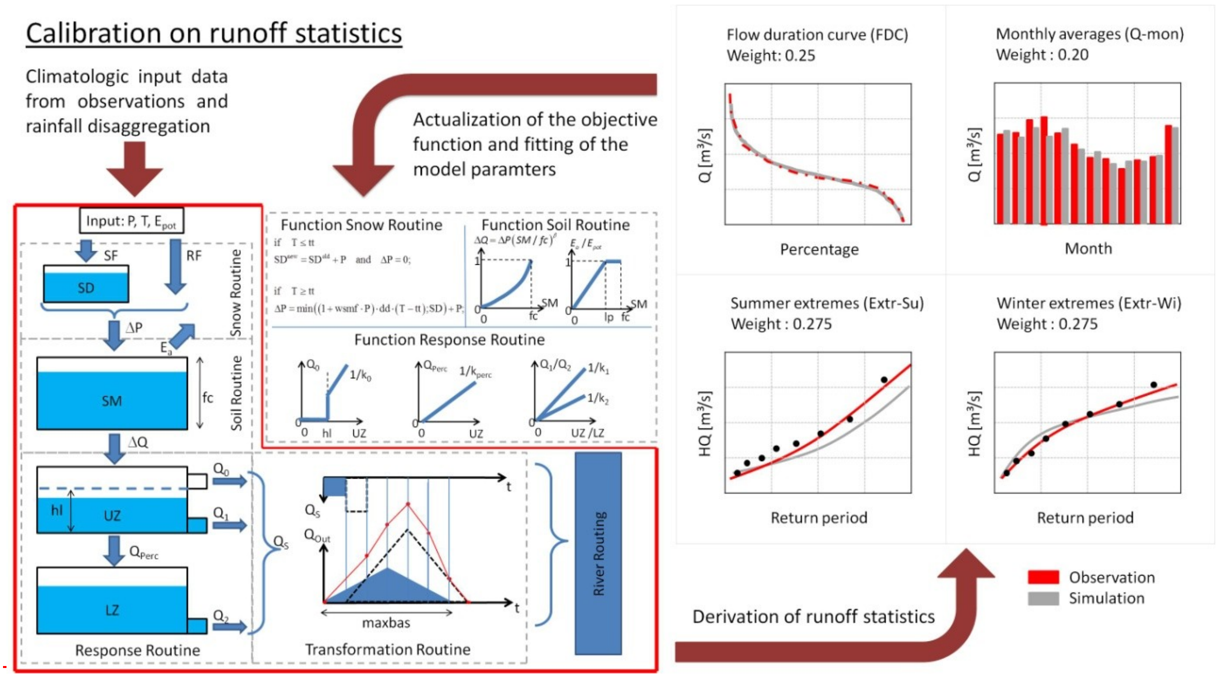

Fig. 4. Flow chart and applied calibration procedure for HBV (Wallner and Haberlandt, 2015)

For Extr-Su and Extr-Wi, a two-parametric Gumbel-distribution is fitted to the annual series of extreme values. L-moments are used for parameter estimation to reduce the sensitivity against outliers (Hosking and Wallis, 1997). Since Although extreme values occur only in a few time steps, their reproduction in the discharge time series is the main aim of the simulation on an hourly basis. However, since the extreme values represent only a small fraction of the discharge time series, also FDC and Q-mon of the simulated discharge time series are used to represent the more frequent discharge values. Q-mon accounts for the temporal dependency on the inter-annual variation of the discharge. FDC and Q-mon are calculated from averaged daily discharge values in order to reduce computation time.As mentioned before, the analyzes of FDC and Q-mon allows no direct validation of the rainfall products, but enables an overall plausible simulation of rainfall-runoff processes.

~~Hence,~~ FDC and Q-mon are calculated from averaged daily discharge values in order to reduce computation time.

For the goodness-of-fit analyzes of simulated (Sim) and observed (Obs) statistics, the Nash-Sutcliffe-efficiency NSE (Nash and Sutcliffe, 1970) is used. A perfect fit would result in *NSE=1*, while assuming the average of the observed data for all time steps would result in *NSE=0*. The equation for the *NSE* is given in Eq. 4~~5~~ and the corresponding quantiles for Extr-Su, Extr-Wi and FDC and months for the Q-mon, respectively, are given in Eq. 5~~6~~.

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{Obs}(t)-Q_{Sim}(t))^2}{\sum_{t=1}^{n}(Q_{Obs}(t)-\overline{Q_{Obs}})^2} \; \cancel{\rightarrow max}$$

(4~~5~~)

$$t = \begin{cases} \{0.05, 0.25, 0.5, 0.75, 0.95, 0.975\} & for\ FDC \\ \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} & for\ Q-mon \\ \{0.2, 0.5, 0.8, 0.9, 0.95, 0.98, 0.99\} & for\ Extr-Su\ and\ Extr-Wi \end{cases}$$

(5~~6~~)

The goodness-of-fit of all runoff statistics are summarized in the objective function O_{stat}, which should be minimized ~~used for~~ during the calibration:

$$O_{stat} = 1 - (0.275 \cdot NSE_{Extr-Su} + 0.275 \cdot NSE_{Extr-Wi} + 0.2 \cdot NSE_{FDC} + 0.25 \cdot NSE_{Q-mon}) \; \cancel{\rightarrow min}$$

(6~~7~~)

For the optimization simulated annealing is used. The parameters modified during the optimization with the corresponding ranges are given in the supplementary material S2~~Table 5~~.. The periods for calibration and validation are listed in Table 6 for each catchment.

~~**Table 5. HBV model parameters modified during calibration with limiting ranges**~~

~~Parameter~~	~~Unit~~	~~Explanation~~	~~Minimum~~	~~Maximum~~
~~*wsmf*~~	~~[mm⁻¹]~~	~~Wet snow melt factor~~	~~1~~	~~4~~
~~*tt*~~	~~[°C]~~	~~Threshold temperature~~	~~-1.5~~	~~1.5~~
~~*dd*~~	~~[mm °C⁻¹ d⁻¹]~~	~~Degree day factor~~	~~0.5~~	~~5~~
~~*fc*~~	~~[mm]~~	~~Field capacity~~	~~50~~	~~300~~
~~*lp*~~	~~[-]~~	~~Limit for potential evapotranspiration~~	~~0.1~~	~~0.95~~
~~*ß*~~	~~[-]~~	~~Empirical factor for runoff calculation from the soil layer~~	~~0.5~~	~~4~~
~~*hl*~~	~~[mm]~~	~~Threshold value for surface runoff~~	~~1~~	~~30~~
~~*k0*~~	~~[d]~~	~~Storage coefficient surface runoff~~	~~0.25~~	~~5~~
~~*k1*~~	~~[d]~~	~~Storage coefficient interflow~~	~~3~~	~~40~~
~~*k2*~~	~~[d]~~	~~Storage coefficient baseflow~~	~~50~~	~~500~~

~~*kperc*~~	~~[d]~~	~~Storage coefficient perculation~~	~~3~~	~~40~~
~~*maxbas*~~	~~[h]~~	~~Length of the triangular unit hydrograph impulse~~	~~3~~	~~10~~
~~*mx*~~	~~[-]~~	~~Weighting factor of Muskingum method~~	~~0.1~~	~~0.4~~
~~*mk*~~	~~[h]~~	~~Retention constant of Muskingum method~~	~~0.25~~	~~10~~

Table 6. Calibration and validation period for all catchments

Gauge	Calibration period		Validation period	
	Start	End	Start	End
Pionierbrücke	01.11.1952	31.10.1977	01.11.1977	31.10.2003
Tetendorf	01.11.1986	31.10.1993	01.11.1993	31.10.2000
Reckershausen	01.11.1974	31.10.1990	01.11.1990	31.10.2006

b) WaSiM

WaSiM (Schulla, 1997, 2015) is a physically based and distributed hydrological model which has been designed to study climate change and land-use change impacts on the water balance and floods in meso-scale catchments (e.g., Niehoff et al. 2002, Bormann and Elfert, 2010). WaSiM was formerly known as WaSiM-ETH, but has since been renamed (Schulla, 2015) and hence the new abbreviation is used throughout the manuscript. WaSiM is flexible regarding the resolution of spatial input data. In general, elevation, land-use, and soil data need to be prepared as gridded raster datasets. The spatial resolution of WaSiM applications covers several scales ranging from tens of meters to a few kilometers. For this study a spatial resolution of 150 m x 150 m was chosen.

For the areal rainfall estimation a combined Inverse Distance Weight (IDW) and elevation dependent regression approach is applied. This approach does not only account for a horizontal interpolation but also addresses the typically observed increase in precipitation with increasing elevation which proves helpful given that the catchment spans an altitudinal range of several hundred meters.

A set of alternative hydrological process representations for each of the following sub-models is included in the model in order to cover different user needs and meteorological data requirements: (i) evapotranspiration, (ii) snow, (iii) interception, and (iv) soil water. This list is not exhaustive since other processes can be also addressed using the model. Here, only the processes utilized in this study are described. Potential evapotranspiration is computed using the Penman-Monteith approach (e.g., Monteith, 1965) taking into account look-up tables of parameters defined for different land-use classes. Seasonal snow cover dynamics is simulated using a temperature threshold for phase partitioning and a temperature index model for snowmelt calculations. A bucket approach is applied to consider interception of rainwater. The soil water dynamics including actual evapotranspiration, infiltration, lateral outflow (interflow), and percolation is simulated in a numerical scheme which is based on the Richards equation. The lowermost nodes in each grid cell which are subject to saturation represent the groundwater storage in the model. A linear storage approach is applied here to simulate the outflow from the groundwater.

Since WaSiM is more complex than HBV with respect to computational needs, a different strategy for model calibration was chosen. As the number of both adjustable parameters and iterations is limited due to limited computational resources, a lexicographical approach was set-up for model calibration (Gelleszun et al., 2015). In this way, the optimization of parameters is divided into subsequent steps that are associated to different processes. In a first step, the parameters of the soil water balance and runoff generation (i.e. recession of hydraulic conductivity along the soil profile and the flow density) have been calibrated through maximizing *NSE*. Then, the baseflow recession is improved through minimizing the root mean square error *RMSE* of the lowermost part of the flow duration curve (two parameters). Both calibration steps have been performed using hourly meteorological time series and observed discharge time series from the period 2009-2012. As highly resolved meteorological observations are only available from 2000 onwards, an additional calibration step has been carried out using disaggregated rainfall time series in order to better match the long-term water balance characteristics through slightly modifying canopy resistance parameters of the evapotranspiration model. Without these pre-calibration steps an underestimation of the mean discharge and hence the water balance was identified. An incorrect representation of the water balance introduces other uncertainty sources which hence superpose the effects of the different versions of spatial rainfall. However, these pre-calibration was focused only on the water balance itself and not on the objectives used in Eq. (67).

4. Results & Discussion

For the discussion of the results, the section chapter is divided into two parts. The first part deals with the interpretation of the rainfall spatial variability, while the influence on simulated discharges is discussed in the second part.

4.1 Rainfall

For the disaggregation of daily rainfall time series to hourly values, the micro-canonical cascade model of Müller and Haberlandt (2015) is used. This model was previously validated in the beforementioned study for the the Aller-Leine river basin, which is also considered in this study. Since the focus of this investigation is the spatial variability of the generated rainfall, the interested reader is referred to their investigation for a detailed analyzes of point results. In their study, observed high-resolution time series have been aggregated to daily values and afterwards disaggregated back to hourly values, which enables comparisons between observed and disaggregated time series at the same location. The main findings for hourly resolution from Müller and Haberlandt (2015) are shown in Table 7 as relative errors *r* between disaggregated (*Dis*) and observed (*Obs*) time series over all investigated stations (*n*=9 in their case) :

$$r = \frac{1}{n} \times \sum_{i=1}^{n} \frac{(r_{Dis} - r_{Obs})}{r_{Obs}} \qquad (8)$$

Slight underestimation of dry spell duration (relative error of -6 %), fraction of dry intervals (-3 %), wet spell duration (-12 %) and amount (-9 %) can be identified, while average intensity is slightly overestimated (4 %). While the autocorrelation function also shows underestimations, the extreme values are well represented (not shown here).

