https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions

Conditional simulation of surface rainfall fields using modified phase annealing

Jieru Yan[1], András Bárdossy[2], Sebastian Hörning[3], and Tao Tao[1]

[1]College of Environmental Science and Engineering, Tongji University, Shanghai, China.
[2]Institute of Modeling Hydraulic and Environmental Systems, Department of Hydrology and Geohydrology, University of Stuttgart, Stuttgart, Germany.
[3]Centre for Natural Gas, Faculty of Engineering, Architecture and Information Technology, the University of Queensland, Brisbane, Australia.

Correspondence: Jieru Yan (yanjieru1988@163.com)

Abstract. The accuracy of quantitative precipitation estimation (QPE) over a certain region and period is of vital importance across multiple domains and disciplines. However, due to the intricate tempo-spatial variability and the intermittent nature of precipitation, it is challenging to obtain QPE with adequate accuracy. This paper aims at simulating rainfall fields honoring both the local constraints subject by the point-wise rain-gauge observations and the global constraints subject by the field measurement from weather radar. The employed conditional simulation method is the modified phase annealing (PA), which is practically an evolvement of the traditional simulated annealing (SA). Yet, unlike SA which implements perturbations in the spatial field, PA implements perturbations in the Fourier space, making it superior to SA in many aspects. The modification of PA is reflected in two aspects. First, taking advantage of the global characteristic of PA, the method is only used to deal with global constraints, and the local ones are handed over to residual kriging. Second, except for the system temperature, the number of perturbed phases is also annealed during the simulation process, making the influence of the perturbation more global at initial phases. The impact of the perturbation decreases gradually as the number of the perturbed phases decreases. The proposed method is used to simulate the rainfall field for a 30-min-event using different scenarios: with and without integrating the uncertainty of the radar-indicated rainfall pattern and with different objective functions.

1 Introduction

Quantitative precipitation estimation (QPE) over a certain region and period is of vital importance across multiple domains and disciplines. Yet the intricate temporal-spatial variability, together with the intermittent nature of precipitation in both space and time, has hampered the accuracy of QPE.

The point-wise observations of precipitation measured by rain-gauges are accurate but only available at limited locations. Meanwhile, precipitation-related measurements produced by meteorological radars have become standard outputs of weather

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

offices in many places in the world. 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. Despite the various sources of errors, weather radar has been widely acknowledged as a valid indicator of precipitation patterns (e.g., Mendez Antonio et al., 2009; Fabry, 2015). Considering the pros and cons of the two most usually seen sources of precipitation information, 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.

Under the context of merging radar and rain-gauge data, we consider two types of constraints: the local constraints subject by the point-wise rain-gauge observations and the global constraints subject by the field measurement from weather radar. This paper focuses on simulating surface rainfall fields conditioned on the two types of constraints. There exist a variety of geostatistical methods aiming at simulating conditional Gaussian fields with prescribed covariance function, such as turning bands simulation, LU decomposition-based methods, sequential Gaussian simulation, etc. (see Deutsch et al., 1998; Chilès and Delfiner, 2012; Lantuéjoul, 2013, for details). The common goal of these methods is to ensure that the simulated realizations do comply with the additional information available (Lauzon and Marcotte, 2019). The additional information could be observed values of the simulated targets, measurements that are related linearly or non-linearly to the simulated targets, third- or higher-order statistics (Guthke and Bárdossy, 2017; Bárdossy and Hörning, 2017), etc.

The conditional simulation method used in this work is phase annealing (PA). It is first proposed in Hörning and Bárdossy (2018) and is essentially an evolvement of the general-purpose, meta-heuristics method, simulated annealing (SA) (Kirkpatrick et al., 1983; Geman and Geman, 1984; Deutsch, 1992; Deutsch et al., 1994). It utilizes the sophisticated optimization scheme of SA in the search for the global optimum. Yet, compared to SA, the distinction or evolvement of PA lies in that the perturbation, or swapping in the nomenclature of SA, is implemented in the Fourier space. Or, to be more exact, the perturbation is implemented on the phase component of the Fourier transform. While the power spectrum is preserved, such that the spatial covariance is invariant at all iterations according to the well-known Wiener-Chintchin theorem. Compared to SA, PA alleviates the singularity problem, namely the undesired discontinuities or poor-embedding of the conditional points within the neighborhood (Hörning and Bárdossy, 2018). And in general, PA has a much higher convergence rate compared to SA.

A remarkable feature of PA is that it is a global method: any perturbation imposed on the phase component is reflected on the entire field. Yet admittedly, if the perturbation is implemented at lower frequencies, the impact is more global and vice versa. The global characteristic of PA imparts it a perfect methodology for global constraints. However, PA is found to be insufficient in the treatment of local constraints. Note that by local constraints, we refer primarily to point equality constraints, when the total number of the constraints is far less than that of the grid points. In the algorithm of PA, the local constraints are normally ensured by inserting a component measuring the dissimilarity between the simulated and the target values. On the other hand, one could argue that if the information on the measurement error is explicitly known, then this piece of information could be considered in the simulation and the insufficiency of PA in handling the local constraints can be utilized in turn. However, in the general case, the local constraints can only be approximated by PA.

Respecting the fact that the specialty of PA is the treatment of the global constraints, we separate the global from the local. In particular, PA is only used to handle the global constraints, and local ones are handled separately by residual kriging at each

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

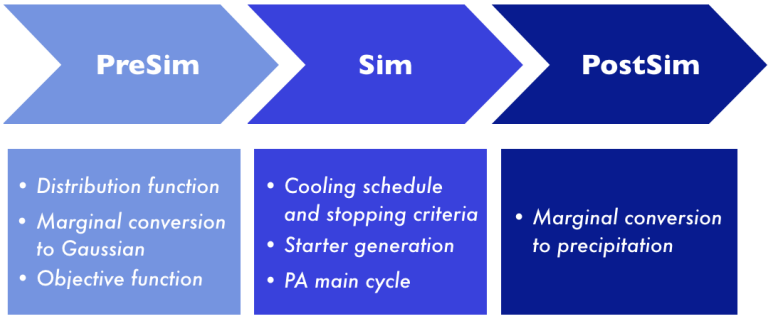


Figure 1. Flowchart of the procedure to simulate surface rainfall fields using the algorithm of PA.

iteration. And as an extension of PA, except for annealing the system temperature, the number of perturbed phases is annealed in parallel to render the algorithm work more globally at initial phases of the simulation. The global impact of the perturbation is weakened as the number of perturbed phases decreases.

This paper is divided into six sections. After the general introduction, the methodology of PA is introduced in Section *two*, including three stages: pre-simulation, simulation, and post-simulation. Section *three* provides two options to integrate the uncertainty of the radar-indicated rainfall pattern into the simulation. Section *four* introduces the study domain and the two types of data used in this study. In Section *five*, the proposed algorithm is used to simulate the rainfall field for a 30-min-event, where different simulation scenarios are applied. Section *six* ends this paper with conclusions.

2 Methodology

Figure 1 summarizes the procedure of simulating surface rainfall fields using the algorithm of PA, including three stages: pre-simulation (PreSim), simulation (Sim) and post-simulation (PostSim). Each stage and the corresponding sub-stages are described in the following subsections in the same sequence as shown in the flowchart.

2.1 Pre-simulation

2.1.1 Distribution Function of Surface Rainfall

PA is embedded in Gaussian space, or, more specifically, standard normal space. The univariate distribution function is essential to transform the simulated Gaussian fields into rainfall fields.

