

## Reply on RC2 concerning the specific remarks

Jieru Yan et al.

---

Author comment on "Conditional simulation of spatial rainfall fields using random mixing: a study that implements full control over the stochastic process" by Jieru Yan et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-56-AC3>, 2021

---

As Prof. Uijlenhoet did not number the specific remarks and listed them in the order that they appear in the manuscript, for consistency, we use line/figure/table number to refer to the corresponding remark hereafter.

Considering the relatively large amount of the specific remarks, we have classified them into two categories: (a) the deficiencies raised where the corresponding replies are not that necessary, but the relevant improvements should be made, such as the inappropriate reference (Remark L. 45), lack of reference (Remark L. 192), ambiguous statement (Remark L. 295-296), suspicion of using undefined abbreviations (Remark Fig. 9, Table 1); (b) the kind of deficiencies/ambiguities that need to be specified or discussed. We have summarized the second kind of remarks into the following 6 topics:

1) The role of random mixing (Remark L. 135-204) and the role of the marginal distribution function in the approach to obtain spatial rainfall fields, respectively, as well as the issue raised concerning the intermittency in the computation of the marginal distribution function (Remark L. 91-97).

Random mixing (RM) is an excellent tool that performs conditional simulation in Gaussian space, but it is not irreplaceable. Another simulation method could be used instead. RM is employed in this work due to (a) the relatively high efficiency (see Comment 4 in RC1 for detailed information on the computational time), which makes mass production of realizations possible, and (b) code availability. A Python package for conditional simulation of spatial random fields using RM is available (Hörning & Haese, 2021), where the authors give practical demonstrations on the usage of the method.

The marginal distribution function, on the other hand, is the foundation of the approach. The statistics of intermittent precipitation are non-Gaussian, and such properties restrict the usage of well-established stochastic models that assume Gaussianity (including RM) (Pulkkinen et al., 2019). Specifically, the marginal distribution function serves two functions: (a) data transformation: the simulated Gaussian fields are transformed to precipitation fields utilizing the normal-quantile transformation (Bogner et al., 2012), where the marginal distribution function is required; (b) definition of the constraints in Gaussian space.

The referee has raised an issue concerning the intermittency in the computation of the marginal distribution function (Remark L. 91-97). We fully agree that radar and rain

gauges have different space-time sampling properties, and one should not expect the same probability of finding zeros by both sensors. Yet one should choose from the two sources of information to evaluate the intermittency. We have used radar data, as the intermittency computed from a limited number of rain gauge observations is less reliable. The issue raised by the referee is a redundancy error. The starting point (0, the ratio of the dry area /intermittency), where the marginal distribution function starts to be positive, has already been set. Practically, we have not used the sampled zeros (at the gauge locations) by both sensors in the computation of the marginal distribution function. The enforcement (L. 91-97 ) is not necessary in the first place. Nevertheless, the redundant step does not influence the estimated marginal distribution function.

2) The interchangeable use of "correlation function" and "variogram" (Remarks L. 118-124 & L. 212-213).

Thank you for the reminder. Otherwise, we might have ignored the lack of clarity before the interchangeable use. We have made simple statistics: the word "variogram" has appeared five times and is relatively concentrated (L. 120-123 and the caption in Fig. 2), while the "correlation function" showed up 14 times. Hence it might be easier if we stick with "variogram". Yet on second thought, we found the complete cutoff of "variogram" is not easy. The simulation is developed under the assumption that the correlation function of the process is stationary, and the correlation between the process at any two locations is only a function of the vector connecting the two locations (i.e., second-order stationarity). Rather than estimating the covariance function, it is common in geostatistics to work with the variogram, i.e., the variance of the difference as a function of the vector. It has been shown that the estimation of the variogram is more stable than the estimation of the correlation function directly (Calder & Cressie, 2009). Namely, one applies the simulation using the correlation function as the measure of spatial dependence, yet the spatial dependence of the simulated product is normally examined on its variogram. Thus, we consider preserving the part where the "variogram" present, but providing a clearer description to explain why "variogram" instead of "correlation function".

The referee has expressed concern about the distinction between the empirical variogram (evaluated from the truncated Gaussian field) and the true variogram. The relevant discussion is missing in the manuscript. It is not a problem for the approach. RM is a geostatistical simulation method, and the fundamental element is the spatially correlated random field. Similarly, as the case in Kriging, the choice of variogram has a limited effect on the results (Verworn & Haberlandt, 2011). Further, the variogram computed from the truncated Gaussian field (transformed radar data) gives an excellent hint to approach the true variogram.

3) The target of the proposed approach (Remark L. 217-218).

The proposed approach is aimed at estimating spatial rainfall fields of short accumulation time: 15 min, 10 min, or even 5 min (if the good quality of rain gauge data can be maintained at such fine time resolution). In the last case, there is no aggregation of radar data; thus, there should be no dispute about employing the log-normal distribution as the model of the CDF of the rainfall field. Further, we think that slight aggregation of radar data (say 2 or 3 timesteps) should not change the type of distribution function remarkably. Under the above assumption, we think it is not inappropriate to use the log-normal distribution function as the model of the CDF of the rainfall field. Admittedly, the statement (L. 217-218 in the manuscript) is ambiguous, and a specification should be provided.

4) Concern raised due to the employment of synthetic data (Remarks L. 232-233 & L. 238-240 & L. 265).