Table 7. Relative error of rainfall characteristics resulting from the disaggregation model for the same investigation area from Müller and Haberlandt (2015)

Rainfall characteristic	Relative error r [%]
Wet spell duration [h]	-12
Standard deviation	-29
Skewness	-26
Wet spell amount [mm]	-9
Standard deviation	-18
Skewness	-19
Dry spell duration [h]	-6
Standard deviation	-7
Skewness	9
Fraction of dry intervals [%]	-3
Average intensity [mm/h]	4

In Fig. 45 the bivariate characteristics are shown for V1, V2 and V3 in comparison with the observations for Pionierbrücke (results for the other two catchments are in the supplementary material S3 and S4). For the V1 case (the disaggregated time series without any subsequent steps), the probability of occurrence and the correlation coefficients are underestimated, whereas the continuity ratio is overestimated.

For the V2 case, the probability of occurrence and the correlation coefficients could be improved. While values for the probability of occurrence and correlation coefficient for rainfall intensities > 4 mm are similar to observations, a slight underestimation can be identified for correlation coefficients for rainfall intensities ≤ 4 mm for some station pairs. For the continuity ratio, V2 results are varying. This is due to the definition of the criterion, taking into account station *k* with respect to station *l*, but not vice versa. This definition leads to different values for the same station pair, because different time steps are taken into account. Therefore, for $C_{k,l}$ an improvement can be identified during simultaneous worsening of $C_{l,k}$.

It should be noted that the resampling algorithm has not been validated in the context of distances smaller than 20 km and temporal resolution of 1 h. Although the spatial rainfall characteristics are underestimated after the disaggregation (V1), a major improvement can be identified moving all station pairs into the cloud of observations (except some of the continuity ratio).

The simultaneous rainfall of V3 leads to the best values for the continuity ratio, comparable to those from observations. However, slight overestimations can be identified for both coefficients of correlation. For the probability of occurrence, high overestimations can be identified (approximately 50 %). Although the same diurnal cycles are used for all stations, probability of occurrence is less than 1 due to the fact that rainfall does not necessarily occur at all stations on a wet day.

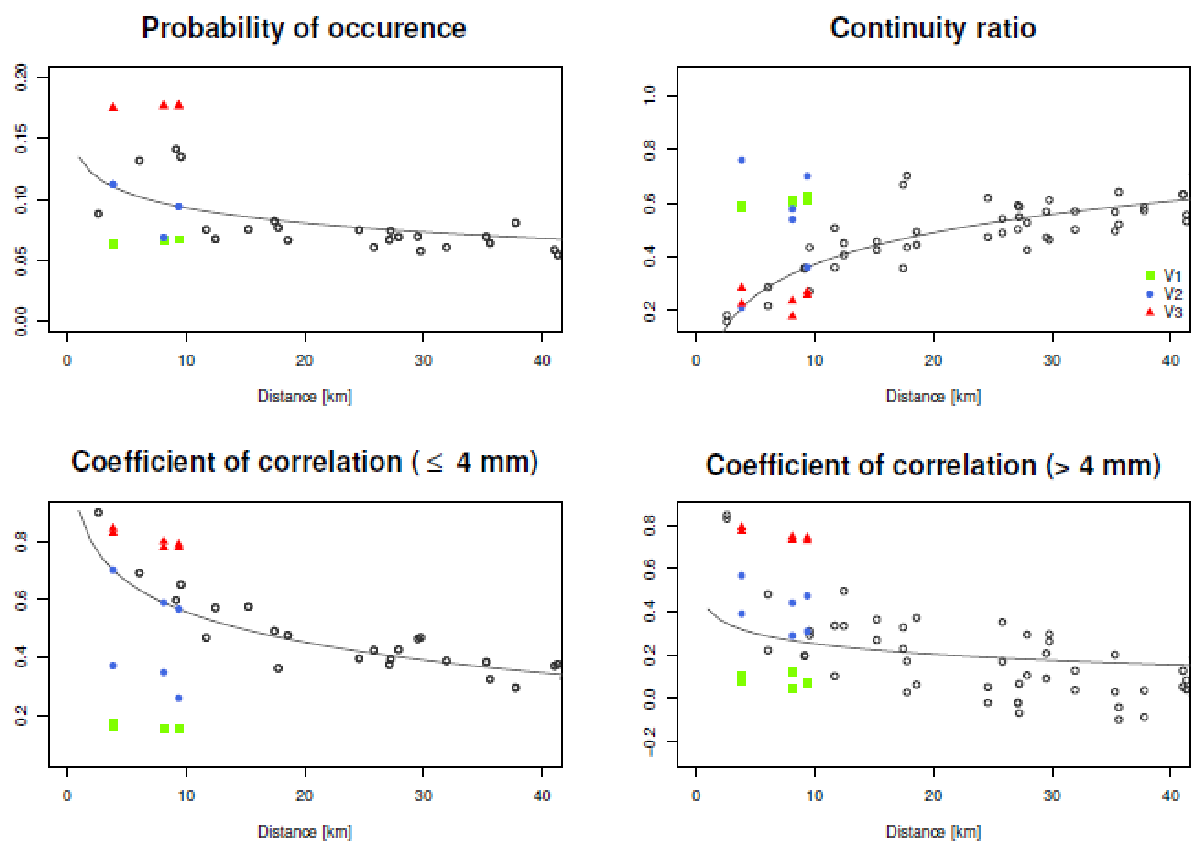

Fig. 4~~5~~. Bivariate spatial rainfall characteristics of V1, V2 and V3 in comparison to observations for the catchment Pionierbrücke (for one realization, black circles represent observations - for details the reader is referred to Müller and Haberlandt (2015)).

Additionally, the influence of the spatial consistence on resulting areal rainfall intensities is investigated. In ~~Fig. 6~~ the supplementary material S5, areal rainfall intensities resulting from V1, V2 and V3 are shown for one subcatchment of Pionierbrücke. Since only one observed high-resolution time series (Reckershausen: two) is available for each catchment, no comparison between areal rainfall intensities between observed and disaggregated time series (resulting from three stations for each catchment) can be carried out. Areal rainfall intensities resulting from disaggregated time series can only be compared among each other. V1 leads to the lowest rainfall intensities, V3 to the highest. Areal rainfall intensities of V2 lie between V1 and V2. The “random” rainfall occurrence in V1 leads to smaller rainfall intensity values as was indicated by the probability of occurrence (see Fig. 5). Accordingly, the parallelization of V3 leads to the highest areal rainfall intensities. Therefore, the results for the spatial bivariate characteristics and the areal rainfall intensities are consistent. The findings are similar for the other subcatchments in Tetendorf and Reckershausen.

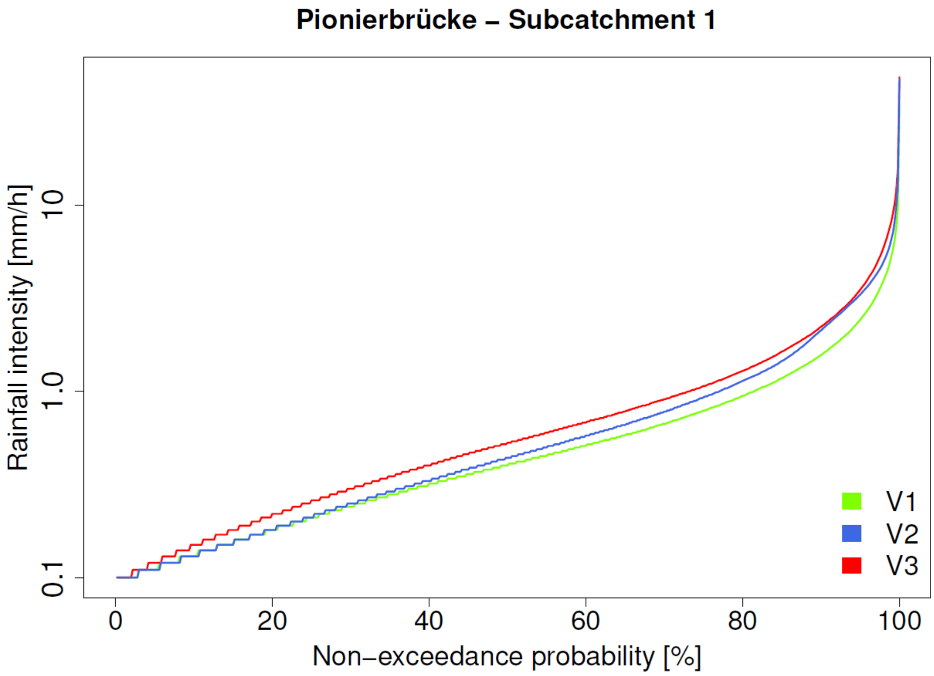

Fig. 6. Non-exceedance curve of areal rainfall intensities for V1, V2 and V3 for one subcatchment of Pionierbrücke (for one realization)

Additionally, the extreme values of the areal rainfall intensities have been analyzed, since those can have a significant influence on the resulting runoff. In Fig. 57, the annual rainfall extremes for another subcatchment in Pionierbrücke are illustrated using the Weibull-plotting position (similar for all subcatchments). As identified for all areal rainfall intensities, V1 also leads for the extreme values to the lowest values for each return period. V2 and V3 result in similar values for return periods < 18 years. The clear difference of higher values for V3 over the whole spectrum of non-exceedance probability from Fig. 6 cannot be identified for the extreme values (see S5). Although for V3, where the diurnal cycle of the station with the highest daily rainfall amount is transferred to the time series of all other stations, V3 does not lead to the highest extreme values. The reason for this is that the highest daily rainfall amount does not necessarily lead to the highest rainfall intensity on the final disaggregation level with an hourly time step. As an example, a rainfall station A with a daily total rainfall amount of 50 mm has a maximum intensity during this day of 8 mm/h, whereas station B with a daily total rainfall of 40 mm has a higher maximum intensity of 15 mm/h. As such, V3 can also lead to a smoothing of the rainfall intensities, at least for peak intensities. However, for higher return periods (> 18 years), V3 leads to higher values than V2.

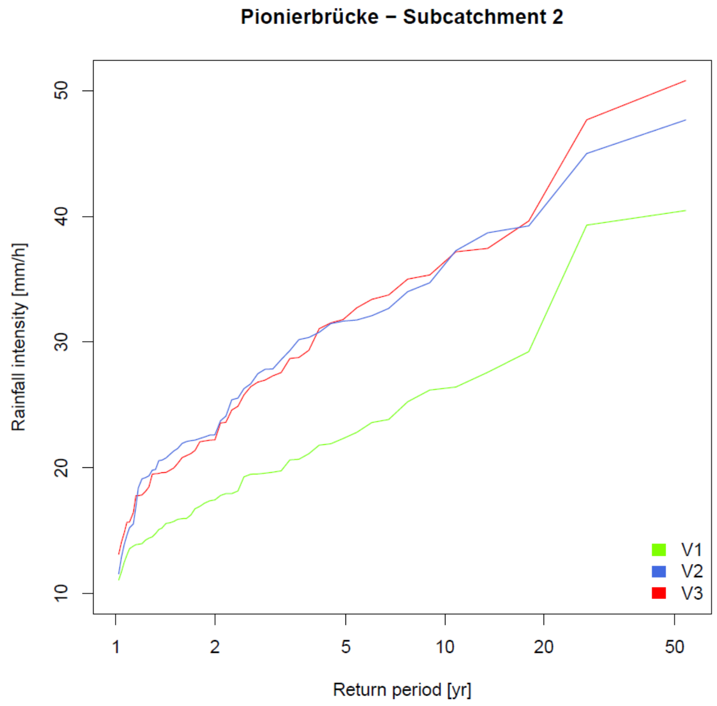

Fig. 5~~7~~. Annual rainfall extremes of the areal rainfall intensities for subcatchment 2 in Pionierbrücke (for one realization, based~~)~~ on n=53 annual extreme values from 01.11.1950-31.10.2003).

