Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Empirically-based generator of synthetic radar-rainfall data

GABRIELE VILLARINI, WITOLD F. KRAJEWSKI & GRZEGORZ J. CIACH

IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa, USA

gabriele-villarini@uiowa.edu

Abstract To fully characterize the uncertainties associated with radar-rainfall (RR) estimates, Ciach *et al.* (2007) developed an empirically based model, in which the relationship between true and radar-rainfall can be described by a deterministic distortion function and a random component. This model has the flexibility to account for different spatio-temporal resolutions, distances from the radar, synoptic conditions, and space–time dependence of the errors. Based on this model, two possible scenarios are presented and described: an ensemble generator and a static estimation of probability maps. In the former, given a time series of hourly radar-rainfall fields, a user can generate an ensemble of synthetic RR data congruent with the error model's characteristics. As far as the second scenario is concerned, given hourly RR maps, it is possible to generate fields with the probability of exceedence of some arbitrary thresholds by the true rainfall.

Key words radar-rainfall uncertainties; ensemble forecasting, NEXRAD, radar hydrology

INTRODUCTION

The capability to accurately predict streamflow is significantly dependent on the accuracy of precipitation estimates. According to NRC (2002), one of the biggest obstacles to advancing in hydrological predictability is precipitation measurements. These measurements are difficult to obtain because rainfall is a phenomenon characterized by high variability in space and time (National Academy Press, 1998). As shown in several studies (see Winchell *et al.* (1998) and references therein), it is important to account for this variability in order to obtain better streamflow predictions. Even though raingauges can accurately measure rainfall, they provide measurements not able to capture the spatial variability of the precipitation systems (e.g., Michaud & Sorooshian, 1994). On the other hand, radar estimates are characterized by high spatio-temporal resolution over a domain as large as 10^5 km^2. However, radar-rainfall (RR) estimates are affected by several sources of uncertainties (see Krajewski & Ciach (2003) for a discussion about the problem). As noted by Winchell *et al.* (1998), "…meaningful hydrologic predictions are not possible unless the uncertainty associated with the radar-derived precipitation can be quantified and corrected for". Despite the relevance of the topic, a comprehensive statistical characterization of these errors is still lacking: all the existing models describing radar-rainfall uncertainties are based on assumptions and conjectures rather than on empirical evidence (e.g. Krajewski & Georgakakos, 1985; Carpenter & Georgakakos, 2004). To address the problem, Ciach *et al.* (2007) proposed and developed a datadriven model to describe the relationship between true rainfall (RA) and radar-rainfall. This model has the flexibility to account for different spatio-temporal resolutions, distance from the radar, synoptic conditions, and space–time dependence of the errors. In this paper, based on their results we will present an empirically-based generator of synthetic radar-rainfall fields which are congruent with the model's statistical error characterization. Additionally, based on the same model we present a second scenario in which we generate maps with the probability of exceedence for some arbitrary thresholds by the true rainfall for a given hourly radar-rainfall accumulation (static estimation). These maps allow us to answer questions like "What is the probability of exceedence by the true rainfall of some specified hydrologically relevant thresholds?"

The paper is structured as follows. In the next section we briefly describe the error model, followed by the description of the generator of synthetic RR fields and the probability maps. The final section summarizes the important points made in this article and discusses possible applications of the work.

OVERVIEW OF THE RADAR-RAINFALL ERROR MODEL

To characterize the relation between RR and RA, Ciach *et al.* (2007) proposed a model in which the true rainfall can be described as the product of two components: a deterministic distortion function, accounting for the systematic biases, and a random component, which describes the random errors. The estimation of these two components is achieved through a nonparametric approach and true rainfall is approximated by rain-gauge measurements. Their results are based on a large sample of six years of Level II data (e.g. Crum *et al.*, 1993) from the Oklahoma City radar site (KTLX) processed through Build 4 of the Open Radar Product Generator Precipitation Processing System (PPS; Fulton *et al.*, 1998) and raingauge measurements from the Oklahoma Mesonet, and Agricultural Research Service Micronet. The data set was split into three seasons (cold, warm, and hot) and the radar umbrella was divided into five zones (the reader is pointed to Ciach *et al.* (2007) for a discussion on the selection of the seasons and of the zones). The model has the flexibility to account for radar range effects, synoptic conditions, different space–time resolutions, and the spatial and temporal dependence of the errors. Strictly speaking the results in Ciach *et al.* (2007) are valid for KTLX radar only and hourly (and larger) time scale and approximately 4 km resolution, and NEXRAD Z-R relationship but there is no reason why they should be significantly different for similar rainfall regimes and radar hardware (S-band).