The scenario is as follows: based on K rain-gauge observations and a regular grid of radar quantiles (representing the pattern of surface rainfall), the distribution function of surface rainfall is generated. The procedure is first introduced in Yan and Bárdossy (2019) and we specify the modified version here.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

(a) $\forall$ rain-gauge observations r_k, the collocated quantiles in the radar quantile map U is determined and denoted as u_k. The two datasets are then sorted in ascending order, i.e., $r_1 \leq \cdots \leq r_K$ and $u_1 \leq \cdots \leq u_K$.

(b) The quantile corresponding to zero precipitation (denote as u_0), is determined. u_0 is actually the ratio of the dry area in the domain of interest. It is recommended to estimate u_0 from the original radar display in dBZ, where a typical value around 20 dBZ could be used as the threshold.

(c) Let $G(r)$ be the distribution function of surface rainfall and a linear interpolation is applied for rainfall values of less than the maximum recorded gauge observations, i.e., r $\leq r_K$:

$$G(r) = \frac{u_k - u_{k-1}}{r_k - r_{k-1}}(r - r_{k-1}) + u_{k-1} \tag{1}$$

with r_{k-1}, r_k being the two nearest neighbors of r ($r_{k-1} \leq r \leq r_k$), and u_{k-1}, u_k being the quantiles corresponding to r_{k-1}, r_k, respectively.

(d) Extrapolate for rainfall values $r > r_K$. A modification is made here: the minimum of the exponential and linear extrapolation is used as the result, as expressed in Equation 2. As we've learned from practice that the exponential extrapolation tends to over-estimate the rainfall extremes. Thus, a linear component is used to restrict the extrapolation result.

$$G(r) = \min\left(1 - e^{-\lambda r},\ \frac{u_K - u_{K-1}}{r_K - r_{K-1}}(r - r_K) + u_K\right) \tag{2}$$

where, the only parameter λ of the exponential distribution is determined from the last pair (r_K, u_K):

$$\lambda = -ln(1 - u_K)/r_K \tag{3}$$

Note that the above-mentioned radar quantile map U could be obtained from the original radar display in dBZ or the radar-based precipitation estimates. For the latter case, as we only use the radar-indicated spatial ranks, there is no requirement on the accuracy of Z-R relation, given the monotonic relationship of the two quantities.

It is worth mentioning that if the two datasets, r_k and u_k, have a Spearman's rank correlation of less than 1, the re-ordering in Step (a) will destroy the original collocating configuration. In fact, the different ordering of the two datasets reflects the conflict of radar and gauge data, and the misfit is utilized later in the quantification of the uncertainty of the radar-indicated rainfall pattern in the subsequent section.

2.1.2 Marginal Conversion to Gaussian

As PA is embedded in Gaussian space, all the constraints, including the point equality targets r_k $[\mathrm{mm}]$ subject by rain-gauge observations and the quantile map U $[-]$ subject by weather radar, are converted to the standard normal marginal using the quantile-quantile transformation, as formulated in Equation 4 and 5.

$$Z^\star = \Phi^{-1}(U) \tag{4}$$

$$z_k = \Phi^{-1}(G(r_k)) \quad \text{for } k = 1, \cdots, K. \tag{5}$$

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

where Φ^{-1} is the inverse of the standard normal distribution function; G is the distribution function of surface rainfall obtained according to the procedure described in Section 2.1.1.

2.1.3 Objective Function

We impose two kinds of constraints: the local and global constraints. As has been explained in the introductory section, PA is a powerful method to handle global constraints. In order not to mess up the logic behind PA and to fulfill the point equality constraints (local constraints) exactly, residual kriging is implemented at each iteration to fix up the simulated values at the observational locations. Thus, the objective function only needs to measure the fulfillment of the global constraints.

We impose two global constraints: field pattern and directional asymmetry. Note that both constraints are evaluated from the fields of the standard normal marginal and are compared with the marginal-converted Gaussian field $Z^\star$, which we term ***the reference field*** hereafter.

The *first global constraint*, field pattern, requires that the simulated field should be similar to the reference field. The similarity of the two fields is quantified by the Pearson's product-moment correlation coefficient:

$$\rho_{Z,Z^\star} = \frac{\mathrm{cov}(Z, Z^\star)}{\sigma_Z \sigma_Z^\star} \tag{6}$$

In the ideal case, $\rho_{Z,Z^\star}$ equals 1, and we use the difference, $(1 - \rho_{Z,Z^\star})$, to measure the distance from the ideal.

The *second global constraint* is the directional asymmetry, as expressed in Equation 7. Directional asymmetry is first introduced in Bárdossy and Hörning (2017) and Hörning and Bárdossy (2018). It is a third-order statistic and the physical phenomenon revealed by this statistic could be significant for advection-dominant processes, such as storms.

$$[A(h)]_Z = \frac{1}{N(h)} \sum_{x_i - x_j \approx h} \left(\Phi\left(Z(x_i)\right) - \Phi\left(Z(x_j)\right)\right)^3 \tag{7}$$

where, $[A(h)]_Z$ (abbreviated as A_Z, hereafter) is the asymmetry function evaluated from the simulated Gaussian field Z; $N(h)$ is the number of the pairs fulfilling $x_i - x_j \approx h$ and Φ is the cumulative standard normal distribution function. This statistic could be computed efficiently on regular grids using fast Fourier transformations (Marcotte, 1996; Lauzon and Marcotte, 2019). We use the entire field to compute the directional asymmetry function and compare the simulated one with the reference asymmetry function, i.e., the directional asymmetry function evaluated from the reference field, abbreviated as $A_{Z^\star}$. There exist multiple choices to define the distance of the two asymmetry functions, and we've used the L^∞ norm, $\|A_Z - A_{Z^\star}\|_{L^\infty}$.

Different schemes could be used to combine the two components, linear or non-linear, and we've chosen the maximum of the two components. Finally, the objective function we've used to quantify the fulfillment of the two global constraints is expressed as:

$$\mathcal{O}(Z) = \max\left((1 - \rho_{Z,Z^\star}),\ (w/A_{\text{scal}}) \cdot \|A_Z - A_{Z^\star}\|_{L^\infty}\right) \tag{8}$$

where, w is the relative weight of the component directional asymmetry, and A_{scal}, as expressed in Equation 9, is the scaling factor that scales the L^∞ norm of the difference of the two asymmetry functions between 0 and 1.

$$A_{\text{scal}} = \|A_{Z^\star}\|_{L^\infty} \tag{9}$$

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

2.2 Simulation

2.2.1 Cooling Schedule and Stopping Criteria

The system temperature of PA decreases according to the cooling schedule as the optimization process goes on. And the lower the system temperature, the less likely a bad perturbation is being accepted. A bad perturbation is defined when the perturbation does not decrease the objective function value. A reasonable cooling schedule is capable of preventing the optimization of being trapped prematurely at a local optimum. Yet one should be aware that it is always a compromise between the statistical guarantee of the convergence and the computational cost: the slower the temperature decreases, the higher probability of the convergence; however, cooling slowly also means more iterations and therefore, higher computational costs.

Comparative studies of the performance of SA using the most important cooling schedules, i.e., multiplicative monotonic, additive monotonic and non-monotonic adaptive cooling, have been made by, e.g., Nourani and Andresen (1998), Martín and Sierra (2009), etc. The results show that the annealing works properly when the cooling curve has a moderate slope at the initial and central stages of the process and tends to have a softer slope at the final stage. Lots of cooling schedules satisfy these conditions. As PA utilizes the optimization strategy of SA, the rules also apply for PA. Our choice of cooling schedule is the exponential multiplicative monotonic one.

There could be multiple choices of stopping criteria for an optimization algorithm, such as (a) the total number of iterations implemented; (b) the predefined objective function value; (c) the rate of decrease of the objective function; (d) the number of continuous temperature cycles without improvement, if a discrete cooling schedule is being used; (e) the predefined threshold of the initial objective function value, and so forth.