It can be summarized that V1, V2 and V3 lead to different results regarding spatial characteristics and areal rainfall intensities

4.2 Rainfall-runoff-model results

In this section~~chapter~~, all rainfall-runoff simulation results are presented. The section ~~chapter~~ is organized as follows: in a) the rainfall runoff-model results using HBV are shown for all catchments for V1, V2 and V3 with three rain gauges as input for each. In b) HBV-model results for different station densities for catchment Reckershausen are presented. HBV-model results without parameter calibration are shown for all catchments in c), while WaSiM-model results are presented in d) for catchment Pionierbrücke.

a) HBV- simulation results with calibration using three rain gauges as input

The parameterization was carried out by a split-sampling with a calibration and validation period for each catchment. The results for Reckershausen, Pionierbrücke and Tetendorf are shown in Fig. 6~~8~~, 8~~10~~ and 9~~11~~ for the calibration period. For Reckershausen, only results using three rain gauges as input are shown here. For Extr-Su and Extr-Wi, flood quantiles are shown for a return period of 100 years. However, the extrapolation is limited by the length of the simulated runoff time series. As per Maniak (2005), a maximum return period of three times the runoff time series length should be used to avoid too high statistical uncertainties caused by extrapolation. This results in 75 years for Pionierbrücke, 21 years for Tetendorf and 45 years for Reckershausen. The discussion of the results is limited to those and more frequent return periods. For a quantitative analysis, *NSC*-values for all criteria and for each catchment are given in Table 8. As mentioned before, *NSC*-values are based on a few supporting points (see Eq. 56). Also, theoretical Gumbel-distribution functions with two parameters are compared, which can be similar although the population used of each distribution function are different. Hence, values of 0.99 or even 1.00 can be achieved. On the other hand, small deviations from the observations can lead to even negative *NSC*-values (see e.g. the discussion of the simulation results for Reckershausen).

For Reckershausen, the Extr-Su and Extr-Wi are similar to those from observations (Fig. 68). While for summer all observed flood quantiles are within the range of Extr-Su ($0.99 \leq NSC \leq 1.00$), for Extr-Wi a slight overestimation occurs for V2 and V3.

For the validation period, flood quantiles for both, Extr-Su and Extr-Wi, are overestimated. The overestimation is higher in winter (approx. 20 m^3/s for HQ_{50}) than in summer (approx. 10 m^3/s). One possible cause can be the higher yearly maximums in the calibration period. It is assumed that parameters, calibrated to achieve high floods, tend to generate larger discharges even if lower yearly maxima are observed. This is also indicated by the results for FDC and Q-mon. Although both are represented well in the calibration period ($0.88 \leq NSC_{FDC} \leq 0.90$, $0.96 \leq NSC_{Q\text{-}mon} \leq 0.99$), both criteria are overestimated in the validation period ($0.57 \leq NSC_{FDC} \leq 0.63$, $0.81 \leq NSC_{Q\text{-}mon} \leq 0.89$). In the validation period the range and hence the uncertainty for both, Extr-Su and Extr-Wi, is smaller for V2 and V3 in comparison to V1.

The simulation results of Extr-Su of the validation period for the catchment Reckershausen show the sensitivity of the *NSC* as a goodness-of-fit criterion. V1 and V3 lead to positive *NSC*-values (0.60 and 0.31), while V2 leads to a negative value of NSC=-0.05. However, from a visual inspection (see Fig. 79), differences between all three approaches are small and less intense as one might expect from the *NSC*-value itself. The high sensitivity of the *NSC* makes a direct interpretation of its values more difficult (Schaefli and Gupta, 2007, Criss and Winston, 2008). However, for the calibration process, a high sensitivity leads to an improvement of the simulation results.

Values for the objective function are given in Tab. 9. For Reckershausen, the objective function values are very similar for V1, V2 and V3 for both, calibration and validation period. Especially by taking into account that the value for the objective function depends on four *NSC*-values.

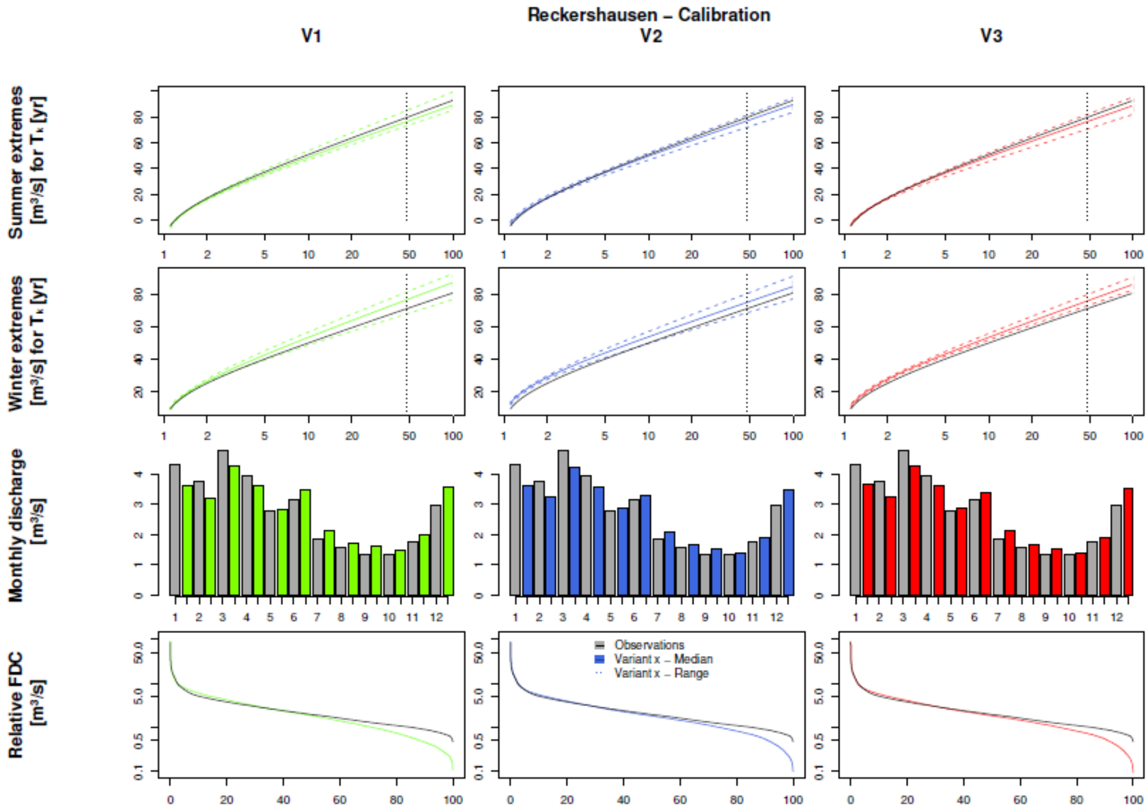

Fig. 68. Runoff simulation results with HBV for Reckershausen, calibration period

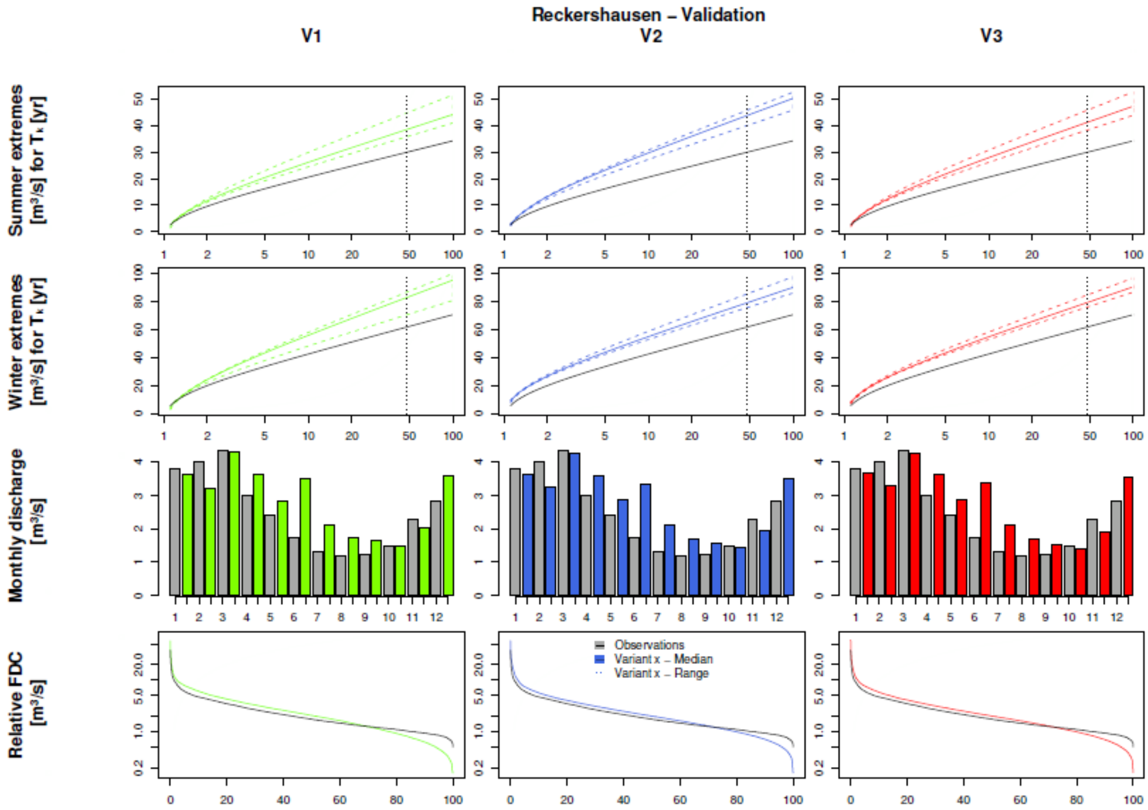

Fig. 79. Runoff simulation results with HBV for Reckershausen, validation period

For Pionierbrücke it should be mentioned that at points during the calibration (see the FDC in Fig. 810) and validation periods a simulated discharge of Q = 0 m³/s was obtained. Zero discharge implies that all storages have been emptied. This occurs only for Pionierbrücke and is due to the very steep conditions in the mountainous catchment (see Fig. 1) and hence the low soil depth and storage capacity. In the observed time series the minimum value is Q = 0.1 m³/s. The underestimation is as well caused by the selection of criteria selected for the objective function used for calibration. The main aim is to represent the extreme flows, while the shape of the intra-annual cycle of monthly average discharges and of the FDC are only implemented to achieve an overall realistic mean discharge behavior. For the FDC, four quantiles greater than 0.5 and only two quantiles smaller than 0.5 are used. Smaller quantiles are not of interest in these simulations, since discharge values in that range belong to dry periods with low flows, for which daily values of rainfall are sufficient for simulations and hence no rainfall disaggregation would be necessary. For the FDC, V3 leads to a slightly better fit to observations for non-exceedance probabilities smaller than 35 %, but to a worse fit between 35 % and 60 % non-exceedance probability. However, FDC is underestimated, independent from the applied rainfall data setproduct, for non-exceedance probabilities higher than 60 %. The underestimation identified by the FDC can also be identified for Q-mon in winter and in the underestimation of the Extr-Su and Extr-Wi. The results for the validation period are very similar and not shown here.

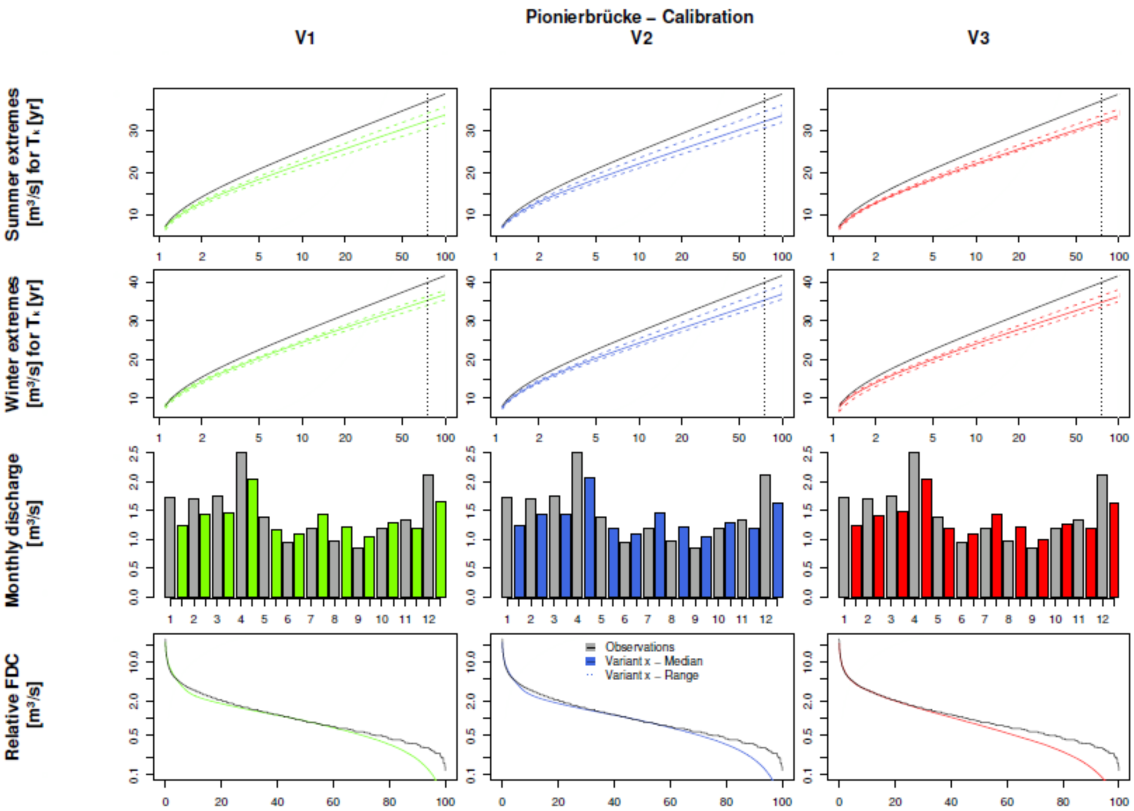

Fig. 810. Runoff simulation results with HBV for Pionierbrücke, calibration period

On the contrary, for Tetendorf FDC and Q-mon (except September and October) are overestimated by all rainfall data sets products (Fig. 911). However, for Q-mon the shape of the intra-annual cycle is well-represented. For the extreme values it should be mentioned again, that the analyzes are only valid for return periods more frequent than 21 years. For Extr-Su, underestimations occur for return periods more frequent than 5 years for all variants in the calibration period (less than 2 years in the validation period). For Extr-Wi, the median of V1 represents the observed values well, while for V2 and V3 the median leads to overestimations for return periods frequent than 5 years. However, observations are still in the range of the simulation results, whereby the range is wider for V1 and V3 in comparison to V2. In total, the resampling in V2 leads to a reduction of the overestimation of the observed summer extreme values, but to a stronger overestimation for winter extremes in comparison to V1 and V3.

