Dear Paweł Licznar,

thank you for your comments and suggestions, that will improve our manuscript. Please find below the answers or our comments on the questions you have raised.

- Page 7, lines 197-200. Why do you mention about 4-hour dry duration period between following maxima if you are deriving annual maximum series (AMS)? Don't you derive just only one maxima for each single calendar year? You have most probably months gaps between following maxima values in your series?

  Months gaps are mainly found on past recordings, but on the most recent period we have winter recordings. Depending on the duration level that we are considering, the maxima of two consecutive years should be distinguished by at least a 4-hour dry duration (or the duration interval itself) to make sure that we are not counting the same maxima in two consecutive years. We will add this explanation in the updated manuscript.

- Page 6, lines 185-186. Why only 30 realizations of disaggregation were made? In most multiplicative random cascades applications, cascades generators are run 100 times

  It was evaluated how the validation results change with increasing number of realisations. It was found that the relative error does not improve significantly after 30 realisations, as was also reported by Müller & Haberlandt (2018). Therefore, 30 disaggregation runs were made. We will add this to the manuscript.

- Page 6, lines 181-183. Daily rainfall is disaggregated up to 15 min time scale. If this is true, why in Fig. 3 you present the relative error results only to 1 h time scale? The evolution of error seen in Fig 3 shows visible underprediction of extreme rainfall for shorter time durations. Since this is cascading process most probably the magnitude of underprediction for sub-hourly durations of 30-min and 15-min must be much bigger. Most probably such problem could be explained by errors of cascade generator parametrization (for a detailed discussion of this please refer to:
  Licznar P., Å omotowski J., Rupp D. E.: Random cascade driven rainfall disaggregation for urban hydrology: An evaluation of six models and a new generator. Atmospheric Research 99, 2011, s. 563-578).

  During the many-step improvement of the parameterisation, it was found that the results for the sub-hourly durations are too uncertain for being further used. Therefore, in the end, the disaggregation was carried out only down to 1h duration. We will correct and clarify this in the manuscript.

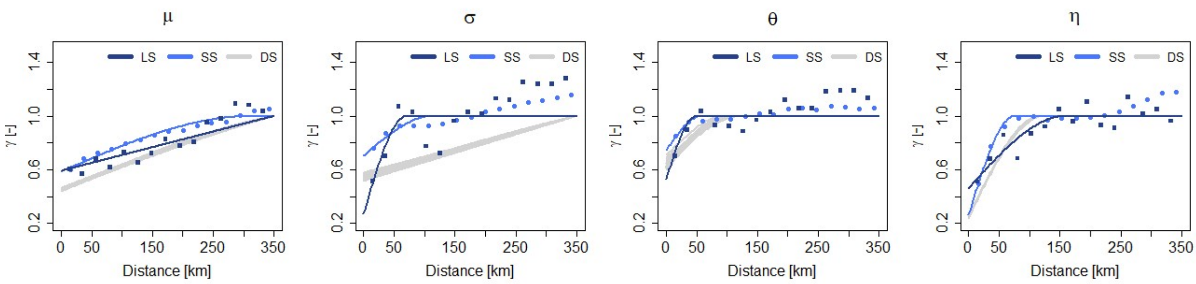
  Page 12, lines 342-343 "Based on a k clustering approach (Ward, 1963) 9 homogeneous regions were identified and are shown in Figure 8". This is unclear since usually you must enter the k-means algorithm with already known (assumed) number of clusters. How did you select k=9? Was there any objective approach to estimate the optimal clusters number? Addition of some dendrogram might help here. I believe most probably you should explain your selection of clusters number by analyzing how values of the CaliÅ ski and Harabasz Index (the CHIndex) and total within sum of squares (wss), depend on the adopted number of clusters k. Examples of such approach could be found at:
  Karol MikoÅ ajewski, Marek Ruman, Klaudia Kosek, Marcin Glixelli, Paulina DzimiÅ ska, Piotr ZiÄ tara, PaweÅ Licznar: Development of cluster analysis methodology for identification of model

rainfall hyetographs and its application at an urban precipitation field scale. Science of the Total Environment. 2022, vol. 829, art. 154588, s. 1-20

Our approach to estimate the number of clusters was somehow subjective. We tested different number of clusters between 2 and 20 and compared the homogeneity indicator H1 for the identified regions. The maximum number of clusters of 20 was chosen to ensure a sufficient number of stations and thus a sufficient number of observation years per region. We assumed 9 clusters as homogeneity was valid for all of the 9 identified clusters and the clusters were spatially continuous and physically reasonable. We will clarify this in the manuscript.