Our choice of the stopping criterion is a combination of (a) and (b). Specifically, we search for two parameters, the initial temperature (also the maximum temperature) T_0 and the final temperature (also the minimum temperature) $T_{\min}$ by decreasing the temperature exponentially and discretely. At each fixed temperature, N perturbations, say 1000, are implemented and the corresponding acceptance rate and improvement rate are computed. The *acceptance* means the perturbation is being accepted by the system and the system state is being updated, whether the perturbation brings improvement to the system; while the *improvement* means the perturbation does decrease the objective function value. In short, a perturbation bringing improvement must be an accepted perturbation, yet an accepted perturbation is not necessarily bringing improvement to the system. We refer to the experiment, where N perturbations are being done and the two mentioned rates are being computed, at a fixed temperature as a ***temperature cycle***.

T_0 is firstly set by starting a temperature cycle with an initial guess of T_0. If the acceptance rate is lower than the predefined limit, say 98%, then increase the temperature, and vice versa. The goal is to find a T_0 with a relatively high acceptance rate. Then, $T_{\min}$ is set by decreasing T_0 exponentially until the predefined stopping criterion is met. We use the criterion when the predefined objective function value has been reached. It is clear, in our case, that if more iterations are implemented, better realizations (realizations with a smaller objective function value) could be produced. Yet, it is again a compromise between the satisfying destination and the computational cost.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

Note that ever since the determination of T_0, the total number of temperature cycles m (to anneal T_0 to $T_{\min}$) is recorded. Thus the total number of perturbations could be estimated as $L = mN$. A continuous cooling schedule is then computed from T_0, $T_{\min}$ and L. The temperature at Iteration l is computed by:

$$T_l = T_0 \cdot \alpha_T^l \quad \text{for } l = 0, \cdots, L-1. \tag{10}$$

where, α_T is the attenuation factor of system temperature, which can be computed by

$$\alpha_T = \exp\left(\ln\left(\frac{T_{\min}}{T_0}\right)/(L-1)\right) \tag{11}$$

Inspired by the work by Lauzon and Marcotte (2019), in parallel with the temperature annealing, we also anneal the number of phases being perturbed in the Fourier space: starting with a relatively large number N_0 (say 5 to 20% of the total number of the valid Fourier phases) and decreasing all the way down to 1. The logic behind this is to make the perturbation have a more global influence initially when the distance from the destination is relatively large, and vice versa. The number of phases to be perturbed at Iteration l is computed by

$$N_l = N_0 \cdot \alpha_N^l \quad \text{for } l = 0, \cdots, L-1 \tag{12}$$

where, α_N is the attenuation factor with respect to the number of perturbed Fourier phases. It can be computed by

$$\alpha_N = \exp\left(-\ln(N_0)/(L-1)\right) \tag{13}$$

2.2.2 Starter Generation

PA requires a starting Gaussian random field with the prescribed spatial covariance, abbreviated as *starter*. Various methods could be used to generate such a field, e.g. fast Fourier transformation for regular grids (Wood and Chan, 1994; Wood, 1995; Ravalec et al., 2000), turning band simulation (Journel, 1974), or the Cholesky transformation of the covariance matrix.

If fast Fourier transformation (FFT) is utilized in the generation of the starter, the inherent periodic property of FFT should be treated with carefulness. Specifically, the simulation should be embedded in a larger domain. The original domain is enlarged in all directions by a finite range, i.e. the range bringing the covariance function from one to approaching-zero. As we use random fields of the standard normal marginal, the maximum covariance function value (equaling the variance) is one. If covariance models with asymptotic ranges (e.g. exponential, Matérn, Gaussian covariances, etc) are employed, the extension in domain size could be significant (Chilès and Delfiner, 2012).

2.2.3 PA Main Cycle

The starting point of PA is a Gaussian random field Z_K with the prescribed spatial covariance, as described in Section 2.2.2. Here a little modification, in our case, is that the values at the observational locations are fixed by residual kriging, as indicated by the subscript "K". The procedure of residual kriging is the same as explained in the following (Step 6). The procedure of the PA algorithm applied in this paper is specified as follows:

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

1. The discrete Fourier transform (DFT) of Z_K is computed by:

$$\mathcal{Z}_K = \mathcal{F}\{Z_K\} \tag{14}$$

2. The system temperature (T_l) and the number of perturbed phases (N_l) at Iteration l are computed by Equation 10 and 12.

3. N_l vectors (u_n, v_n) are generated, where $n = 0, \cdots, N_l - 1$. u_n, v_n are randomly drawn from the two discrete uniform distributions $[1, \cdots, U]$ and $[1, \cdots, V]$, respectively, where U, V are the highest frequencies considered for perturbation in the two spatial directions.

   Note that DC frequency should be excepted from the perturbation. One could use the entire frequency range or select a sub-area to impose the perturbation.

4. N_l phases (θ_n, $n = 0, \cdots, N_l - 1$) are randomly drawn from the uniform distribution, $[-\pi, \pi)$.

5. The Fourier coefficients at the selected locations, (u_n, v_n), are drawn and expressed as

$$\mathcal{Z}_K[u_n, v_n] = a_n + jb_n \tag{15}$$

   These coefficients are then updated in terms of the Fourier phases, using θ_n as expressed in Equation 16; while the amplitudes remain unchanged:

$$\sqrt{a_n^2 + b_n^2} \cdot (\cos(\theta_n) + j \sin(\theta_n)) \tag{16}$$

   The updating is also applied to the symmetrical locations of (u_n, v_n) in the Fourier space by the conjugation:

$$\sqrt{a_n^2 + b_n^2} \cdot (\cos(\theta_n) - j \sin(\theta_n)) \tag{17}$$

   The perturbed Fourier transform is denoted as $\tilde{\mathcal{Z}}$, and the corresponding perturbed spatial random field is obtained by the inverse DFT:

$$\tilde{Z} = \mathcal{F}^{-1}\{\tilde{\mathcal{Z}}\} \tag{18}$$

6. Due to the perturbation, $\tilde{Z}$ no longer satisfies the point equality constraints exactly (note the removal of the subscript "K"). Thus, kriging is applied for the residuals, $z_k - \tilde{Z}(x_k)$, where z_k are the point equality constraints in Gaussian space, as defined in Equation 5. The results of kriging, r, are superimposed on $\tilde{Z}$:

$$\tilde{Z}_K = \tilde{Z} + r \tag{19}$$

   As kriging is a geostatistical method that depends only on the configuration of the data points, the weight matrix of individual grid point does not change at iterations. One needs to compute the weight matrices for all the grid points for only once. Thus, residual kriging is cheap to use and causes almost no additional computational cost.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

7. $\tilde{Z}_K$ is then subject to the objective function defined in Equation 8. If $\mathcal{O}(\tilde{Z}_K) < \mathcal{O}(Z_K)$, the perturbation is accepted. Otherwise, the perturbation is accepted with the probability:

$$P = \exp\left(\frac{\mathcal{O}(Z_K) - \mathcal{O}(\tilde{Z}_K)}{T_l}\right) \tag{20}$$

8. If the perturbation is accepted, the system state is updated, namely Z_K, $\mathcal{Z}_K$ and $\mathcal{O}(Z_K)$ are replaced by $\tilde{Z}_K$, $\tilde{\mathcal{Z}}_K$ and $\mathcal{O}(\tilde{Z}_K)$, respectively.

9. If the stopping criterion is met, stop and Z_K is a realization in Gaussian space, satisfying the predefined optimization criterion. Otherwise, go to Step 2.

2.3 Post-Simulation

The realization Z_K in Gaussian space is transformed to the original marginal of precipitation using the quantile-quantile transformation:

$$R = G^{-1}(\Phi(Z_K)) \tag{21}$$

where Φ is the cumulative standard normal distribution function and G^{-1} is the inverse of the distribution function of surface rainfall, as obtained in Section 2.1.1. The resultant R is a realization of the surface rainfall field.