Since for Tetendorf seasonal differences regarding V2 were identified, the spatial rainfall characteristics of the objective function applied for the resampling process have been re-analyzed, differing between the summer and winter half years. The results regarding both periods as well as the estimation over the complete year are shown in Fig. 10 for all bivariate spatial rainfall characteristics based on all 24 hourly stations in Lower Saxony that have been used before for the estimation of these characteristics (Müller, 2016)~~are shown in Fig. 12 for all bivariate spatial rainfall characteristics~~. For the continuity ratio, probability of occurrence and both volume classes of correlation coefficients, differences can be identified, based on the different geneses of rainfall in summer and winter. The probability of rainfall occurrence is lower in summer due to a higher amount of convective rainfall events. However, the distance-dependent curve progression is very similar between the seasonal and annual estimated spatial characteristics. Since spatial characteristics are just moved closer to the regression line by V2 (without a perfect fit, see Fig. 4~~5~~), an improvement of the spatial rainfall characteristics by introducing slightly different season-dependent regression lines cannot be expected and is hence not applied.

As main reasons for the seasonal differences, the short validation and calibration period are considered. Short periods mean a small amount of days with rain and hence a small amount of relative diurnal cycles to swap during the resampling, limiting the ability of the algorithm to improve the spatial characteristics. The usage of time series of V2 as input for HBV and the additional short time for the calibration process lead to the seasonal differences.

For longer calibration and validation periods (Reckershausen and Pionierbrücke) the results for V1, V2 and V3 are very similar regarding the runoff statistics. An influence of the chosen method for the implementation of spatial consistence cannot be recognized.

Table 8. ***NSC*-values for all catchments and all criteria for calibration (Cal) and validation (Val) period**

Catchment	Criteria	V1		V2		V3	
		Cal	Val	Cal	Val	Cal	Val
Reckershausen	*Extr-Su*	0.99	0.60	1.00	-0.05	0.99	0.31
	Extr-Wi	0.97	0.43	0.97	0.58	0.97	0.58
	FDC	0.88	0.57	0.90	0.63	0.90	0.61

	Q-mon	0.96	0.81	0.99	0.89	0.98	0.85
Pionierbrücke	*Extr-Su*	0.89	0.95	0.88	0.91	0.89	0.94
	Extr-Wi	0.91	0.88	0.91	0.86	0.89	0.83
	FDC	0.61	0.17	0.61	0.16	0.61	0.17
	Q-mon	0.99	1.00	0.99	1.00	0.99	0.99
Tetendorf	*Extr-Su*	0.32	-0.79	0.68	0.78	0.21	-0.61
	Extr-Wi	0.87	0.70	0.64	-4.36	0.47	0.88
	FDC	0.79	0.82	0.84	0.65	0.71	0.78
	Q-mon	0.86	0.93	0.78	0.92	0.83	0.92

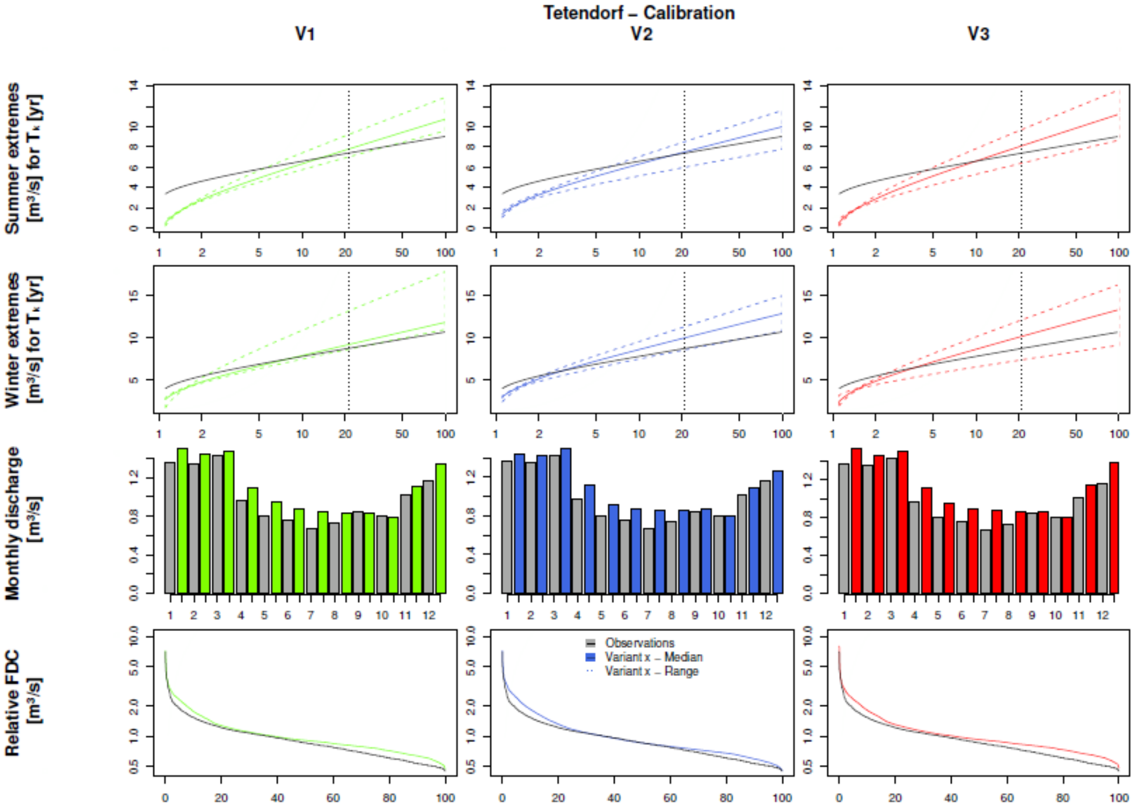

Fig. 9~~11~~. Runoff simulation results with HBV for Tetendorf, calibration period

Table 9. O_{stat}-values for all catchments and all criteria for calibration (Cal) and validation (Val) period

Catchment	**V1**		**V2**		**V3**	
	Cal	**Val**	**Cal**	**Val**	**Cal**	**Val**

Reckershausen	0.04	0.39	0.03	0.48	0.03	0.40
Pionierbrücke	0.13	0.21	0.13	0.23	0.14	0.23
Tetendorf	0.29	0.58	0.27	1.49	0.44	0.50

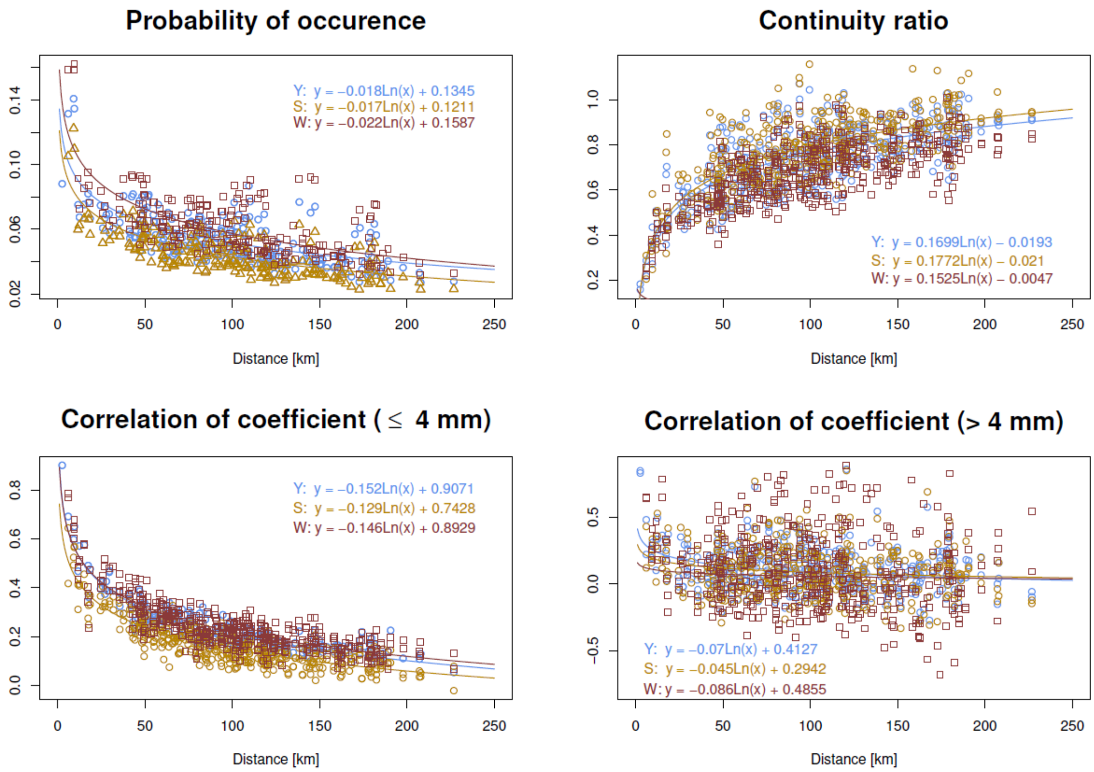

Fig. 10~~12~~. Bivariate spatial characteristics estimated for summer (S) and winter (W) seasonal as well as over the whole year (Y)

b) HBV- simulation results calibration using different numbers of rain gauges as input

A possible reason for the non-visible influence of the chosen method for the implementation of spatial consistence in the simulated runoff statistics is the low rain gauge network density. With a low network density, it is not possible to reflect the spatial rainfall variability and hence the influence of V1, V2 and V3 cannot be identified. The influence of the spatial rainfall variability on the runoff can only be determined by rainfall-runoff simulations.~~Investigations to the influence of spatial rainfall patterns can be found in Krajewski et al. (1991), Ogden and Julien (1993), Obled et al. (1994) and Nicotina et al. (2008).~~

Therefore, for Reckershausen, different numbers of rain gauges are applied for the calculation of the areal rainfall used as input for HBV. Areal rainfall is estimated by 3 rain gauges (representing a network density of 0.9 gauges per 100 km²) as carried out in a), 5 (1.6 gauges/100 km²) and 8 rain gauges (2.5 gauges/100 km²). The results are shown for V2 in Fig. 11~~3~~ for the calibration and in Fig. 13~~4~~. for the validation period. The results for V1 and V3 are very similar and not shown here. However, for a quantitative analysis the *NSC*- and O_{stat}-values are shown in Table 10 and Table 11.