ENSEMBLE GENERATOR

Following the notation from Krajewski & Georgakakos (1985), the synthetic radar-rainfall field $G(x, y)$ can be written as:

$$\{G(x,y)\}_{s,t} = \{S(x,y)\cdot \varepsilon(x,y)\}_{s,t} \tag{1}$$

where $S(\cdot)$ is the deterministic distortion function, $\varepsilon(\cdot)$ is the random component, x and y are the field-point coordinates, while the subscripts s and t account for different season and time scale. The coordinates x and y uniquely identify each zone. Notice that the zero-rainfall areas are preserved, due to the multiplicative nature of the generator.

Given the original hourly radar rainfall field (DPA) field $RR(x, y)$, the deterministic distortion function was parameterized as:

$$\{S(x,y)\}_{s,t} = \{a[B \cdot RR(x,y)]^b\}_{s,t} \tag{2}$$

where a and b are the deterministic distortion model parameters and B is the overall bias.

The random component $\varepsilon(x, y)$ can be described by a Gaussian distribution with mean equal to 1, standard deviation $\sigma_e(RR(x,y))$, spatial correlation $\rho_s(\Delta s)$ and temporal correlation $\rho_t(\Delta t)$. The standard deviation of the random component was parameterized as:

$$\{\sigma_e(RR(x,y))\}_{s,t} = \left\{c + \frac{d}{[B \cdot RR(x,y)]^e}\right\}_{s,t} \tag{3}$$

where c, d and e are the random component parameters.

The correlation in space and time was parameterized using a two-parameter exponential function:

$$\{\rho_s(\Delta s)\}_{s,t} = \left\{\exp\left[-\left(\frac{\Delta s}{a_0}\right)^{b_0}\right]\right\}_{s,t} \tag{4}$$

$$\{\rho_t(\Delta t)\}_{s,t} = \left\{\exp\left[-\left(\frac{\Delta t}{a_1}\right)^{b_1}\right]\right\}_{s,t} \tag{5}$$

where a_0, and b_0, are the values of the parameters for the correlation in space while a_1, and b_1 are the values of the parameters for the correlation in time, and Δs and Δt are the separation lags in space and time respectively.

In the literature, several methods have been proposed to generate correlated Gaussian random fields (e.g. Gneiting *et al.*, 2005). For our purposes, we have selected the Cholesky decomposition (CD) method (e.g. Cressie, 1993). The advantages of this method are its exactness (the ensemble of simulated fields has exactly the required correlation structure) and its flexibility to account for correlation in space and time and variance dependent on RR estimates. This methodology is based on the Cholesky decomposition of the variance–covariance matrix. One condition for the use of CD is that the matrix has to be definite positive and symmetric. In this case, this condition is satisfied because the variance–covariance matrix from an exponential correlation function is definite positive and symmetric (e.g. Journel & Huijbregts, 1978).

Let D be the domain over which we want to simulate the random field. In our case it is a regular grid with side n. As written above, the random process $\{\boldsymbol{\varepsilon}(\mathbf{s}); \mathbf{s} \in D\}$ has mean:

$$\boldsymbol{\mu}(\mathbf{s}) \equiv 1, \qquad \mathbf{s} \in D \tag{6}$$

As far as the covariance is concerned, we assume a separable spatio-temporal model and we can write it as:

$$C(\mathbf{s}_i - \mathbf{s}_j;\ \Delta t) = C(\mathbf{s}_i - \mathbf{s}_j) \cdot C(\Delta t) \qquad \mathbf{s}_i, \mathbf{s}_j \in D \tag{7}$$

To account for dependence of the variance on RR estimates, we can multiply the correlation matrix by $\{\sigma_e(RR(\mathbf{s}_i))\cdot\sigma_e(RR(\mathbf{s}_j))\}$ and the variance-covariance matrix at time t_0 can be written as:

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho_s(\|\mathbf{s}_{1,2}\|)\rho_t(t_0) & \cdots & \sigma_1\sigma_{n\cdot n}\rho_s(\|\mathbf{s}_{1,n\cdot n}\|)\rho_t(t_0)\\ \sigma_2\sigma_1\rho_s(\|\mathbf{s}_{2,1}\|)\rho_t(t_0) & \sigma_2^2 & \cdots & \sigma_2\sigma_{n\cdot n}\rho_s(\|\mathbf{s}_{2,n\cdot n}\|)\rho_t(t_0)\\ \cdots & \cdots & \cdots & \cdots\\ \sigma_{n\cdot n}\sigma_1\rho_s(\|\mathbf{s}_{n,n}\|)\rho_t(t_0) & \sigma_{n\cdot n}\sigma_2\rho_s(\|\mathbf{s}_{n\cdot n,2}\|)\rho_t(t_0) & \cdots & \sigma_{n\cdot n}^2\end{bmatrix} \tag{8}$$

At time t_1 we have:

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho_s(\|\mathbf{s}_{1,2}\|)\rho_t(\|t_{1,0}\|) & \cdots & \sigma_1\sigma_{n\cdot n}\rho_s(\|\mathbf{s}_{1,n\cdot n}\|)\rho_t(\|t_{1,0}\|)\\ \sigma_2\sigma_1\rho_s(\|\mathbf{s}_{2,1}\|)\rho_t(\|t_{1,0}\|) & \sigma_2^2 & \cdots & \sigma_2\sigma_{n\cdot n}\rho_s(\|\mathbf{s}_{2,n\cdot n}\|)\rho_t(\|t_{1,0}\|)\\ \cdots & \cdots & \cdots & \cdots\\ \sigma_{n\cdot n}\sigma_1\rho_s(\|\mathbf{s}_{n\cdot n,1}\|)\rho_t(\|t_{1,0}\|) & \sigma_{n\cdot n}\sigma_2\rho_s(\|\mathbf{s}_{n\cdot n,2}\|)\rho_t(\|t_{1,0}\|) & \cdots & \sigma_{n\cdot n}^2\end{bmatrix} \tag{9}$$

where $s_{i,j}=s_i-s_j$ and $t_{i,0}=t_0-t_i$.

Applying the Cholesky decomposition to the variance-covariance matrix, we obtain $\Sigma = LL'$. It is possible to simulate $\boldsymbol{\varepsilon}$, which satisfy (6) and (8), as:

$$\boldsymbol{\varepsilon}=\boldsymbol{\mu}+\mathbf{L}\boldsymbol{\eta} \tag{10}$$

where $\boldsymbol{\eta}$ is a vector of uncorrelated random variables with mean equal to zero and variance of 1:

$$\boldsymbol{\eta}\equiv(\eta(\mathbf{s}_1),\ \ldots,\ \eta(\mathbf{s}_{n\cdot n}))' \qquad \mathbf{s}_1,\ldots,\mathbf{s}_{n\cdot n}\in D \tag{11}$$

If the generated value of $\varepsilon(\mathbf{s})$ is negative, we set it to zero. In Fig. 1 we have plotted the original DPA map (top panel), and four radar-rainfall fields consistent with the error model (middle and bottom panels). The patterns of all of the four simulated RR fields seem plausible and consistent with the RR map.

Therefore, given $RR(x, y)$ at t_0, to generate an ensemble of m synthetic radar-rainfall fields we have to multiply $S(x, y)$ by m realization of $\varepsilon(x, y)$ obtained from decomposing the variance-covariance matrix in (8). Given a DPA map at t_1, we multiply the deterministic distortion function at t_1 with the simulated Gaussian fields from the decomposition of (9).