3 Uncertainty of Radar-indicated Rainfall Pattern

As has mentioned in Section 2.1.1, the misfit between the collocated radar quantiles u_k and rain-gauge observations r_k is reflected by the Spearman's rank correlation ρ^S: the closer ρ^S to 1, the more accordance (or the less conflict) of the two datasets.

The conflict of the two could be partially explained by the fact that weather radar is measuring at some distance above the ground (a few hundred to more than a thousand meters aloft). It is, therefore, reasonable to suspect the correctness of comparing the ground-based rain-gauge observations with the collocated radar data by assuming the vertical descent of the hydrometeors. In fact, hydrometeors are very likely to be laterally advected during their descending by the wind. And the wind is quite frequently occurring concurrently with precipitation. To take the possible wind-induced displacement into account, the procedure described in Yan and Bárdossy (2019) is adopted. Here we recall the procedure briefly in three steps.

(1) A rank correlation matrix $\boldsymbol{\rho}_{ij}^S$ is generated by (first) shifting the original radar quantile map U with all the vectors defined by a regular grid $\boldsymbol{h}_{ij}$, i.e., locations of all the grid cells in $\boldsymbol{h}_{ij}$ and (then) calculating the rank correlation between the rain-gauge observations and the collocated radar quantiles in the shifted map. Note that the radar quantile map shifted by the vector in $\boldsymbol{h}_{ij}$ is denoted as $\boldsymbol{U}_{ij}$.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020


Hydrology and Earth System Sciences
Discussions

Both $\boldsymbol{h}_{ij}$ and $\boldsymbol{\rho}_{ij}^{S}$ have the same resolution as the radar map. The center of $\boldsymbol{h}_{ij}$ is $(0,0)$, with the corresponding entry in $\boldsymbol{\rho}_{ij}^{S}$ denoted as ρ_{00}^{S}, meaning that zero shift is imposed on U. The further a grid cell in $\boldsymbol{h}_{ij}$ from the center, the larger shift is imposed on U. One should limit the number of grid cells in $\boldsymbol{h}_{ij}$, depending on the radar measurement height.

(2) A probability matrix $\boldsymbol{P}_{ij}$ is generated from the rank correlation matrix $\boldsymbol{\rho}_{ij}^{S}$:

$$\boldsymbol{P}_{ij} = \begin{cases} 0 & \text{for } \rho_{ij}^{S} \leq \rho_{00}^{S} \\ H(\rho_{ij}^{S}) & \text{else} \end{cases} \tag{22}$$

where, H is a monotonic function. It could be a power function, an exponential function or a logarithmic function, etc. The first derivative (dH/dx) matters, as a large dH/dx means that more weights are assigned to those displacement vectors (or shifted fields) producing a higher rank correlation. Our choice of function H is simply: $H(x) = x^2$.

(3) $\boldsymbol{P}_{ij}$ is then scaled, as expressed in Equation 23, to ensure the sum of all entries equals 1, such that P_{ij} has the same property as a probability mass function.

$$\boldsymbol{P}_{ij} = \frac{P_{ij}}{\sum_i \sum_j P_{ij}} \tag{23}$$

$\boldsymbol{P}_{ij}$ quantifies the uncertainty of the radar-indicated rainfall pattern by indicating the probability of individual shifted quantile fields. The logic behind it is that only those displacement vectors which increase the accordance of radar and gauge data are accepted and given positive probability.

There are two options to integrate the information carried by $\boldsymbol{P}_{ij}$ into the simulation:

The first option, the expectation of the shifted quantile fields $\boldsymbol{U}_{ij}$ is computed using the probability matrix. Then, the marginal-converted expected quantile field (as expressed in Equation 24) is used as the reference field when applying the PA algorithm.

$$Z^{\star} = \Phi^{-1}\left(\sum_i \sum_j \boldsymbol{P}_{ij} \cdot \boldsymbol{U}_{ij}\right) \tag{24}$$

One could also place the marginal conversion operator, $\Phi^{-1}(\cdot)$, inside the summation, as expressed in Equation 25. Yet, the distinction is tiny.

$$Z^{\star} = \sum_i \sum_j \left(\boldsymbol{P}_{ij} \cdot \Phi^{-1}(\boldsymbol{U}_{ij})\right) \tag{25}$$

The second option, those (marginal-converted) shifted quantile fields, $\Phi^{-1}(\boldsymbol{U}_{ij})$, bearing a positive probability are taken as the individual inputs (the reference fields) of PA and the algorithm described in the section *Methodology* is applied for all these inputs. The results are multiple realizations of the surface rainfall field, denoted as $\boldsymbol{R}_{ij}$. The expectation of these realizations is calculated using the probability matrix, as expressed in Equation 26. And $\bar{R}$ is the expected realization which takes the

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

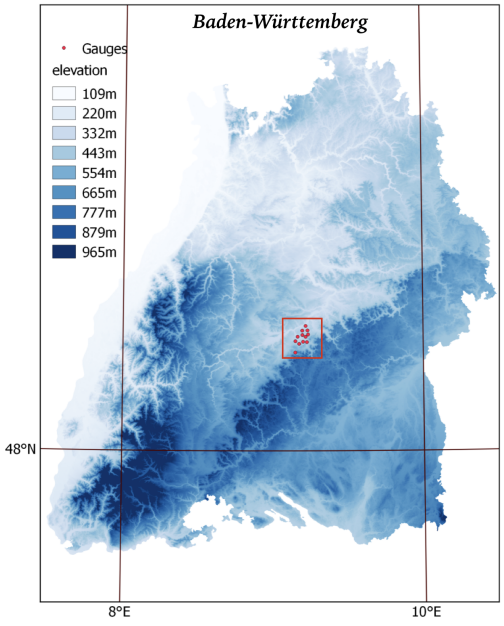


Figure 2. The elevation map of Baden-Würtemberg with the study domain marked by the red square and the rain-gauges marked by the red dots.

uncertainty of the radar-indicated rainfall pattern into account.

$$\bar{R} = \sum_i \sum_j \boldsymbol{P}_{ij} \cdot \boldsymbol{R}_{ij} \tag{26}$$

Both options are capable of producing realizations of the surface rainfall field with the uncertainty of the radar-indicated rainfall pattern integrated, yet the results are distinct, as presented in the section *Application*.

4 Study Domain and Data

The study domain is located in Baden-Würtemberg in the southwest of Germany, as shown in Figure 2. It is a square domain with the side length of $19\mathrm{km}$. The domain is discretized to a 39×39 grid with the spatial resolution of $500\mathrm{m} \times 500\mathrm{m}$. A gauge network consisting of 12 pluviometers is available within the domain, as denoted by the red dots in Figure 2.

The radar data used in this study is the raw data measured by Radar Türkheim, around $45\mathrm{km}$ from the domain center. The temporal resolution is $5\mathrm{min}$. Radar Türkheim is operated by the German Weather Service (DWD). The raw data are subject to a processing chain consisting of (1) clutter removal (Gabella and Notarpietro, 2002); (2) attenuation correlation (Krämer and Verworn, 2008; Jacobi and Heistermann, 2016); (3) re-projection from polar to Cartesian coordinates and (4) clip the square data for the study domain. All these quality control steps are operated under the environment of *wradlib*, an open-source library for weather radar data processing (Heistermann et al., 2013).