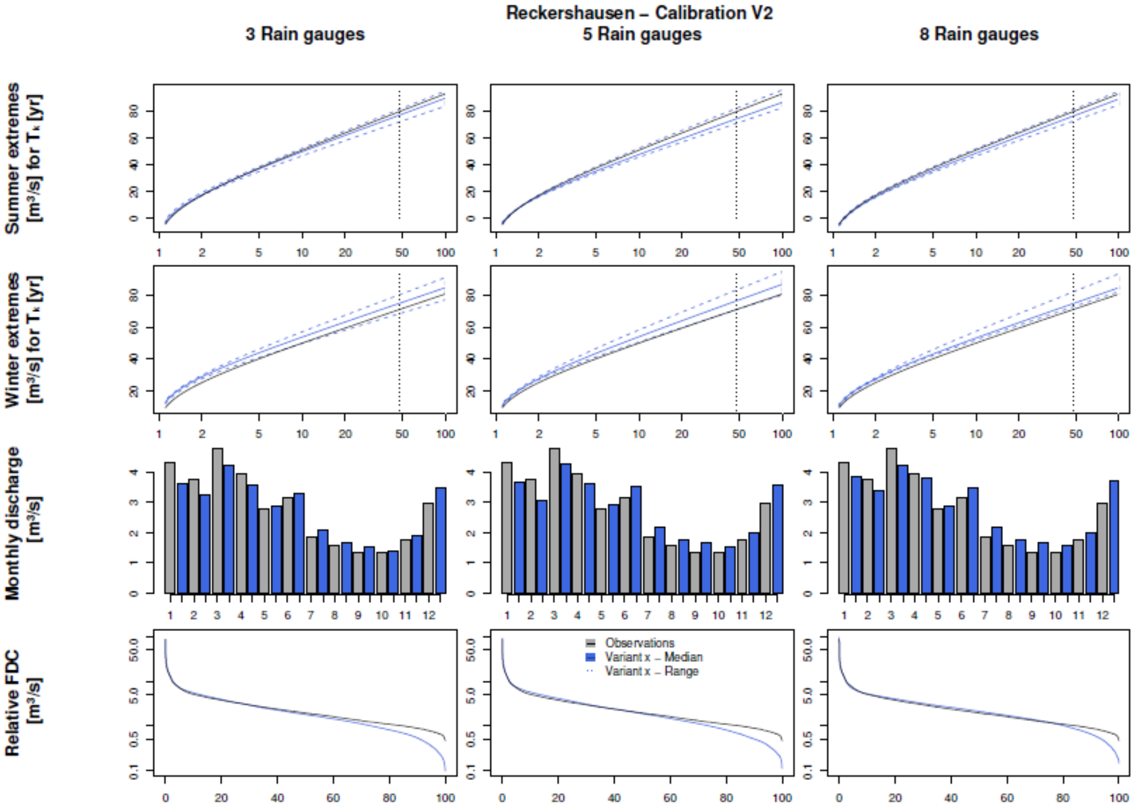

Fig. 113. Runoff simulation results for V2 with 3, 5 and 8 rain gauges with HBV for Reckershausen, calibration period

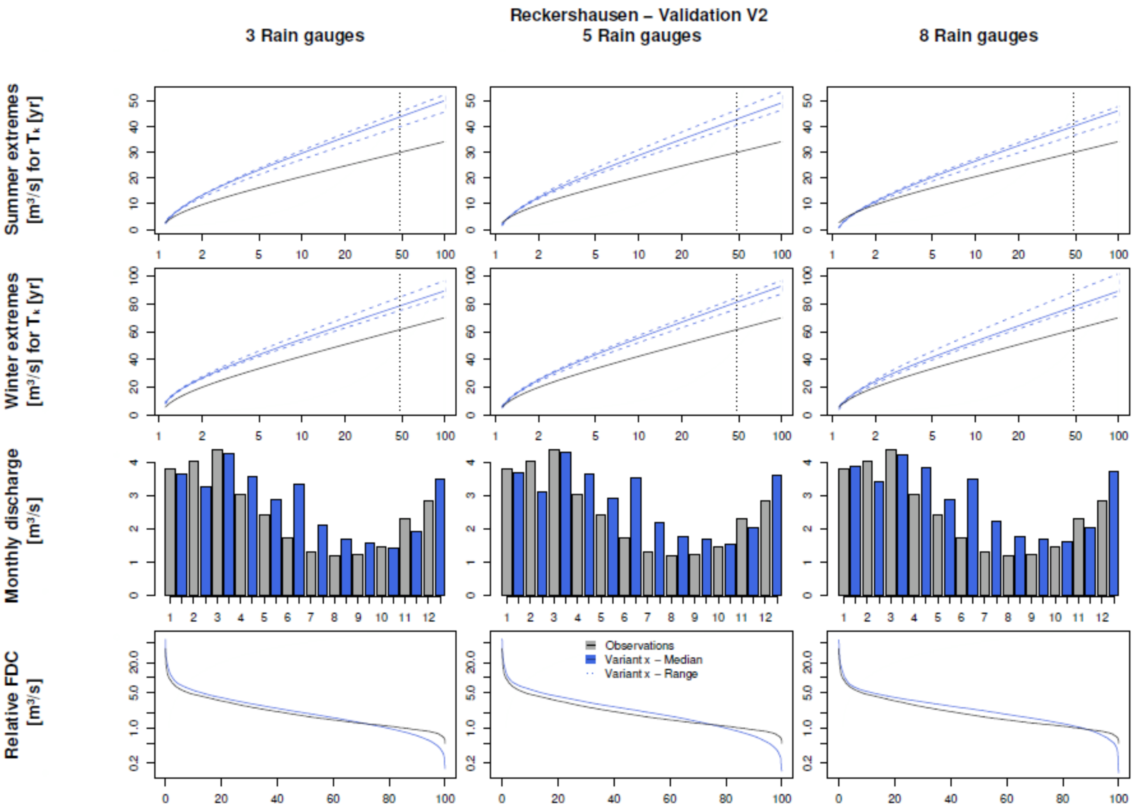

Fig. 124. Runoff simulation results for V2 with 3, 5 and 8 rain gauges with HBV for Reckershausen, validation period

Again, independent of the number of rain gauges used for the estimation of the areal rainfall, the results from the calibration period (Fig.1311) represent the observations better than those from the validation period (Fig. 1412). In the validation period, Extr-Su and Extr-Wi are overestimated as well as the majority of Q-mon and the FDC. Minor differences can be identified between the different rain gauge network densities, but no general conclusion is possible, e.g. the overestimation of Extr-Wi in the calibration period is increasing with an increasing network density. However, in the validation period the overestimation is decreasing with an increasing number of rain gauges from 3 to 8. Also for Q-mon or the FDC, no systematic improvement can be identified. This is an unexpected finding, because with the additional information from the daily total rainfall amounts, an improvement of at least the continuum characteristics was expected. Also for the *NSC*- and O_{stat}-values no systematical improvement can be identified: O_{stat}(V2, 3 rain gauges)= 0.03, O_{stat}(V2, 5 rain gauges)= 0.04, O_{stat}(V2, 8 rain gauges)= 0.03 (see Table 10 and Table 11).

It can be summarized, that the number of rain gauges has only a minor, but no systematic influence on runoff statistics for the catchments used in this investigation. ~~Investigated number of rain gauges were 3, 5 and 8, respectively 0.9, 1.6 and 2.5 rain gauges/100 km².~~ This contradicts conclusions from other studies. Seliga et al. (1992) recommend for spatial rainfall applications information every 5 km² (20 rain gauges/ 100 km²). So an improvement by an increasing station density up to this threshold should have been expected. For a French catchment with an area size of 71 km², Obled et al. (1994) investigated the influence of using 5 or 21 rain gauges, representing rain gauge network densities of 7 and 22 rain gauges/100 km². With 21 rain gauges Obled et al. improved their results significantly. Nevertheless, they conclude that the improvement is based on the better estimation of the total rainfall amount, not on its spatial distribution. Xu et al. (2013) investigated the influence of station density on a Chinese catchment with an area size of 94 660 km² and daily rainfall time series, hence a direct comparison of network densities is not possible. Nevertheless, they point out that the distribution of rain gauges inside the catchment is of importance. A distribution covering regions with different rainfall behaviors in a catchment can lead to better simulation results with only a few rain gauges in comparison to a less efficiently distributed network with more rain gauges. In the actual study, the rain gauges for each network density scenario have been selected in a way to cover the catchment area and its rainfall representatively (see Fig. 2). This could be one reason why an increase in rain gauge network density shows no systematic improvement in this study.

Table 10. *NSC*-values for all catchments and all criteria for calibration (Cal) and validation (Val) period

Number of rain gauges	Criteria	V1		V2		V3	
		Cal	Val	Cal	Val	Cal	Val
3	*Extr-Su*	0.99	0.6	1	-0.05	0.99	0.31
	Extr-Wi	0.97	0.43	0.97	0.58	0.97	0.58
	FDC	0.88	0.57	0.9	0.63	0.9	0.61
	Q-mon	0.96	0.81	0.99	0.89	0.98	0.85
5	*Extr-Su*	0.98	-0.24	0.98	0.09	0.99	-0.23

	Extr-Wi	0.97	0.68	0.96	0.48	0.98	0.65
	FDC	0.86	0.53	0.87	0.53	0.86	0.55
	Q-mon	0.99	0.91	0.98	0.86	0.99	0.91
8	*Extr-Su*	0.99	0.75	0.99	0.46	1	0.54
	Extr-Wi	0.96	0.62	0.98	0.64	0.97	0.59
	FDC	0.91	0.57	0.89	0.54	0.89	0.6
	Q-mon	0.99	0.88	0.99	0.94	0.98	0.88

Table 11. O_{stat}-values for all catchments and all criteria for calibration (Cal) and validation (Val) period

Number of rain gauges	V1		V2		V3	
	Cal	Val	Cal	Val	Cal	Val
3	0.04	0.39	0.03	0.48	0.03	0.40
5	0.05	0.51	0.04	0.49	0.04	0.51
8	0.04	0.28	0.03	0.34	0.04	0.33

c1) HBV-simulation results without calibration using three rain gauges as input

Another possible reason for the small differences between V1, V2 and V3 is the calibration of the rainfall-runoff model parameters for each of the rainfall data setsproducts. Parameters are allowed to vary between V1, V2 and V3, and hence damp the effects of the different degrees of spatial consistence. To exclude the calibration as a possible reason for the damping behavior, a calibration with a neutral rainfall data setproduct offering the same spatial rainfall coverage without giving preference to one of the investigated versions would be recommended. This would enable a direct comparison between V1, V2 and V3 without re-calibration of the models. Since high-resolution time series do not exist with the required spatial network density, radar data could be a possible solution. However, radar time series exist only for lengths which are too short for model simulations and subsequent derived flood frequency analyses.

To avoid re-calibrations, a pragmatic solution is chosen. The arithmetic means of the lower and upper limit for each parameter in Table 5 have been applied as a neutral parameter set. For the validation period simulation results based on this neutral parameter set have been analyzed. Although a splitting in calibration and validation period is not necessary if no calibration is carried out, comparisons are possible between the simulation results with and without calibrated parameters. The results are shown in Fig. 15 13 for Reckershausen, results are similar for Pionerbrücke and Tetendorf. For a quantitative evaluation *NSC*-values for all catchments are provided in the supplement material S6listed in Table 11 and O_{stat}-values in Table 12S7.

For Pionierbrücke and Tetendorf simulation results are worse without calibration (e.g. for Pionierbrücke, V1: $O_{stat,not\ calibrated}$=1.14 and $O_{stat,calibrated}$=0.21). For Reckershausen a slight improvement can be identified without calibration. The calibrated parameters led in the validation period to an overestimation of extreme values for both seasons as well as an overestimation of FDC and Q-mon (e.g. for V3: $O_{stat,not\ calibrated}$=0.28 and $O_{stat,calibrated}$=0.40). For all catchments, Extr-Su are underestimated by every version of spatial consistence. The Extr-Wi are also underestimated for Reckershausen and Pionierbrücke, but overestimated for Tetendorf. For all catchments, an intra-annual cycle of Q-mon can be identified. For Reckershausen, Q-mon is similar to observations, while for Pionierbrücke underestimations and for Tetendorf overestimations can be identified in winter. The FDC is not represented well for any of the catchments.

Although a neutral set of parameters has been applied, the differences in the simulation results between V1, V2 and V3 are still small. For Pionierbrücke the values of the objective function show the same range without and with calibration (1.10 (V2)≤$O_{stat,not\ calibrated}$=≤1.14 (V1) respectively

0.21 (V1)≤$O_{stat,calibrated}$≤0.23(V2, V3)). The similarity of the simulation results exists even if the model parameters are not calibrated and a neutral parameter set is used.