- Page 313, lines 306-307. In your manuscript you are only declaring selection and application of some semivariogram model. It is simple spherical model. Did you measured the goodness of spherical model fit? It is of value to have at least a single example illustrating empirical semivariogram and fitted theoretical spherical model. I am raising this question since here you might encounter two problems. First, high (singular) extreme values like ones reported for event in Münster destroy the regular picture of empirical variance dependance vs. stations distance. Second, usually to model the complicated picture of spatial distribution of extremes (or their distributions parameters) one needs to implement more advanced semivariogram model, being a superposition of two or three simple models (eg. double spherical plus nugget). It is explained by the influence of different scale process on local rainfall maxima (from large scale forcing up to local turbulence scale). For example, the rainfall maxima field in Poland is usually the sum of the outcomes of three types of processes operating at various spatial scales. The spatial scales are probably connected with a convective/orographic, a frontal and a 'climatological' genesis of high precipitation. It is explained in details in monograph (https://repozytorium.amu.edu.pl/bitstream/10593/3938/1/Stach_analiza.pdf). Once more it is in Polish but with English summary.

We have compared different fits for the empirical variograms, i.e. exponential, spherical and gaussian, but the spherical model with a nugget delivered the best fit to the data. The fitting of the variogram model parameters is done automatically by weighted least square fit. Since the automatic fit relies on the initial values of the model parameters, we defined the initial values with trial and error, and accepted a fit that was adequate qualitatively. In the updated manuscript, we will include illustration of the variograms used (see Figure below), estimated from the three datasets available: long series (LS), short series (SS) and daily disaggregated series (DS). Note that the variograms are normalized in order to ensure a comparison between the different datasets. From the Figure below the differences of the spatial dependency. We will discuss the differences more in the updated manuscript.



Regarding the Münster event, yes, we have noticed that it affects the structure of the variogram and most particularly the nugget of the variogram. We are considering this in another manuscript where we are investigating different types of uncertainty in the proposed regionalisation procedure. However, we will try to address this better in the updated version of this manuscript.

- What I am missing in all your manuscript is the proposal or clear discussion on how your method could contribute to confidence intervals estimation to be provided for designers at the new KOSTRA. To my best knowledge at the current KOSTRA Atlas they are imposed subjectively and calculated as a certain percentage of maxima. This approach is in general against the physics of rainfall maxima in nature. Rainfall series are multifractals and rainfall maxima are singularities. The more rare (higher frequency) maxima the higher order singularity it is and it becomes more uncertain. With increasing frequency and decreasing time duration confidence intervals should expand.

  Yes, you are right. The current KOSTRA product evaluates confidence intervals as dependable on only the return period. For return periods lower than 5 years, the confidence intervals is ± 10%, up to 50 years ±15% and up to 100 years ± 20%. As mentioned above, we are currently working in another manuscript were we extensively investigate the uncertainty sources and derive the confidence intervals. In out proposed methodology, the confidence intervals vary with strongly on durations (with confidence intervals of 5min being much higher than longer durations), return period, and location (for instance distance from a long recording location). We will discuss this shortly at the end of this manuscript as an outlook that we are currently working on. Note that the kriging simulations have the advantage of tackling the spatial uncertainty through different simulations, which provides a useful tool in assessing the overall uncertainty of regionalisation.

- It is worth to explain or at least discuss why you have selected new 5 km by 5 km spatial resolution for the new KOSTRA. This resolution is different than in old KOSTRA (71 km2). Here I would advise to refer to the paper: Hengl T., 2006, Finding the right pixel size, Computers and Geosciences, 32 (9), 1283-1298, DOI: 10.1016/j.cageo.2005.11.008.

  We were considering two types of resolution 1 km by 1km, and 5 km by 5km. We choose this resolution in an agreement with German Weather Service DWD, which has other hydrometeorological data at same spatial resolution (HYRAS products). Nevertheless, the study you have suggested is interesting, and we will include this in the discussion about the choice of the spatial resolution.

- Page 8, lines 221-222. I would say that another possible reason could be a “step response error” typical for electronic gauges. It is real problem not only in case of simple tipping-bucket gauges series but also for series recorded by the first generation of weighing type gauges. This issue is discussed at: Licznar, P., De Michele, C., and Adamowski, W.: Precipitation variability within an urban monitoring network via microcanonical cascade generators, Hydrol. Earth Syst. Sci., 19, 485-506, doi:10.5194/hess-19-485-2015, 2015.

  We had given a possible reason why analogue gauges might not register the total amount of precipitation of very high intensity falling in short time intervals. The step response error of the electronic gauges is a counterargument that we will include in our reasoning.

References:

Müller, H. and Haberlandt, U.: Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology, J. Hydrol., 556, 847–864, doi:10.1016/J.JHYDROL.2016.01.031, 2018