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

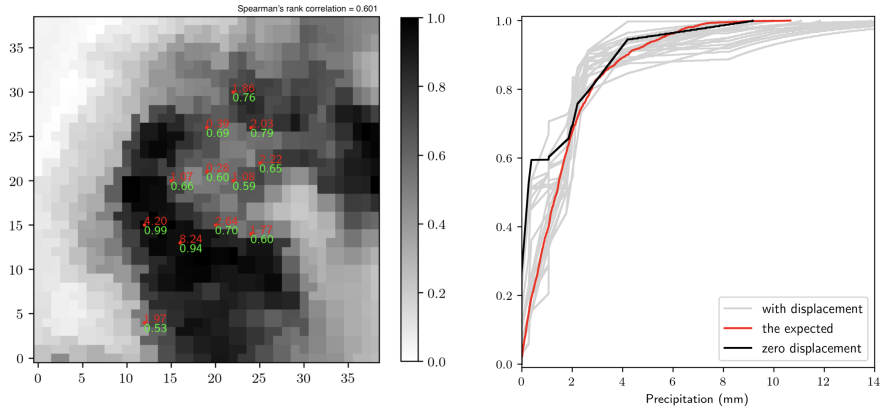


Figure 3. (Left) Original radar quantile map for the 30-min event *20170504 13:20 - 13:50*, with the rain-gauge observations [mm] labeled in red and the collocated radar quantiles [−] labeled in lime green. (Right) Distribution functions of surface rainfall: Grayish - radar maps with displacement; Red - the expectation of the grayish; Black - original radar map with zero displacement.

5 Application

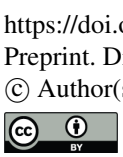
We select a 30-min event to apply the algorithm of PA. The event is selected, not only due to the relatively prominent rainfall intensity reflected by the rain-gauge data, but more importantly, the event is properly captured by a few rain-gauges unevenly distributed in the domain of interest, as shown by the red dots in the left figure in Figure 3. From the figure, it is clearly seen that the sampled quantiles (indicated by the text in lime green) cover the entire range $[0, 1]$: not only the smalls or the bigs but the sampled quantiles are more or less evenly distributed. One might have noticed that the smallest sampled quantile is 0.53. Yet, in this case, u_0 (the quantile corresponding to zero precipitation) is 0.26, and the quantile value of 0.53 corresponds to the rainfall value of 0.28mm after the re-ordering.

The conflict of radar and gauge data is obviously reflected in Figure 3, for example, the collocation of 4.22mm rainfall with the quantile 0.99 in contrast with the collocation of 8.24mm rainfall with the quantile 0.94. The Spearman's rank correlation of the gauge and the collocated radar data is 0.601, as labeled in the title of the figure. The distribution function of the surface rainfall field based on the two original datasets is shown by the black line in Figure 3 (right).

However, using the algorithm described in Section 3 to evaluate the uncertainty of the radar-indicated rainfall pattern, one could obtain multiple distribution functions of surface rainfall, as shown by the grayish lines in the right figure of Figure 3. Note that only the distribution functions associated with those shifted fields possessing a positive probability (as indicated by the probability matrix) are shown in the figure.

With these distribution functions, one could transform the corresponding shifted quantile fields into rainfall fields. And with the probability matrix, the expectation of these rainfall fields can be computed, as shown in Figure 4 (right). The expected quantile map is shown on the left, with the rain-gauge observations and the collocated radar quantiles labeled. A remarkable

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

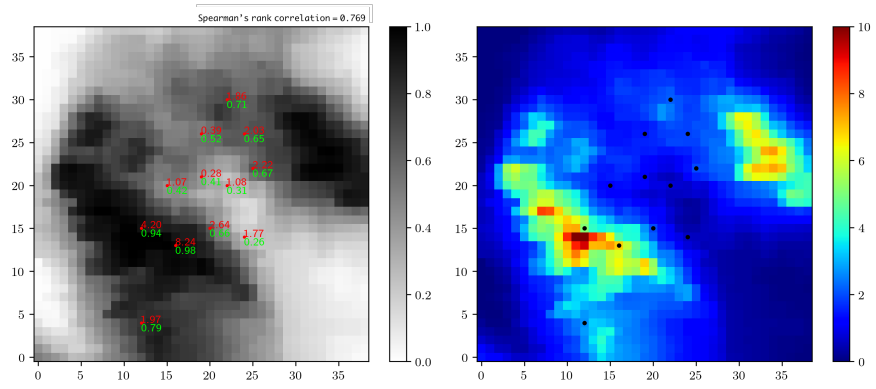


Figure 4. (Left) The expected quantile field $[-]$, with the rain-gauge observations labeled in red and the collocated radar quantiles labeled in lime green. (Right) The expected rainfall field $[\mathrm{mm}]$.

improvement of the concordance of the two datasets is reflected by the increased Spearman's rank correlation (0.769, as labeled in the title of the figure).

We open up two simulation sessions, depending on the different objective functions used when applying the PA algorithm. In the ***first session***, the objective function contains only the component field pattern: $\mathcal{O}(Z) = 1 - \rho_{Z,Z^\star}$. In the ***second session***, the component directional asymmetry comes into play and the objective function expressed in Equation 8 is employed, where the relative weight of the component directional asymmetry w equals 0.5. Technically, for both simulation sessions, the process stops when the objective function reaches below 0.05.

5.1 Simulation Session I

We present an evolvement in terms of the simulation strategy, where the algorithm of PA is applied differently by:

(1) using the original quantile map as the reference;

(2) using the expected quantile map as the reference, i.e., integrating the uncertainty of the radar-indicated rainfall pattern via Option 1 in Section 3;

(3) simulating independently using those shifted quantile maps with a positive probability as the reference and computing the expectation, i.e., integrating the uncertainty of the radar-indicated rainfall pattern via Option 2 in Section 3.

Stage 1, simulating using the original quantile map as the reference means the uncertainty of the radar-indicated rainfall pattern is not integrated. Figure 5 shows the mean of 100 such realizations on the left and the corresponding standard deviation map on the right. The standard deviation map reflects the variability between different realizations.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

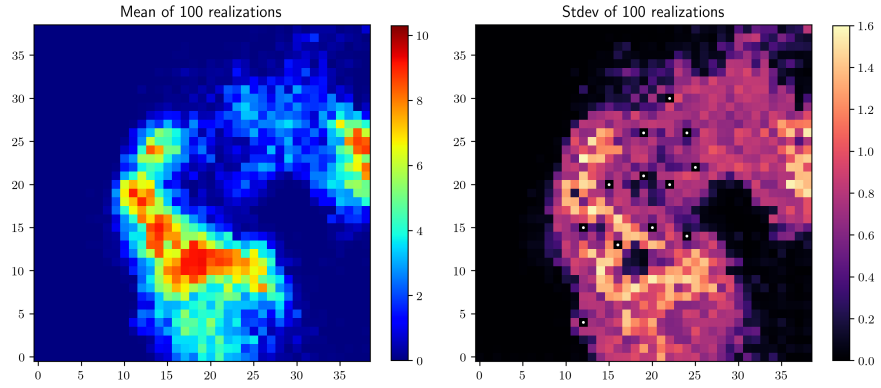


Figure 5. (Results Stage 1) Mean and standard deviation of 100 realizations obtained using the original radar quantile map as the reference. Unit: [mm]

Stage 2, simulating using the expected quantile map as the reference integrates the uncertainty of the radar-indicated rainfall pattern via the first option as described in Section 3. Similarly, the mean of 100 such realizations and the corresponding standard deviation map are shown in Figure 6. Comparing Figure 5 and 6, one could observe the obvious displacements of the peaks in both the rainfall map and the standard deviation map. Compared to Figure 5, a visible reduction in the standard deviation could be observed in Figure 6, showing the reduced estimation uncertainty by integrating the uncertainty of the radar-indicated rainfall pattern.