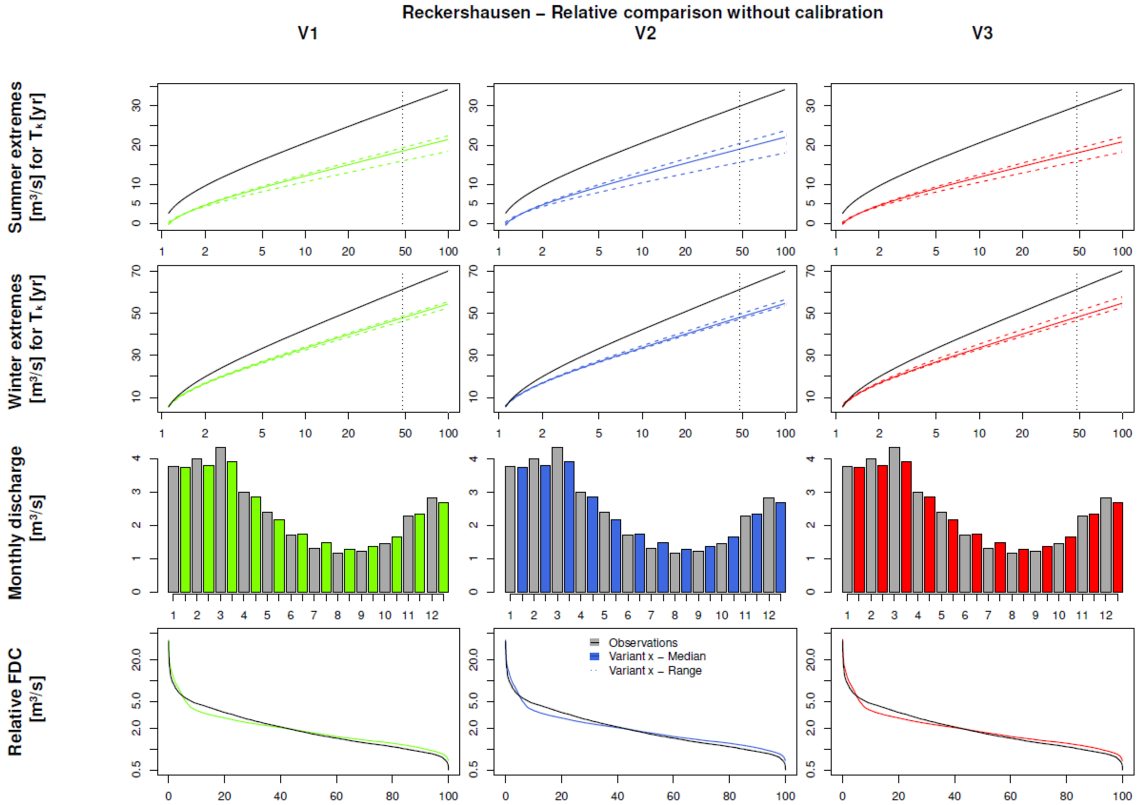

Fig. 135. Runoff simulation results with HBV without calibration for Reckershausen, validation period

Table 12. NSC-values for all catchments and all criteria without calibration for validation period

Catchment	Criteria	V1	V2	V3
Reckershausen	*Extr-Su*	0.20	0.26	0.14
	Extr-Wi	0.76	0.77	0.77
	FDC	0.97	0.97	0.97
	Q-mon	0.99	0.99	0.99
Pionierbrücke	*Extr-Su*	-1.68	-1.58	-1.59
	Extr-Wi	0.01	0.10	0.06
	FDC	-0.07	-0.07	-0.07
	Q-mon	0.96	0.96	0.96
Tetendorf	*Extr-Su*	0.52	0.54	0.55
	Extr-Wi	-7.78	-7.78	-8.41
	FDC	0.19	0.19	0.19
	Q-mon	-0.06	-0.05	-0.11

Table 13. O_{stat}-values for all catchments and all criteria without parameter calibration for validation period

Catchment	V1	V2	V3
Reckershausen	0.27	0.25	0.28
Pionierbrücke	1.14	1.10	1.11
Tetendorf	2.79	2.79	2.96

c2) WaSiM- simulation results without calibration using three rain gauges as input

For the comparison of V1, V2 and V3, WaSiM (Schulla, 1997, 2015) as a physically based and distributed hydrological model is used as an additional rainfall-runoff model. The application of more than one model increases the reliability of the simulation results and excludes the possibility of being model-dependent. As mentioned before, a pre-calibration of the rainfall-runoff model was carried out focusing only on the water balance itself and not on the objectives used in Eq. (7). Only the parameters mentioned in the model description (see 3.2.b) are calibrated. As far as possible, the same parameter values as in HBV in the uncalibrated case (c1) have been applied. The investigation with WaSiM is carried out only for catchment Pionierbrücke, since here the highest differences in simulation results are expected due to the short reaction time of the catchment.

The results are shown in Fig. 16 14 for the calibration period, Fig. 17 15 for the validation period and a quantitative analysis is given in Table 1412. For the calibration and the validation period Extr-Su and Extr-Wi are simulated slightly higher with V2 and V3 in comparison to V1. This is consistent with the areal rainfall extremes presented for Pionierbrücke in Fig. 57. In addition, the range for both criteria is higher for V2 and V3 in comparison to V1, whereby V2 leads to even wider ranges than V3 in some cases (e.g. Extr-Win the validation period). In this context it should be repeated, that a relative comparison is carried out and under- or overestimations are not points of interest. The *NSE*-values for both Extr-Su and Extr-Wi are very similar for V2 and V3 (e.g. $NSC_{Extr\text{-}Wi,Cal,V2}$=0.98and $NSC_{Extr\text{-}Wi,Cal,V3}$=0.99), but show differences to V1 ($NSC_{Extr\text{-}Wi,Cal,V1}$=0.90). Hence, in WaSiM a slight effect of the spatial consistence of rainfall is visible from the simulation results. Possible reasons for the differences are the spatial resolution (150 m x 150 m for each raster cell) and the IDW algorithm with an altitudinal rainfall adjustment, which was carried out by a linear regression model. However, for FDC and Q_{mon}, values for V1, V2 and V3 are again very similar. While for the calibration period the O_{stat}-values are similar for all rainfall data setsproducts, in the validation period the O_{stat}-values for V2 and V3 ($O_{stat, Val, V2}$=0.45 and $O_{stat, Val, V3}$=0.46) are much closer to each other than to V1 ($O_{stat, Val, V1}$=0.30).

Table 124. *NSC*- and O_{stat}-values for Pionierbrücke without parameter calibration using WaSiM

Criteria	V1		V2		V3	
	Cal	Val	Cal	Val	Cal	Val
Extr-Su	0.95	0.96	0.97	0.95	0.96	0.95
Extr-Wi	0.90	0.77	0.98	0.21	0.99	0.26
FDC	0.86	-0.15	0.87	-0.20	0.88	-0.27
Q-mon	0.99	0.99	0.99	0.99	1.00	0.99
O_{stat}	0.07	0.30	0.04	0.45	0.04	0.46

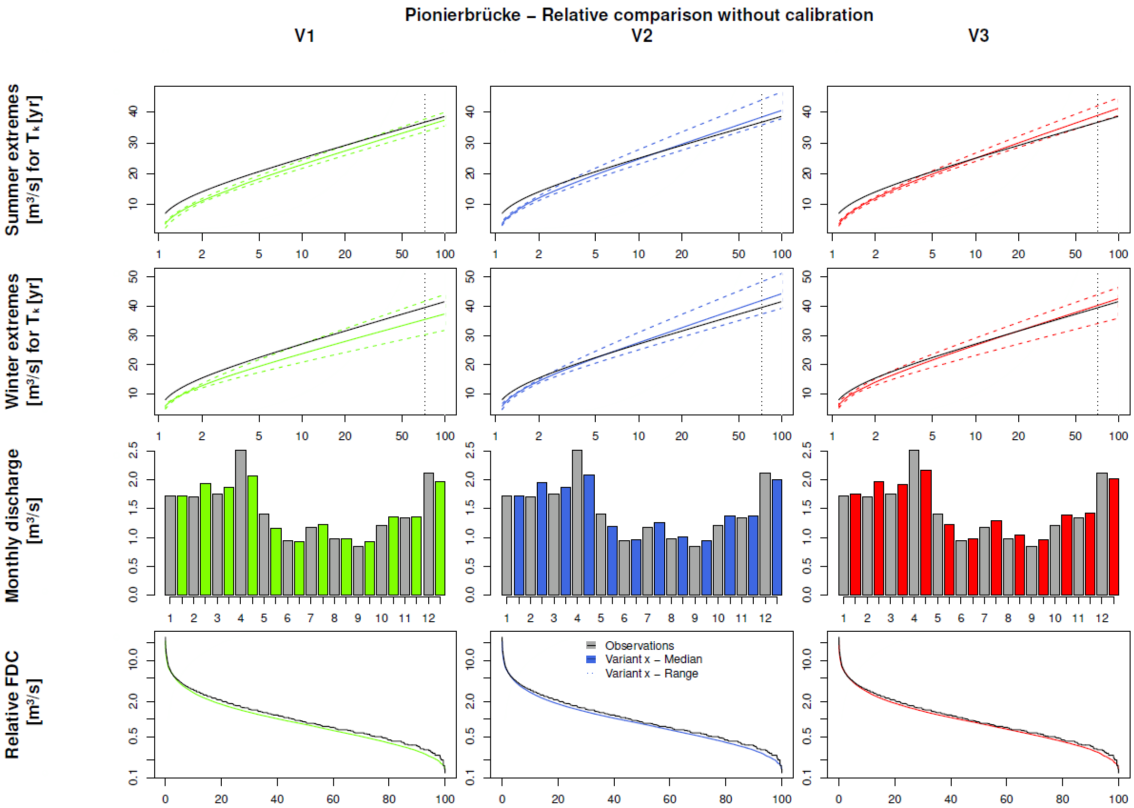

Fig. 146. Runoff simulation results with WaSiM without calibration for Pionierbrücke, calibration period

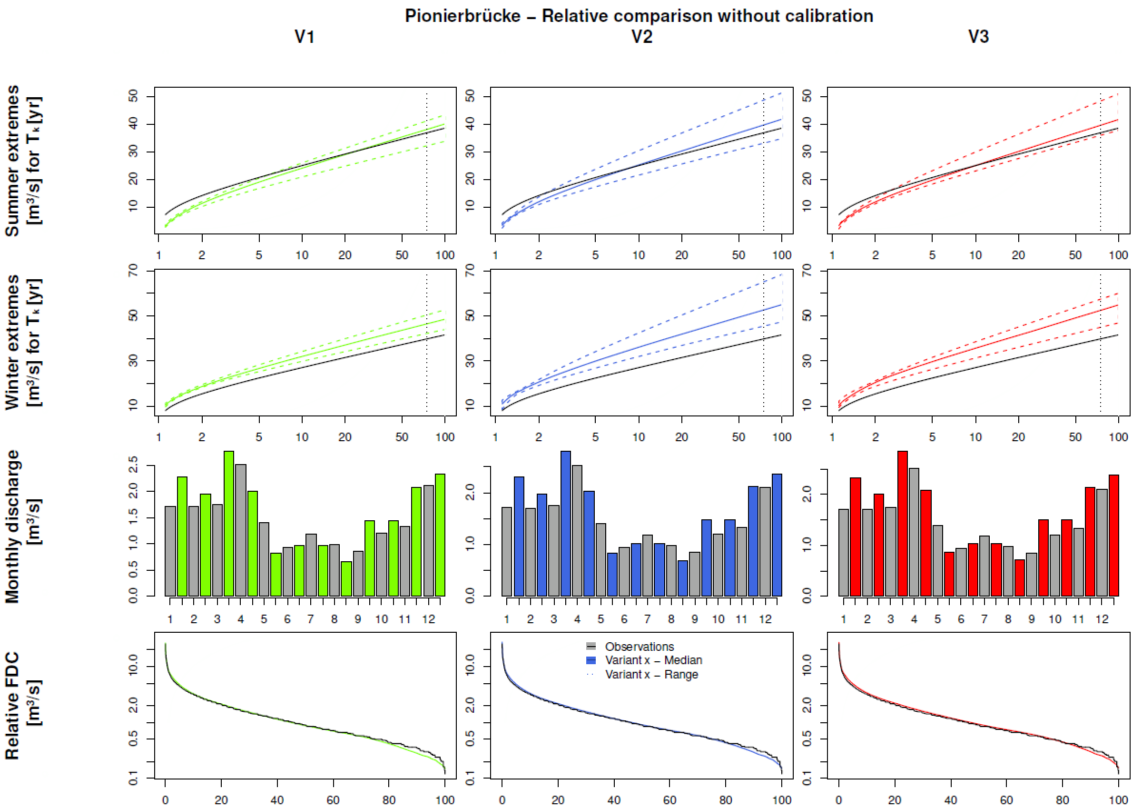

Fig. 15~~7~~. Runoff simulation results with WaSiM without calibration for Pionierbrücke, validation period

5. Discussion of rainfall-runoff simulation results

The rainfall-runoff simulation results with HBV after calibration of the parameters show that with all three rainfall data setsproducts, V1, V2 and V3, the Extr-Su and Extr-Wi, the FDC, and Q-mon can be represented with a comparable quality. The differences between the three methods are very small for the majority of all cases. Possible reasons for these small differences, which are discussed below, are:

- small differences between the three rainfall data setsproducts

- dampening of those differences by the calibration of the rainfall-runoff model parameters

- dampening behavior of the catchments

- choice of the rainfall-runoff model and its ability to represent differences of the three rainfall data setsproducts

Small differences between V1, V2 and V3 would lead to small differences in rainfall-runoff simulation results. However, the differences between the three methods are apparent. For the bivariate spatial characteristics (Fig. 45), the areal rainfall intensities (see S5Fig. 6) and the areal rainfall extremes (Fig. 57), differences can be identified among all three methods, which should as well be reflected by the runoff statistics results.

