

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<sup>2</sup>), 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.

[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).

[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).

[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...”

[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.

[2 17] “higher” – finer.

[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”.

[3 8] “investigations” replace with “studies”.

[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.

[3 20] “by amongst others” – change to “by others, as”.

[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.

[3 25] “to take into account different genesis” – not a clear sentence.

[3 27] “investigation area” – I believe “study area” is a more common phrase to use.

[3 27] replace “chapter” with “section”.

[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.

[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.

[6 12] “has been shown not to be that sensitive as model input” – a reference is needed here.

[7 6] “as per” – replace with “as in”.

[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.

[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.

[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.

[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.

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

[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 subsection 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”.

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

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

[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.

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

[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)?

[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.

[Fig. 4 and Table 5] Can be moved to the Supplementary Information (SI).

[11 23] to [11 26] A repetitive.

[12 1] Please define the periods for summer and winter.

[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?

[Eq. 5] Remove the “max”.

[Eq. 6] Can be moved to the SI.

[Eq. 7] Remove “min”. What is the logic behind the weights?

[Table 5] Move to 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.

[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?

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

[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).

[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.

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

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

[17 14] “runoff”.

[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?

[Fig. 7] What is the size of sub-catchment 2? From Figure 1 I would estimate around 20 km<sup>2</sup>. If this is the case, I would argue that V2 and V3 are likely overestimating the extreme rainfall intensities. For example, ~38 mm h<sup>-1</sup> for a 10-year return period over a ~20 km<sup>2</sup> 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.

[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.

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

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

[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.

[27 14] A repetitive.

[Table 12 and 13] SI.

[31 8] to [31 15] There are some repetitive here is well.

[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.

[32 10] But FDC compare daily discharges, right? I guess that at hourly scale the differences are clearer.

[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.

[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.