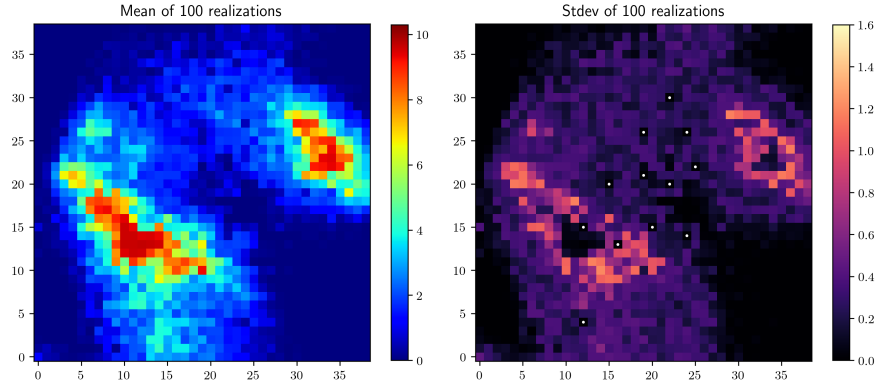


Figure 6. (Results Stage 2) Mean and standard deviation of 100 realizations obtained using the expected quantile map as the reference. Unit: [mm]

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions

Stage 3 involves a simulation strategy that is a little bit complicated than before. The simulation is applied independently using the shifted quantile map associated with a positive probability as the reference field. And the single simulation cycle is applied for all the components possessing a positive probability. Finally, the expectation of the individual realizations is computed. We term the simulation result from Stage 3 ***the expected realization***, hereafter. Using this simulation strategy, 100 expected realizations are generated, but at a much higher computational cost. The time elapsed to obtain a single realization is practically the same, yet to obtain an expected realization, one needs (in this case) to simulate 23 contributing realizations. Thus, to obtain 100 expected realizations, 2300 single realizations are required in total. In the similar mode as before, the mean of the 100 expected realizations and the corresponding standard deviation map are shown in Figure 7. Comparing Figure 6 and 7, the location of the rainfall peak is almost the same. Yet, compared to Figure 6, the reduction in the standard deviation is remarkable in Figure 7, showing the further reduction in the estimation uncertainty.

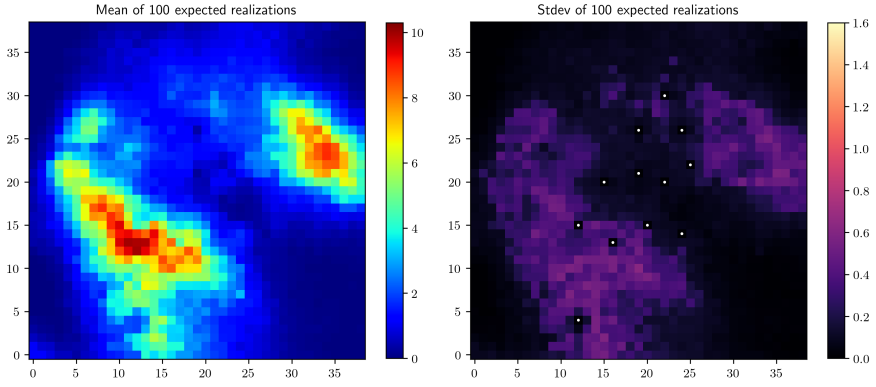


Figure 7. (Results Stage 3) Mean and standard deviation of 100 expected realizations. Unit: [mm]

Besides, the mean-field, as well as the standard deviation map of the results from Stage 3, is much smoother compared to the other alternatives shown previously. To show the similarity between different expected realizations, as well as the smoothness, Figure 8 displays four randomly selected expected realizations.

5.2 Simulation Session II

In the previous session, the objective function only contains the component field pattern. In this session, the component directional asymmetry (abbreviated as asymmetry hereafter) comes into play in the objective function. We use the entire field for the computation of the asymmetry. It should be noted that the evaluation cost of the objective function in this session is higher compared to that of the previous session, and so does the time needed to generate a single realization, if the same stopping criterion is employed. As pointed out at the beginning of this section, we use the same stopping criterion for both sessions.

We still adopt the three-stage evolvement when applying the PA algorithm as in Simulation Session I. The differences between realizations from Session I and Session II do exist but are not that remarkable. Therefore, in the presentation of the

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

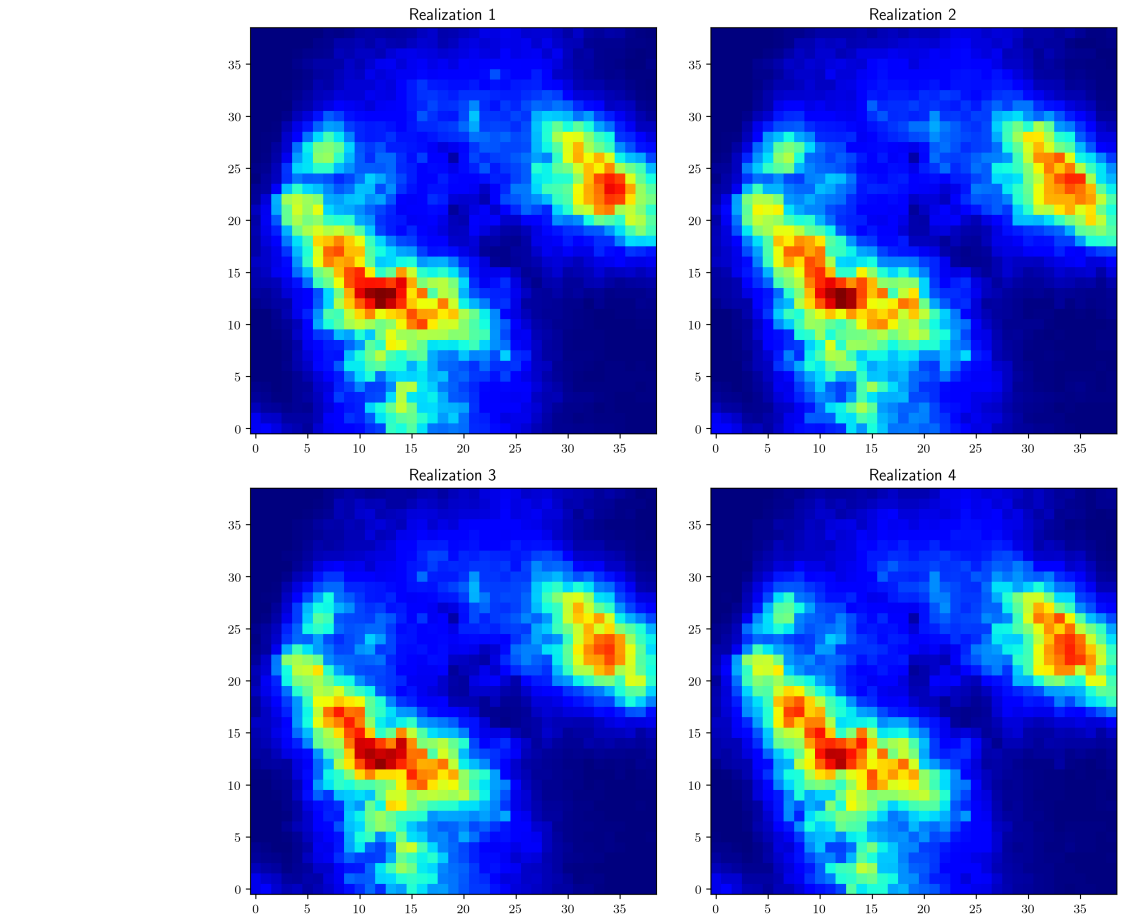


Figure 8. Four randomly selected expected realizations, sharing the same color map scale. Unit: [mm]

results from Session II, we omit the results from Stage 1 and 2, and only display the results from Stage 3 in Figure 9 and 10 (both on the right). The same results from Session I are presented on the left just for the sake of comparison and identifying the effect of adding the component asymmetry in the objective function. In the similar mode as in Session I, the mean of 100 expected realizations and the corresponding standard deviation map are shown. From Figure 9, the difference of the two mean-fields is not that noticeable. A slight reduction in the standard deviation of Session II could be observed from Figure 10, but again, not that remarkable. This is due to the fact that the two components, field pattern and asymmetry, share a special relationship: high similarity in terms of the pattern between the reference and the simulated field suggests high similarity in the asymmetry of the two as well. Yet, it might not work in reverse. Thus, the addition of the component asymmetry in the objective function seems to have a small influence on the final results.