Another cause can be the separate calibration of the rainfall-runoff model parameters for each method. The applied calibration strategy has the capability to harmonize the different rainfall data setsproducts with the runoff statistics used for calibration. For the discussion of this harmonization effect, the simulation results for Reckershausen during the calibration (Fig. 1311) and validation periods (Fig. 1412) are used. During the calibration period, higher values for Extr-Su and Extr-Wi can be found in the observed runoff data. Hence, the parameters calibrated in this period tend to lead to higher runoff values. This is proven by the simulation results of the validation period with an overestimation of all runoff statistics. Only by the usage of an uncalibrated parameter set the calibration can be excluded from the list of possible causes.

The dampening behavior of the investigated catchments depends on the size and the concentration time of a catchment (Andres-Domenech et al., 2015). Also, catchments act as a filter, so that rainfall as an input signal is dampened during its transformation to runoff by several processes (e.g. interception, losses due to storage filling, transport processes). Mandapaka et al. (2009) have analyzed for (sub-) catchments of different sizes the runoff response from different rainfall scenarios with a total amount of 10 mm. For catchments with an area less than 10 km², a strong dependence of the duration, the intensity and the spatial distribution of the rainfall is identified. With increasing area size, the influence of these factors is reduced and for catchments with 1000 km², it is almost completely dampened. Since the catchment areas in the actual study range between 44 km² and 321 km², i.e. considerably larger than 10 km², this could be a possible reason why the differences in the runoff results are so small. On the other hand, the results of Seliga et al. (1992) and Obled et al. (1994) show that an increasing station network density lead to an improvement of rainfall information and hence should also lead to an improvement of the runoff simulation results. Ogden and Julien (1993) investigate the time of concentration of a catchment as an influencing factor for the rainfall-runoff processes. If the duration of a rainfall event causing flooding is shorter than the time of concentration, the spatial distribution of the rainfall is influencing the discharge at the catchment outlet. If rainfall events last longer than the concentration time, the influence decreases. However, Nicotina et al. (2008) identify an influence of spatial rainfall patterns only for catchments with areas > 1000 km^2, based on the travel time in the catchment. In the investigated catchments, the concentration time ranges from 1.8 h to 7.4 h, so the temporal and spatial variation should have an influence on the simulated discharges.In Müller and Haberlandt (2018) the rainfall products V1 and V2 and their influence on simulated discharge have been analyzed for 5-minute time steps in an urban hydrological context. Significant differences could be identified between the simulated runoff statistics resulting from V1 and V2 for their artificial sewage system.

Another reason could be the choice of the rainfall-runoff model. Obled et al. (1994) raise the question if it is possible with semi-distributed models to transfer the information of the spatial rainfall patterns into the simulated discharge time series. Obversely, if spatial rainfall patterns are necessary for rainfall-runoff simulations for a catchment with an area size of 71 km^2, as is used in their study, the spatial resolution of semi-distributed models may not be sufficient. Krajeski et al. (1991) also conclude that for the analysis of spatial problems, fully-distributed models may be more suitable and recommend those for further studies. Bárdossy and Das (2008) point out that with an increasing spatial resolution of the applied rainfall-runoff model, the sensitivity of for example the rain gauge density and hence the spatial rainfall patterns may increase as well. The rainfall-runoff simulations were carried out with two models, the semi-distributed HBV model and the fully-distributed WaSiM-model. The spatial resolution is in WaSiM with 150 m x 150 m for each raster cell much higher than in HBV with approx. 20 km^2 per subcatchment. This higher spatial rainfall diversity and hence a numerical diffusion of the rainfall due to a too-coarse spatial resolution is thus avoided. Through the rainfall correction for altitude, an additional increase of the spatial diversity is achieved. While for the simulated discharge time series with HBV almost no differences between the different rainfall ~~data sets~~ products could be identified, for catchment Pionierbrücke in WaSiM slight differences between method V1 and methods V2 and V3 differences regarding the seasonal extreme values can be identified. For both, V2 and V3, subsequent steps after the rainfall disaggregation were applied to implement spatial consistence by simultaneous rainfall occurrence at different rain gauges. This affects the simulated runoff at least for instantaneous peak flows in the summer and winter period. However, the number of subcatchments in HBV and by that the spatial resolution of the rainfall-runoff model can be increased, which is assumed to lead to more diverse results between V1, V2 and V3, similar as resulting from WaSiM.

For Pionierbrücke, as a fast-reacting, mountainous catchment, the absolute differences for the seasonal extreme flows resulting from V1 or the ~~data sets~~ products V2 and V3 for a flood with a return period of 50 years are approx. 5-8 m^3/s during both the calibration and validation periods (see Fig. ~~16~~ 14 and ~~17~~15) using WaSiM. For the other two catchments Reckershausen and Tetendorf, the difference is expected to be smaller since both catchments are larger and cover a less steep area. Thus, no additional simulations with WaSiM have been carried out for these two catchments. In this context it should be mentioned, that WaSiM is a much more complex rainfall-runoff model than HBV with a high demand on meteorological input time series (e.g. precipitation, temperature, humidity, wind speed and global radiation), which has to be available for the whole simulation period on an hourly time step.

6. Summary, Conclusions & Outlook

The aim of this investigation is to investigate the influence of different degrees of spatial consistence in disaggregated time series on simulated runoff statistics. The investigation is carried out for three meso-scale catchments in Lower Saxony, Germany, which differ in terms of their size, land use, soil and slope. For the disaggregation, a multiplicative, micro-canonical cascade model after Müller and Haberlandt (2015) is used. Since the disaggregation process is performed on a per station basis without taking into account neighboring stations, spatial consistence must be implemented afterwards. Here, a resampling algorithm based on Müller and Haberlandt (2015) is applied (named V2) as well as a more pragmatic approach where the same relative diurnal cycle is used for all stations on the one day (Haberlandt and Radtke, 2014, named V3). Nevertheless, investigations without subsequent steps to implement spatial consistence exist as well (Ding et al., 2016) and have been included in this investigation (named V1). The hypothesis tested in this study is that these different rainfall ~~data sets~~ products lead to differences in the derived runoff statistics as well. The following conclusions can be drawn regarding the rainfall ~~data set~~ product differences:

1. The resampling algorithm for the implementation of spatial consistence was applied on an hourly basis for the first time for distances smaller than 20 km for V2. The achieved values for the bivariate spatial rainfall characteristics are comparable to those from observations.

2. The bivariate spatial characteristics are underestimated by V1 and overestimated by V3 respectively.

3. While for the areal rainfall intensities, the exceedance curve leads to an expected order of V1<V2<V3, for the areal rainfall extremes, V2 and V3 result in similar values, both being higher than V1.

The generated rainfall ~~data sets~~ products V1, V2 and V3 have been used as input for rainfall-runoff modeling to evaluate the influence of the above identified differences of rainfall characteristics. An application-based evaluation is important in terms of rainfall generation, since it provides a new perspective and hence new insights into the rainfall data (Müller and Haberlandt, 2016, Müller et al., 201~~7~~8, Sikorska et al., 201~~7~~8). For the simulations, the semi-distributed HBV model (Wallner et al., 2013) and the fully-distributed WaSiM model (Schulla, 1997, 2015) have been implemented. The essential findings are:

1. With the applied calibration process in HBV, a good representation of observed runoff statistics is possible for V1-V3 for the calibration period.

2. The ~~data sets~~ rainfall products V1-V3 result in only small differences in the simulated runoff statistics using HBV. Differences do not increase whether a neutral parameter set without calibration is applied nor if the station density increases.

3. For peak flows in the summer and winter periods, slight differences resulting from V1 and both, V2 and V3, can be identified using WaSiM. V2 and V3 lead to comparable higher flood peaks than V1, which is consistent with extreme value analysis of areal rainfall for this catchment.

4. For the intra-annual cycle and the flow duration curve, no difference resulting from V1-V3 can be identified from either HBV nor WaSiM.

Ding et al. (2016) achieved a good representation of summer and winter peak flows with V1 as input rainfall data and HBV as rainfall-runoff model. Haberlandt and Radtke (2014) applied HEC-HMS as semi-distributed rainfall-runoff model with disaggregated and parallelized rainfall time series (V3) as input data. The continuously simulated runoff time series were analyzed regarding annual extreme flows, which could be reproduced well for all catchments. The findings of both investigations can be confirmed by the actual study.

However, no differences resulting from V1, V2 and V3 regarding the summer and winter extremes are detectable for HBV. ~~This is remarkable, since V1 and V2 and their influence on simulated discharge have been analyzed before for 5-minute time steps by Müller and Haberlandt (2016) in an urban hydrological context. In their investigation, the urban hydrological model SWMM (Rossmann, 2010) has been used, which enables a good representation of the hydrodynamic of the fast-responding sewage system. Significant differences could be identified between the simulated runoff statistics resulting from V1 and V2 for their artificial sewage system.~~

On the other hand, WaSiM results in slight differences for seasonal extreme values for the investigated catchment Pionierbrücke which is in line with previous findings regarding the areal rainfall extreme values. ~~This is presumably caused by the spatial resolution, which is in WaSiM with 150 m x 150 m for each raster cell, being much higher than in HBV, with approx. 20 km² per subcatchment. The higher spatial resolution enables higher spatial rainfall diversity, which is intensified by the rainfall corrected for altitude. A numerical diffusion of the rainfall in space due to a too-coarse spatial resolution is thereby avoided. Semi-distributed rainfall-runoff models like HBV or HEC-HMS with a simple horizontal structure and hence a less complex model-structure (as in e.g. WaSiM) lead to numerical diffusion and hence to a "smudging" of the areal rainfall, resulting in less differences in runoff statistics. Other investigations raise the question if spatial rainfall patterns can be transferred sufficiently into runoff with semi-distributed models and thus with a coarse spatial resolution (Krajeski et al., 1991, Obled et al., 1994, Bárdossy and Das, 2008).~~

However, the differences between the resulting seasonal peak flows simulated with WaSiM from V1, V2 and V3 are still small with approx. 5-8 m³/s (up to 15 %) for floods with return periods of 50 years. It should be noted that V1, V2 and V3 clearly differ regarding the investigated spatial bivariate characteristics of probability of occurrence, coefficient of correlation, continuity ratio and the resulting areal rainfall intensities, especially regarding their extreme values. Hence, the hypothesis formulated before is rejected in this case study. Although several possible causes regarding the applied rainfall-runoff models (parameter calibration, rainfall station density, type and spatial resolution of rainfall-runoff model) have been analyzed, no final conclusion about the reason for the similar runoff statistic can be drawn. It is assumed that the damping behavior of the catchments leads to these small differences in runoff statistics.

These findings suggest that (i) simple model structures might compensate for deficiencies in spatial representativeness through parameterization and (ii) highly resolved hydrological models benefit from improved spatial modeling of rainfall.

Of course, the similarity of the simulated runoff statistics from V1, V2 and V3 is only valid for the investigated catchments. For catchments with other climatic or physiographic attributes, results can be different. Therefore, a systematic investigation of catchments with different hydrological behavior in climates and with different rainfall-runoff models would be necessary (comparative hydrology) to identify catchments, for which the degree of spatial rainfall consistence matters. The actual study could be a starting point for that.

