Comments 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.