Yet, to show the capability of the proposed algorithm in terms of fulfilling the component asymmetry, the mean of 100 simulated asymmetry functions is displayed in the middle in Figure 11, together with the reference asymmetry function on the

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

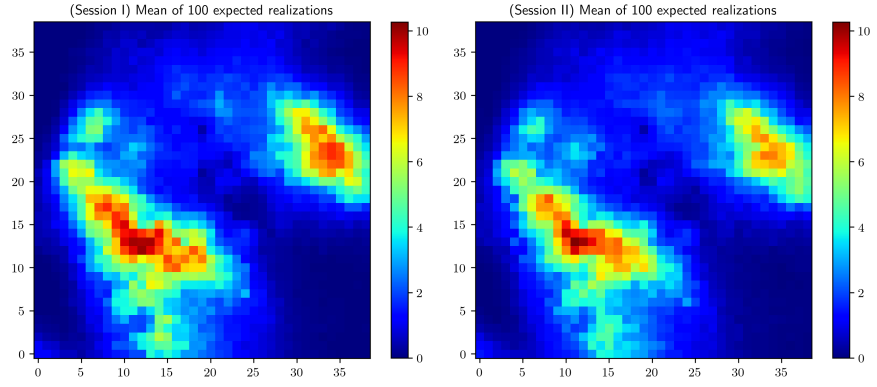


Figure 9. Mean of 100 expected realizations from Session I (left) and Session II (right). Unit: [mm]

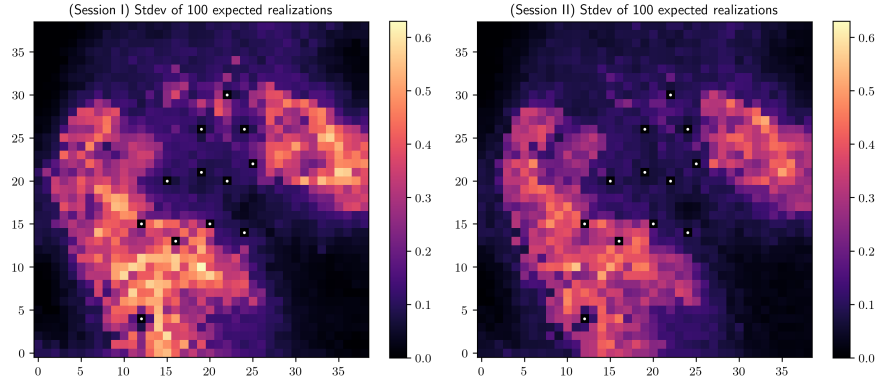


Figure 10. Standard deviation of 100 expected realizations from Session I (left) and Session II (right). Unit: [mm]

left. Note that the reference asymmetry function is the one evaluated from the reference field. From Figure 11, the reference and the mean asymmetry functions are barely distinguishable. The standard deviation map of the 100 simulated asymmetry functions is shown on the right, showing the small variability between different simulated asymmetry functions. Note that the asymmetry functions presented in Figure 11 are computed from realizations of Stage 1. But a similar standard (in terms of the fulfillment of the component asymmetry) could be achieved by the results from any stages.

It is worth mentioning that a trick is played to reduce the computational cost substantially at a fairly low cost of the estimation quality. In Section 3, when using multiple shifted quantile fields to produce the expected realization, one should be aware that these individuals have very different contributions to the final estimate. If we plot the Lorenz curve of the contribution of individual shifted fields, namely the weights indicated by the corresponding entries in the probability matrix, as shown in

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

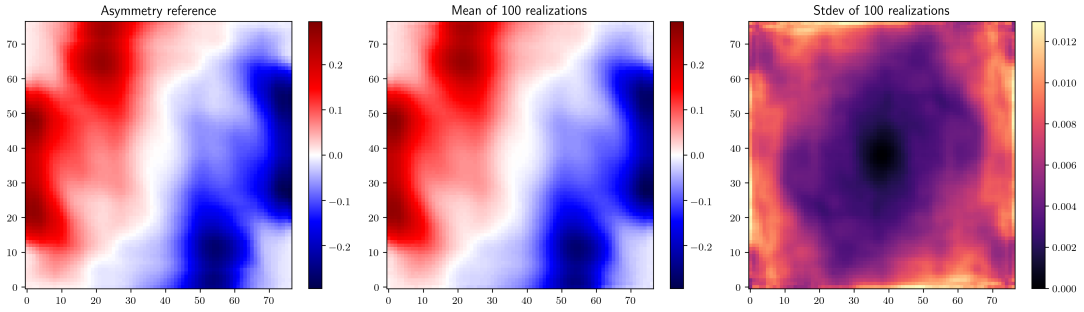


Figure 11. (Left) Reference asymmetry function. (Middle) Mean of the asymmetry functions of 100 realizations from Stage 1. (Right) Standard deviation of the asymmetry functions of 100 realizations from Stage 1. Unit: [mm]

Figure 12, one could see that the lowest several weights contribute very little to the accumulated total. Specifically, the top 9 weights (out of 23) contribute to 90%; the top 13 weights contribute to 95%; the top 18 weights contribute to 99% and the top 19 weights contribute to 99.5% of the accumulated total. Being a little bit conservative, we've chosen to use the top 19 contributors to produce the expected realization. Note that the top 19 weights should be scaled before using, such that the sum of them equals 1. We've tested this trick, using 19 to represent all (23), on the results from Session I (Stage 3). As expected, the difference in the resultant expected realizations is tiny, with the maximum difference of $0.050\mathrm{mm}$; the minimum difference of $-0.058\mathrm{mm}$ and the mean difference of $0.001\mathrm{mm}$. The results, shown in Figure 9 and 10, are produced by implementing the trick, and it does help in saving the computational cost.

6 Conclusions

The focus of this paper is to simulate rainfall fields conditioned on the local constraints subject by the point-wise rain gauge observations and the global constraints subject by the field measurement from weather radar. The innovation of this work comes in three aspects. ***First***, the separation of global and local constraints. Due to the global characteristic of PA, it is powerful in handling the global constraints. Thus, PA is only used to deal with the global constraints and the local ones are handed over to residual kriging. The separation of different constraints makes the best use of PA and avoids its insufficiency in terms of the fulfillment of the local constraints. ***Second***, the extension of the PA algorithm. Except for annealing the system temperature, the number of perturbed phases is also annealed during the simulation process, making the algorithm work more globally at initial phases. The global influence of the perturbation decreases gradually at iterations as the number of perturbed phases decreases. ***Third***, integrating the uncertainty of the radar-indicated rainfall pattern by (A) simulating using the expectation of multiple shifted fields as the reference or (B) applying the simulation independently using multiple shifted fields as the reference and combining the individual realizations as the final estimate.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

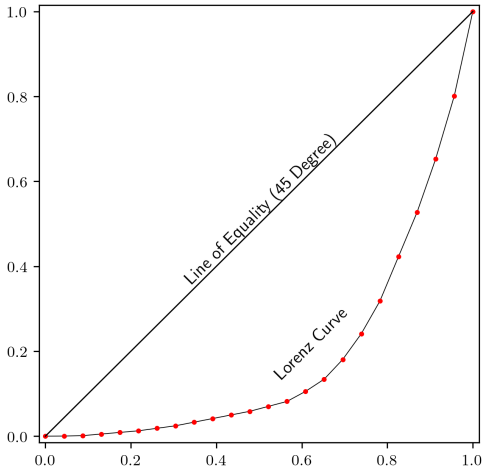


Figure 12. Lorenz curve of the contribution of individual shifted fields, where the x-axis denotes the cumulative share of population ordered by contribution from the lowest to the highest and the y-axis denotes the cumulative share of contribution.