However, the main intention of the actual study was to analyze the impact of rainfall ~~data sets~~ products with different degrees of spatial consistence on simulated runoff statistics. The application of the resampling algorithm (V2) is recommended for spatial application of disaggregated rainfall data since this method leads to the best agreement with the observed spatial rainfall characteristics.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

First of all, the two reviewers Anna Sikorska and Nadav Peleg are gratefully acknowledged. Their suggestions and comments helped to improve the manuscript significantly. The authors thank also the former student Jennifer Ullrich for calibration of the simulated annealing parameters. Thanks also to Ross Pidoto for useful comments on an early draft of the manuscript. A special thank to Bastian Heinrich for technical support during the investigation. We are also thankful for the permission to use the data of the German National Weather Service. The publication of this article was funded by the Open Access fund of Leibniz Universität Hannover.

References

Aarts, E. and Korst, J.: Simulated Annealing and Boltzmann Machines: A stochastic approach to combinatorial optimization and neural computing, John Wiley & Sons, Chichester, 1965.

Andres-Domenech, I., Garcia-Bartual, R., Montanari, A. and Marco, J.B.: Climate and hydrological variability: the catchment filtering role. Hydrol. Earth Syst. Sci. 19, 379–387, 2015.

Bárdossy, A. and Das, T.: Influence of rainfall observation network on model calibration and application. Hydrol. Earth Syst. Sci. 12, 77–89, 2008.

Beven, K. J.: Rainfall-Runoff Modelling: The Primer. John Wiley and Sons, Chichester, 2001.

Breinl, K., Turkington, T., and Stowasser, M.: Simulating daily precipitation and temperature: A weather generation framework for assessing hydrometeorological hazards. Meteorol. Appl., 22 (3), 334-347, 2014.

Bormann, H. and Elfert, S.: Application of WaSiM-ETH model to Northern German lowland catchments: model performance in relation to catchment characteristics and sensitivity to land use change. Adv. Geosci. 27, 1–10. doi:10.5194/adgeo-27-1-2010, 2010.

Buishand, T. A.: Stochastic modeling of daily rainfall sequences, H. Veenmann and Zonen B. V., Wageningen, Netherlands, 1977.

Carsteanu, A. and Foufoula-Georgiou, E.: Assessing dependence among weights in a multiplicative cascade model of temporal rainfall. J. Geophys. Res., 101(D21), 26363–26370. Federal Environment Agency (2009). CORINE Land Cover. DLR-DFD, Hannover, 1996.

Criss, R. E. and Winston, W. E.: Do Nash values have value? Discussion and alternate proposals. Hydrol. Process., 22, 2723-2725, 2008.

Ding, J., Wallner, M., Müller, H. and Haberlandt, U.: Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model. Hydrol. Process. 30 (9), 1431–1448, 2016.

DVWK: Ermittlung der Verdunstung von Land- und Wasserflächen. In: ATVDVWK-Regelwerk (Ed.), DVWK-Merkblatt 238/1996. Deutscher Verband für Wasserwirtschaft und Kulturbau e.V. (DVWK), Bonn, 1996.

DVWK: Verdunstung in Bezug zu Landnutzung, Bewuchs und Boden. Merkblatt ATV-DVWK-M 504. Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall, 2002.

Feldman, A. D.: Hydrological Modeling System HEC-HMS – Technical Reference Manual, US Army Corps of Engineers, 2000.

Förster, K., Hanzer, F., Winter, B., Marke, T. and Strasser, U.: An open-source MEteoroLOgical observation time series DISaggregation Tool (MELODIST v0.1.1). Geosci. Model Dev. 9 (7), 2315-2333, 2016.

Gelleszun, M., Kreye, P. and Meon, G.: Lexikografische Kalibrierungsstrategie für eine effiziente Parameterschätzung in hochaufgelösten Niederschlag-Abfluss-Modellen (Lexicographic calibration strategy for efficient parameter estimation in highly resolved rainfall-runoff models). Hydrol. Wasserbewirtsch. 59, 84–95. doi:10.5675/HyWa_2015,3_1, 2015.

Gires, A., Giangola-Murzyn, A., Abbes, J.-B., Tchiguirinskaia, I., Schertzer, D. und Lovejoy, S: „Impacts of small scale rainfall variability in urban areas: a case study with 1D and 1D/2D hydrological models in a multifractal framework". *Urban Water Journal* 12 (8), 607–617, 2014.

Haberlandt, U., Ebner von Eschenbach, A.D. and Buchwald, I.: A space-time hybrid hourly rainfall model for derived flood frequency analysis. Hydrol. Earth Syst. Sci. 12 (6), 1353–1367, 2008.

Haberlandt, U. and Radtke, I.: Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. Hydrol. Earth Syst. Sci. 18 (1), 353–365, 2014.

Hartwich, R., Behrens, J., Eckelmann, W., Haase, G., Richter, A., Roeschmann, G. and Schmidt, R.: Bodenübersichtskarte der Bundesrepublik Deutschland. Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, 1998.

Hosking, J. and Wallis, J.: Regional Frequency Analysis: an approach based on L-moments. Cambridge University Press, New York, 1997.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P.: Optimization by simulated annealing, Science, 220 (4598), 671-680, 1983.

Kirpich, Z. P.: Time of concentration of small agricultural watersheds. Civil Eng. 10 (6), 362, 1940.

Krajewski, W. F., Lakshmi, V., Georgakakos, K. P. and Jain, S. C.: A Monte Carlo study of rainfall sampling effect on a distributed catchment model. Water Resour. Res. 27 (1), 119–128, 1991.

Maniak, U.: Hydrologie und Wasserwirtschaft. Springer Verlag, Berlin, 2005.

Mandapaka, P. V., Krajewski, W. F., Mantilla, R. and Gupta, V. K.: Dissecting the effect of rainfall variability on the statistical structure of peak flows. Adv. Water Resour. 32, 1508–1525, 2009.

Mandelbrot, B. B.: Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier.J. Fluid Mech. 62 (2), 331–358, 1974.

Marshak, A., Davis, A., Cahalan, R., and Wiscombe,W.: Bounded cascade models as nonstationary multifractals. Phys. Rev. E, 49(1), 55–69, 1994.

Melsen, L. A., Teuling, A. J., Torfs, P. J. J. F., Uijlenhoet, R., Mizukami, N. und Clark, M. P.: Hydrol. Earth Syst. Sci. Opinions: The need for process-based evaluation of large-domain hyper-resolution models. Hydrol. Earth Syst. Sci. 20, 1069–1079, 2015.

Monteith, J.L.: Evaporation and environment, in: Fogg, G.E. (Ed.), The State and Movement of Water in Living Organisms, Symposia of the Society for Experimental Biology. Cambridge University Press, Cambridge, 205–234, 1965.

Müller, H.: Rainfall disaggregation for hydrological modeling, PhD-thesis, Faculty of Civil Engineering and Geosciences, Institute of Water Resources Management, Hydrology and Agricultural hydraulic Engineering (in German), 2016.

Müller, H. and Haberlandt, U.: Temporal Rainfall Disaggregation with a Cascade Model: From Single-Station Disaggregation to Spatial Rainfall. J. Hydrol. Eng. 20 (11), 04015026, 2015.

Müller, H. and Haberlandt, U.: Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology. J. Hydrol. 556, 847-864 accepted, doi: http://dx.doi.org/10.1016/j.hydrol.2016.01.031556, , 20186.

Müller, T., Schütze, M. and Bárdossy, A.: Temporal asymmetry in precipitation time series and its influence on flow simulations in combined sewer systems. Adv. Water Resour. 107, 56-64, 2017.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - a discussion of principles. J. Hydrol. 10, 282–290, 1970.

Nicotina, L. E., Celegon, E. Alessi, Rinaldo, A. and Marani, M.: On the impact of rainfall patterns on the hydrologic response. Water Resour. Res. 44, W12401, 2008.

Niehoff, D., Fritsch, U. and Bronstert, A.: Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany. J. Hydrol. 267, 80–93, 2002.

Obled, C., Wendling, J. and Beven, K.: The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data. J. Hydrol. 159 (1 4), 305–333, 1994.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., Assel, J. v., Kroll, S., Murla-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P. und Veldhuis, M.-C. ten (2015). „Impact of Spatial and Temporal Resolution of Rainfall Inputs on Urban Hydrodynamic Modelling Outputs: A Multi-Catchment Investigation“. *Journal of Hydrology* 531 (2), 389–407, 2015.

Ogden, F. L. and Julien, P. Y.: Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales. Water Resour. Res. 29 (8), 2589–2597, 1993.

Onof, C., Chandler, R. E. und Kakou, A.: Rainfall modelling using Poisson-cluster processes: a review of developments. Stoch. Env. Res. Risk A. 6 (14), 384–411, 2000.

Paschalis, A., Fatichi, S., Molnar, P., Rimkus, S., Burlando, P.: On the effects of small scale space–time variability of rainfall on basin flood response. J. Hydrol. 514, 313-327, 2014.

Pathiraja, S., Westra, S. and Sharma, A.: Why continuous simulation? The role of antecedent moisture in design flood estimation. Water Resour. Res. 48 (6), W06534, 2012.

Peel, M. C., Finlayson, B. L. and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11 (5), 1633–1644, 2007.

Peleg, N., Blumensaat, F., Molnar, P., Fatichi, S., Burlando, P.:Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modelling. Hydrol. Earth Syst. Sci. 21, 1559-1572, 2017.

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., Burlando, P.: Spatial variability of extreme rainfall at radar subpixel scale. J. Hydrol. 556, 922-933, 2018.

Rodriguez-Iturbe, I., Cox, D. R. and Isham, V.: Some models for rainfall based on stochastic point-processes. P. R. Soc. London A410, 269-288, 1987.

Rossman, L.A.: Storm water management model – User's manual version 5.0. Environmental Protection Agency, EPA/600/R-05/040, 2010.

Rupp, D. E., Keim, R. F., Ossiander, M., Brugnach, M. and Selker, J. S.: Time scale and intensity dependency in multiplicative cascades for temporal rainfall disaggregation. Water Resour. Res. 45 (7), W07409, 2009.

Schaefli, B., Gupta, H. V.: Do Nash Values have value?. Hydrol. Process., 21, 2075-2080, 2007.

Seliga, T. A., Aron, G., Aydin, K. and White, E.: Storm runoff simulation using radar-estimated rainfall rates and a Unit Hydrograph model (SYN-HYD) applied to GREVE watershed. American Meteorological Society, 25th International Conference on Radar Hydrology, 587–590, 1992.

Schulla, J.: Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen (Dissertation). ETH Zürich, 1997.

Schulla, J.: Model Description WaSiM (Water balance Simulation Model) - completely revised version of 2012 with 2013 to 2015 extensions. Hydrology Software Consulting J. Schulla, 2015.

Veneziano, D., Furcolo, P., and Iacobellis, V. (2006). Imperfect scaling of time and space-time rainfall. J. Hydrol. 322(1–4), 105–119.

Sikorska, A. E., Vivrioli, D., Seibert, J. (2017). Effective precipitation duration for runoff peaks based on catchment modelling. J. Hydrol., ~~submitted~~556, 510-522..

Sivapalan, M., Blöschl, G.: Transformation of point to areal rainfall: Intensityduration-frequency curves. J. Hydrol. 204 (1–4), 150–167, 1998.

Thober, S., Mai, J., Zink, M. and Samaniego, L.: Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets. Water Resour. Res. 50 (11), S. 8714–8735, 2014.

Veneziano, D., Langousis, A.: The areal reduction factor: a multifractal analysis. Water Resour. Res. 41 (7), W07008, 2005.

Wallner, M., Haberlandt, U. and Dietrich, J.: A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. J. Hydrol. 494 (0), S. 59–71, 2013.

Wallner, M. and Haberlandt, U.: Klimabedingte Änderung von Hochwasserabflüssen im Aller-Leine-Einzugsgebiet - Eine Fallstudie mit HBV-IWW. Hydrol. Wasserbewirtsch. 59 (4), 174–183, 2015.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E. and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves. Hydrol. Earth Syst. Sci. 15 (7), 2205–2227, 2011.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol. 210 (1-4), 178–191, 1998.

Wright, D.B., Smith, J.A., Baeck, M.L.: A critical examination of area reduction factors. J. Hydrol. Eng. 19 (4), 769–776, 2013.

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

Yu, P. S. and Yang, T. C.: Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites. Hydrol. Process. 14 (1), 117–133, 2000.