We thank Scott Sinclair for the comments for improvement. In the following, we present our replies on the general comments and the detailed and editorial comments.

Concerning the general comments

Surely there are many interesting questions to be discussed concerning the applicability and capability of the proposed method, for example, on what conditions does the proposed method allow a better estimation of the rainfall field and the uncertainty of the estimation than the established Kriging methods. The other referee - Prof. Uijlenhoet - has indicated that a proper discussion section is missing in the manuscript, where one could discuss the assumptions and the associated limitations of the proposed method. We think here the referee implicitly raises the same problem.

It would indeed be of value to understand whether this technique has better performance over a range of time/space scales than other methods. And it is correct that the proposed method fixes both the marginal distribution function and the spatial correlation for each field, while for Kriging only the spatial correlation is fixed. However, this is a pure simulation study, where one has full control over the stochastic process (i.e., knowing the true rainfall fields). Under this context, even if we present better performance of the proposed technique over a range of time/space scales than other methods, the persuasiveness is limited, as synthetic data can only partially represent the reality. Yet as described in AC2, the goodness of a pure simulation study is that one has full control, thus can verify the accuracy of the estimates more comprehensively, compared to the case when the true datasets (radar and gauges) are used, where one only verifies the accuracy of the estimates at limited locations using, e.g., leave-n-out cross-validation. The capability of the proposed method is demonstrated in a synthetic experiment, and many realistic questions are to be answered; hence a further study based on realistic datasets is required.

Concerning the detailed and editorial comments

The referee has listed 11 such comments, among which, the 1st, 4th, 7th are editorial comments which require no replies, yet the relevant improvements should be made. In the following we present our replies to the other comments:

> **The 2nd comment**: pg 4, line 95 - is this quality control step justified by any reason other than practical considerations of the method?
>
> The quality control steps are applied to rule out the negative effects of the random error introduced in the radar estimates, which can lead to inconsistency at zeros, as well as the Spearmans rank correlation of the two datasets (gauge observations and the collocated radar quantiles) not being exactly 1. The quality control steps are effective in eliminating the interference from the random error, while for systematic errors, e.g., the range-dependent errors, the effectiveness of the quality control steps is limited.
>
> **The 3rd , 6th and 8th comments** – concerning the algorithm to compute the marginal distribution of the rainfall field
>
> <u>The 3rd comment</u>: We fully agree that at some point the domain must be too large for a single CDF to represent all processes, namely, the spatial CDF can be considered valid at a limited spatial scale, and this scale should be related to factors such as the rainfall regime, local climate, topography, etc. The referee has raised a very interesting question that we have not considered previously. In the experimental context of this study, we could not answer this question properly, and a further study based on realistic datasets is therefore required.
>
> <u>The 6th comment</u>: the choice of the intermittency u_0. Practically, the choice of u_0 should not change the results significantly in the experimental context of this study. We have tested a hydrologically interesting case, $u_0 = 0.36$. When $u_0 > 0.36$ (larger dry-area-ratio), it only means the starting point of the CDF (the intersecting point on y-axis in Fig. 3b) moves up, and there are more zero-samples in both datasets (gauge and radar). When $0 < u_0 < 0.36$ (larger wet-area-ratio), the starting point of the CDF moves down, and there are fewer zero-samples in both datasets. The algorithm will be problematic if $u_0 = 0$, i.e., the entire domain is wet. In that case, there is no $(0, u_0)$, and thus there

should be no enforcement that the CDF intersects at the point $(0, u_0)$ when fitting a theoretical CDF (Line 104).

The 8th comment: the referee has raised a similar question here as the other referee, see RC1 the 3rd comment, "In practice, for mesoscale hydrological studies only small sample sizes of irregular distributed recording rainfall stations are available...". To obtain an accurate CDF using the proposed method, a certain number of rain gauge observations should be given, which are not necessarily uniformly distributed, yet the rain gauge observations should not be too clustered. And even with a small sample size, there are possibilities to improve the applicability of the method, see the reply to the 3rd comment in AC1.

The 5th comment: Lines 163, 164 - how to decide to increase N?

Take the case when the number of rain gauge observations is 25 as an example. We have 25 linear constraints, $K = 25$ in Eqn. (7). One should choose an initial value of $N (N > 25)$ - the number of unconditional random fields - say 50. Then solve the under-determined system defined in Eqn. (7), and find the set of weights with the least norm, $\sqrt{\sum_{i=1}^{N} \alpha_i^2}$, using, e.g., the singular value decomposition. If the norm is above a certain threshold (say 0.1), then increase N by a step of 10, for example, and solve Eqn. (7) again. Repeat the above procedure until the norm ≤ 0.1.

The last 3 comments – concerning the error statistics in estimation of the extremes.

What is shown in Fig. 9, Line 367, and Table 1 is the error statistics in estimating the single extreme for each field. Yet it should be clarified that when converting the simulated Gaussian field Z to rainfall field R using the normal-quantile transformation, Eqn. (15) in the manuscript

$$R = G^{-1}(\Phi(Z))$$

all pixels satisfying $Z(x) > \Phi^{-1}(0.995)$ are converted to $G^{-1}(0.995)$ to get rid of the numerical effects, where $G(\cdot)$ and $\Phi(\cdot)$ are the CDFs of the rainfall field and the standard normal distribution, respectively. From this perspective, one could say the results shown are the extremes above a threshold quantile, though an extremely high threshold. If a lower threshold is used (say 0.95, 0.9, 0.8), the bias in the results from both RM and KED should decrease. And in that case, the contrast of the results from the two methods should not be as significant as shown in Fig. 9, Line 367, and Table 1. In a nutshell, the higher the threshold quantile, the more significant the contrast is.

The last comment: as for whether to include the "extremes of errors" in Table 1. We think the histogram of the errors provides a very intuitive expression of how the errors distribute, and we have shown the most extreme cases in Fig. 9: the scenarios that are most favorable and unfavorable for the methods. Table 1 is derived from the histograms and provides supplementary information on the means and scatterings of the histograms. As for the extreme errors (the referee meant the maximum absolute error if we understand correctly), it is our opinion that this statistic seems not so informative, and the two methods seem to have comparable performances, as observed from Fig. 9.