Overview:

The paper 'Conditional simulation of surface rainfall fields using modified phase annealing' by Jieru Yan and co-authors presents a new method to derive quantitative precipitation estimates by combining in-situ rain gauge observations and weather radar imagery. The ultimate goal is therefore to obtain accurate and exhaustive estimates of surface rainfall along with associated uncertainty. The present manuscript builds on the paper 'Short time precipitation estimation using weather radar and surface observations: With rainfall displacement information integrated in a stochastic manner' published last year in Journal of Hydrology by Yan and Bárdossy, and the main contribution of the present work is to propose a new framework to merge radar and rain gauge observations. I think the topic fits well with the scope of HESS. The paper is generally well-written and well-structured, and the proposed method is properly described except for one point detailed below. Except from this methodological detail, my main concern is the lack of practical guidelines to help practitioners using the proposed framework.

Main concerns:

(1) How is intermittency handled?

The method of surface rainfall simulation is in general described in details, and the dedicated section (Sect 2) is easy to read and to understand. However, within this nice description, there is one gap: how is rain intermittency (i.e. zero measurements) handled?

Regarding what authors call 'Distribution Function of Surface Rainfall' (Sect 2.1.1) it is clear that zeros are processed separately and that the proportion of zeros is assessed from the radar image. At this point I still follow what happens, and I just have a small question: why are zeros estimated using the whole radar image while the distribution of positive rain intensities is assessed using only pixels co-located with rain gauges? What justifies this difference of treatment?

Starting from Sect 2.1.2, I lost track of how the authors deal with zeros. In particular: In Eq (4) and (5) which value is assigned to $z_0 = \Phi^{-1}(u_0)$ with u_0 the quantile corresponding to zero precipitation? At that point if one assign a single value to z_0 the resulting 'reference field' does not have a Gaussian marginal distribution because the value z0 will be over-represented (and in addition there is no value below z0). Hence, in Sections 2.1.2 and 2.1.3, I suppose that the authors use some tricks to deal with intermittency, but I was not able to find which ones. I think different ways can be followed to tackle this problem of intermittency, and it is fair if the authors made an arbitrary choice. But I would like to know what is exactly done.

The way how ϕ is inverted for zeros has many implications for the following steps, in particular:

- The objective function is evaluated on Z, and therefore its value depends on how z_0 is defined. If the intermittency is high, the value of the objective function (and in turn the field towards which the phase annealing converges) can greatly vary.
- In case a single value z_0 is assigned to all locations where no-rain is observed in the radar image, I have the impression that the marginal distribution of the initial reference field should be adjusted accordingly (the 'starter' is not Gaussian anymore, but rather 'truncated Gaussian').
- This also raises the question of how is defined the covariance of the 'starter'? I suppose it is inferred from rain gauge observations transformed into Z-values, but here again you have to decide what to do with zeros. Depending of the method adopted the inferred covariance can be biased.

(2) Lack of practical guidelines.

The manuscript is quite technical, which in a way is Ok since the main goal is to present a new methodology to combine radar and rain gauge data. However, the application stated in the introduction (i.e. computing QPEs) is of great interest for practitioners, and I am afraid that this readership would not find enough information to apply the proposed framework to its own data. Hereafter are some questions the authors could address in order to make their manuscript more 'user friendly':

- How is the covariance of the 'starter' inferred from observations?
- Which gauge density is required to ensure good results?
- What is the maximum extent of the domain of interest that can be considered?
- Why selecting a final resolution of 500m in space and 30 min in time? In particular why not adopting the temporal resolution of the radar (5 min)? I suppose that a bit of temporal aggregation is required to smooth out some the mismatches between radar and rain gauge observations, but in this case why 30 min?
- What are the requirements in terms of rain gauge and radar data quality?

In addition, I think that more examples should be shown in order to convince potential users that the proposed method is the one they need. For this purpose a single 30min example is certainly not enough. One option would be to process a larger dataset (e.g. one month, or ideally one year), and provide the results as supplementary material in a dedicated repository. And in the main paper you could highlight some specific cases for which you have the feeling that your method performs well (and maybe some other cases where it does not… just to mention the limitations).

Last point about the usability of the method. It would be great to provide a utility (any kind of code, even if it is not a user friendly software) to perform radar-rain gauge combination following the method proposed in the paper. This will definitely help potential users to build on your own developments. This may exist in the repository provided in 'Data availability' (you mention some python scripts) but the DOI you provide seems erroneous or at least I failed to access it.

<u>Minor comments:</u>

* In the introduction I am missing some references to support the following statements:

- 'the intricate temporal-spatial variability, together with the intermittent nature of precipitation in both space and time, has hampered the accuracy of QPE.'
- 'However, the problem with radar-based QPE is the non-guaranteed accuracy, which could be impaired by various sources of errors, such as static/dynamic clutters, signal attenuation, anomalous propagation of the radar beam, uncertainty in Z-R relationship, etc.'
- 'the QPE obtained by merging the point-wise rain-gauge observations and the radar-indicated precipitation pattern has become a research hotspot in both meteorology and hydrology.'
- 'the well-known Wiener-Chintchin theorem'.
- 'Yet admittedly, if the perturbation is implemented at lower frequencies, the impact is more global and vice versa'.
- 'However, PA is found to be insufficient in the treatment of local constraints.'

* Paper outlook (l 59-63): Please use numbers to number the sections (e.g. Section two -> Section 2).

* Second global constraint (l 118-130): I have the impression that any global statistic could be used instead of the directional asymmetry. And one could even combine several statistics in the objective function (Eq 8). If it is the case you should mention it since it makes your method much more generic and flexible.

* Section 5: The results using the different options are pretty similar, as acknowledged for instance l355-362. In particular, if adding the component asymmetry in the objective function does not change the result, why spending so much time describing it? You should try to find (and show) at least one case for which this component improves the final result. If you do not, either replace directional asymmetry by a statistic that influences the result, or remove the mentions to the hybrid objective function and keep only the pattern-related term (which is already a nice contribution).

* In all figures where the standard deviation of realizations is displayed, it appears that only the pixels co-located with rain gauges have a low uncertainty. And the uncertainty increases very fast when moving away from the gauges. Hence conditioning to rain gauges seems to have a very local impact. Any comment about it?

Summary and recommendation:

To sum-up, I think the present paper is an interesting and novel contribution in the field of QPE generation from radar and rain gauge observations. I therefore recommend publication after moderate revisions.

Sincerely,

Lionel Benoit