The proposed method is used to simulate the rainfall field for a 30-min-event. The algorithm of PA is applied using different scenarios: with and without integrating the uncertainty of the radar-indicated rainfall pattern; with different objective functions. The capability of the proposed method in fulfilling the global constraints, both the field pattern and the directional asymmetry, is demonstrated by all the results. Practically, the estimates integrating the uncertainty of the radar-indicated rainfall pattern show a reduced estimation variability. And an obvious displacement of the rainfall peak is observed compared to the results without integrating the uncertainty of the radar-indicated rainfall pattern. As for the two options to integrate the uncertainty of the radar-indicated rainfall pattern, (B) seems to be superior to (A) in terms of the substantial reduction in the estimation variability and the smoothness of the final estimates. As for the two simulation sessions using different objective functions, the impact of adding the component directional asymmetry in the objective function is not that prominent. This is due to the special relationship between the two global constraints: high similarity in the field pattern is the sufficient condition for high similarity in the directional asymmetry function (though, the inverse is not true). Yet, compared to the results using the objective function containing solely the component field pattern, a slight reduction in the estimation variability is observed from the results using the objective function adding the component directional asymmetry.

Data availability. Four basic datasets required for simulating the 30-min rainfall event are archived under the DOI: 10.6084/m9.figshare.11515395. The other files or figures are secondary and generated based on the four basic datasets. The displays of the input data are provided in the python script *inputDisplay.py*.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Author contributions. The first author Jieru Yan did the programming work, all the computations and the manuscript writing. The second author András Bárdossy contributed to the research idea, and supervised the research. The third author Sebastian Hörning provided a good opportunity to improve the program in terms of the efficiency. The fourth author Tao Tao provided valuable suggestions for the revision of this article.

Competing interests. The authors declare that no competing interests are present

Acknowledgements. The radar data were kindly provided by the German Weather Service (DWD), and the rain gauge data were kindly provided by Stadtentwässerung Reutlingen.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Hydrology and Earth System Sciences
Discussions

References

Bárdossy, A. and Hörning, S.: Process-Driven Direction-Dependent Asymmetry: Identification and Quantification of Directional Dependence in Spatial Fields, Mathematical Geosciences, 49, 871–891, https://doi.org/10.1007/s11004-017-9682-1, https://doi.org/10.1007/s11004-017-9682-1, 2017.

Chilès, J. P. and Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty, 2012.

Deutsch, C., V, C., and Journel, A.: GSLIB - Geostatistical Software Library and User's Guide, Technometrics, 40, https://doi.org/10.2307/1270548, 1998.

Deutsch, C. V.: Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data, Ph.D. thesis, stanford university Stanford, CA, 1992.

Deutsch, C. V., Journel, A. G., et al.: The application of simulated annealing to stochastic reservoir modeling, SPE Advanced Technology Series, 2, 222–227, 1994.

Fabry, F.: Radar meteorology: principles and practice, Cambridge University Press, 2015.

Gabella, M. and Notarpietro, R.: Ground clutter characterization and elimination in mountainous terrain, in: Proceedings of ERAD, vol. 305, 2002.

Geman, S. and Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, 721–741, 1984.

Guthke, P. and Bárdossy, A.: On the link between natural emergence and manifestation of a fundamental non-Gaussian geostatistical property: Asymmetry, Spatial Statistics, 20, 1 – 29, https://doi.org/https://doi.org/10.1016/j.spasta.2017.01.003, http://www.sciencedirect.com/science/article/pii/S2211675317300258, 2017.

Heistermann, M., Jacobi, S., and Pfaff, T.: Technical Note: An open source library for processing weather radar data (wradlib), Hydrology and Earth System Sciences, 17, 863–871, https://doi.org/10.5194/hess-17-863-2013, 2013.

Hörning, S. and Bárdossy, A.: Phase annealing for the conditional simulation of spatial random fields, Computers Geosciences, 112, 101 – 111, https://doi.org/https://doi.org/10.1016/j.cageo.2017.12.008, http://www.sciencedirect.com/science/article/pii/S009830041730403X, 2018.

Jacobi, S. and Heistermann, M.: Benchmarking attenuation correction procedures for six years of single-polarized C-band weather radar observations in South-West Germany, Geomatics, Natural Hazards and Risk, 7, 1785–1799, 2016.

Journel, A. G.: Geostatistics for conditional simulation of ore bodies, Economic Geology, 69, 673–687, 1974.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by Simulated Annealing, Science, New Series, 220, 671–680, 1983.

Krämer, S. and Verworn, H.: Improved C-band radar data processing for real time control of urban drainage systems, in: Proceedings of the 11th International Conference on Urban Drainage, pp. 1–10, 2008.

Lantuéjoul, C.: Geostatistical simulation: models and algorithms, Springer Science & Business Media, 2013.

Lauzon, D. and Marcotte, D.: Calibration of random fields by FFTMA-SA, Computers Geosciences, 127, 99 – 110, https://doi.org/https://doi.org/10.1016/j.cageo.2019.03.003, http://www.sciencedirect.com/science/article/pii/S0098300418312044, 2019.

Marcotte, D.: Fast variogram computation with FFT, Computers Geosciences, 22, 1175 – 1186, https://doi.org/https://doi.org/10.1016/S0098-3004(96)00026-X, http://www.sciencedirect.com/science/article/pii/S009830049600026X, 1996.

Martín, J. F. D. and Sierra, J. R.: A Comparison of Cooling Schedules for Simulated Annealing, 2009.

https://doi.org/10.5194/hess-2020-4
Preprint. Discussion started: 10 January 2020

Mendez Antonio, B., Magaña, V., Caetano, E., Da Silveira, R., and Domínguez, R.: Analysis of daily precipitation based on weather radar information in México City, Atmósfera, 22, 299–313, 2009.

Nourani, Y. and Andresen, B.: A comparison of simulated annealing cooling strategies, Journal of Physics A General Physics, 31, 8373–8385, 1998.

Ravalec, M. L., Noetinger, B., and Hu, L. Y.: The FFT Moving Average (FFT-MA) Generator: An Efficient Numerical Method for Generating and Conditioning Gaussian Simulations, Mathematical Geology, 32, 701–723, https://doi.org/10.1023/A:1007542406333, https://doi.org/10.1023/A:1007542406333, 2000.

Wood, A. T.: When is a truncated covariance function on the line a covariance function on the circle?, Statistics and Probability Letters, 24, 157 – 164, https://doi.org/https://doi.org/10.1016/0167-7152(94)00162-2, http://www.sciencedirect.com/science/article/pii/0167715294001622, 1995.

Wood, A. T. and Chan, G.: Simulation of stationary Gaussian processes in [0, 1] d, Journal of computational and graphical statistics, 3, 409–432, 1994.

Yan, J. and Bárdossy, A.: Short time precipitation estimation using weather radar and surface observations: With rainfall displacement information integrated in a stochastic manner, Journal of Hydrology, 574, 672 – 682, https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.04.061, http://www.sciencedirect.com/science/article/pii/S0022169419304019, 2019.