This study is a pure simulation study, where one has full control over the stochastic process due to the employment of the synthetic data (as the true rainfall field). Undoubtedly, synthetic data can only partially represent reality, which is one limitation of the approach. Specifically, this study only considered the random error in radar estimates and did not consider systematic effects (e.g., range dependent error) in radar estimates (Remark L. 232-233). The basic assumption of the proposed approach is that the field pattern indicated by radar is similar to the pattern of the precipitation field on the ground. If the systematic effects are prevailing in the radar estimates such that the assumption is not valid (Remark L. 238-240), then the proposed approach is no longer applicable. The above limitation should be discussed in the manuscript, see Topic (6).

Similarly, with a pure simulation study, one has the opportunity to decide the layout of the rain gauge network as well as the corresponding density, which brings both pros and cons. The advantage is that one has full control over the stochastic process, which makes the sensitivity study possible. While the drawback is that one can hardly comprehensively model the rain gauge network encountered in practice, which is usually irregularly distributed with various densities. Now that we have full control, and also because the real-world gauge network is too complicated to model, we made things as simple as possible: square domain and uniformly distributed rain gauges (concerning Remark L. 265). Yet admittedly, a brief discussion is necessary where relevant.

As the temporal aspect of QPE has been discussed in the reply concerning the general remarks, we just skip Remark Fig. 5.

5) The distinction between the proposed method and the Kriging methods involved in the study (Remarks L. 252-253 & L. 306-307) and how the results from the proposed method can be utilized (Remark L. 314).

First, the choice of the two parameters (Remark L. 252-253). The choice of the two parameters is indeed arbitrary because it makes no difference to the results for the proposed approach. In the algorithm, the radar estimates are converted to quantiles, and the corresponding quantile map (or field pattern in the nomenclature of the manuscript) is incorporated. The monotonic transformation (the power function) does not change the quantile map; thus the results of the approach are not affected. Yet the choice of the two parameters does matter for the Kriging methods involved in the study, e.g., an underestimation of the radar estimates leads to an underestimation in the kriging results. As it is not essential for the proposed approach, we do not care too much about the choice of the two parameters. Further, as suggested by Berne & Krajewski (2013) and Curry (2012), radar data are prone to underestimate the precipitation, and underestimated precipitation will not be very useful for many hydrological applications; thus, we have modeled a case when radar underestimates the rainfall.

Considering the distinction of the results from the proposed approach and the Kriging methods (Remark L. 306-307), with the Kriging methods, one obtains an estimate (one rainfall field) that tends to underestimate the peak and overestimate the small, i.e., more middle-ranged rainfall values are present in the estimate. In the algorithm of Kriging, the marginal distribution function is not required. The Kriging results are obtained via geostatistical interpolation where the stationarity of the regional variable is assumed. However, if one evaluates the empirical distribution function from the Kriging-estimates, the departure of the empirical distribution from the true distribution function can be observed.

By comparison, with the proposed approach, one can obtain a bunch of realizations (estimates). The individual estimate gives relatively accurate statistics (mean, variance, covariance). Yet practically, one can hardly obtain a single realization with accurate statistics, and simultaneously with the precise positions of the rainfall peaks. Namely, one

gets the peaks accurately located but erroneous statistics in the mean realization, while accurate statistics but inaccurate positions of the peaks in the individual realizations (concerning Remark L. 314). Nevertheless, the realizations are still useful. In this fully controlled setup, we know exactly how the true rainfall field looks like. Yet when the true field is unknown, one possibility is that we simulate a bunch of realizations that are statistically identical to the true field and feed each realization to applications, e.g., hydrological models, nowcast models. The corresponding result is a range of possible outcomes, which can represent the estimation uncertainty.

6) The lack of a discussion section (L. 405).

It is necessary to have a discussion section, where the assumptions and the associated limitations, the potential suitability of the approach, and the possible improvement in the future can be discussed. We sincerely thank the referee for the valuable suggestion.

#### Literature

Berne, A. and Krajewski, W.F.: Radar for hydrology: unfulfilled promise or unrecognized potential? *Adv. Water Resour.* 51, 357–366, 2013.

Bogner, K., F Pappenberger, and Cloke, H. L.: Technical note: the normal quantile transformation and its application in a flood forecasting system. *Hydrology and Earth System Sciences*,16,4(2012-04-02), 16, 1085-1094, 2012.

Calder C.A. and Cressie N.: Kriging and Variogram Models, in: *International Encyclopedia of Human Geography*, edited by: Rob, K. and Nigel, T., 49-55 pp., Elsevier, 2009. <https://www.sciencedirect.com/science/article/pii/B9780080449104004612>

Curry, G.R.: Radar essentials: A concise handbook for radar design and performance analysis. *Radar Essentials: A Concise Handbook for Radar Design and Performance Analysis*. SciTech Publishing, 2012. DOI:10.1049/SBRA029E.

Hörning, S. and Haese, B.: RMWSPy (v 1.1): A Python code for spatial simulation and inversion for environmental applications. *Environmental Modelling and Software*, 138(12), 104970, 2021.

Pulkkinen, S., Nerini, D., Hortal, A., Velasco-Forero, C., and Foresti, L.: Pysteps: an open-source python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*, 12(10), 4185–4219, 2019.

Verworn, A. and Haberlandt, U.: Spatial interpolation of hourly rainfall-effect of additional information, variogram inference and storm properties. *Hydrol. Earth Syst. Sci.* 15 (2), 569–584, 2011.