STATIC ESTIMATION

As mentioned earlier, the statistical distribution of the random component can be described by the Gaussian distribution. Therefore, we are able to estimate the probability of exceedence of some arbitrary rainfall values or threshold *THRES* by the true rainfall at the pixel level. We can write:

$$P(X\geq THRES)=1-\Phi(X) \tag{12}$$

where $\Phi(X)$ is the normal cumulative distribution function (CDF), X is:

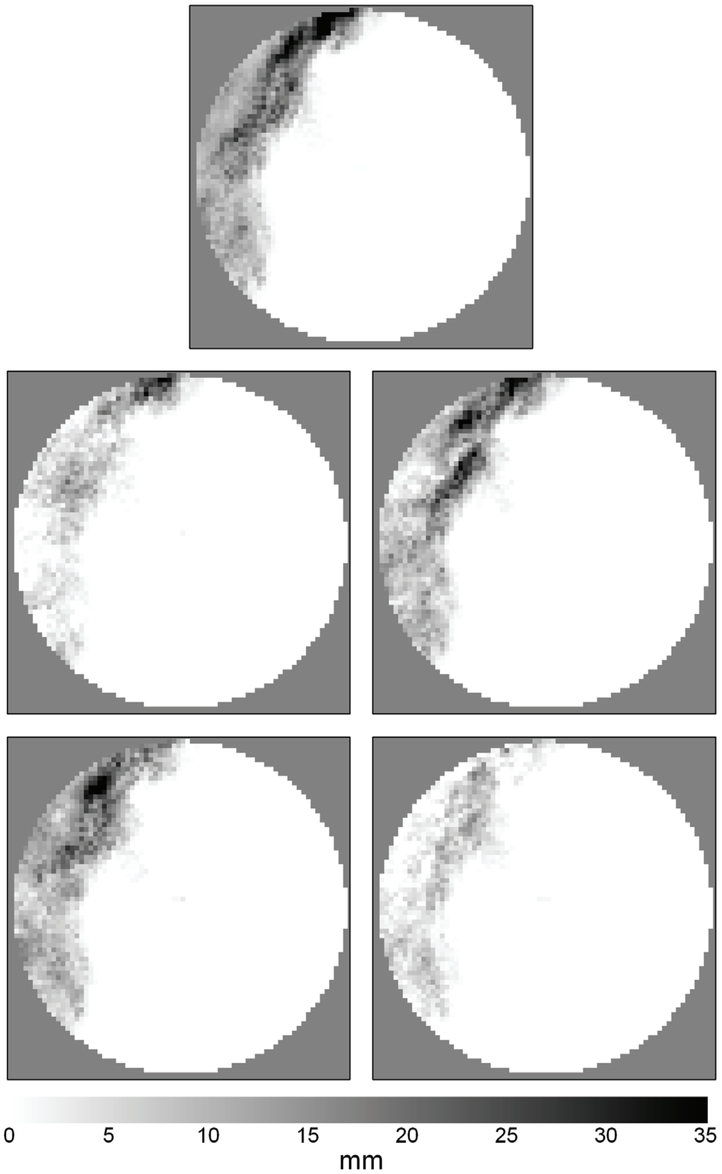

Fig. 1 Original DPA map (upper panel), and four synthetic radar-rainfall fields consistent with the error model's characteristics (middle and bottom panels).

$$X = \frac{x-\mu}{\sigma\sqrt{2}} = \frac{\dfrac{THRES}{DPA^*} - 1}{\sigma\sqrt{2}} = \frac{\dfrac{THRES}{a(B \cdot RR)^b} - 1}{\sqrt{2}\left[c + d(B \cdot RR)^e\right]} \tag{13}$$

where DPA^* represents the deterministic distortion function. As mentioned above, the mean μ of the random component is equal to 1.

One can write the normal CDF in terms of the *erf* function as:

$$\Phi(X) = \frac{1}{2} + \frac{1}{2} erf(X) \tag{14}$$

From (13) and (14), if the values of the deterministic distortion function are larger (smaller) than the threshold, we have $P(X \geq THRES) > 0.5$ (or smaller than 0.5). When DPA^* is equal to *THRES*, $P(X \geq THRES) = 0.5$.

As an example, consider the hourly rainfall map in Fig. 2(a). Figure 2(b) shows the corresponding values of the deterministic distortion function. In this case we have selected a threshold equal to 35 mm. In Fig. 2(c) we have plotted the map with the probabilities of exceedence. As expected, all the values are smaller than 0.5 because $DPA^* < 35$ mm. If we had given a smaller threshold value, the probabilities of exceedence would have been larger. In addition, given certain probabilities thresholds, it is possible to create maps showing only the pixels exceeding them. For the hourly radar-rainfall map in Fig. 2(a), we have plotted the case for $P(X \geq THRES) \geq 0.10$ (Fig. 2(d). We can notice how those pixels correspond to the ones with larger RR values.

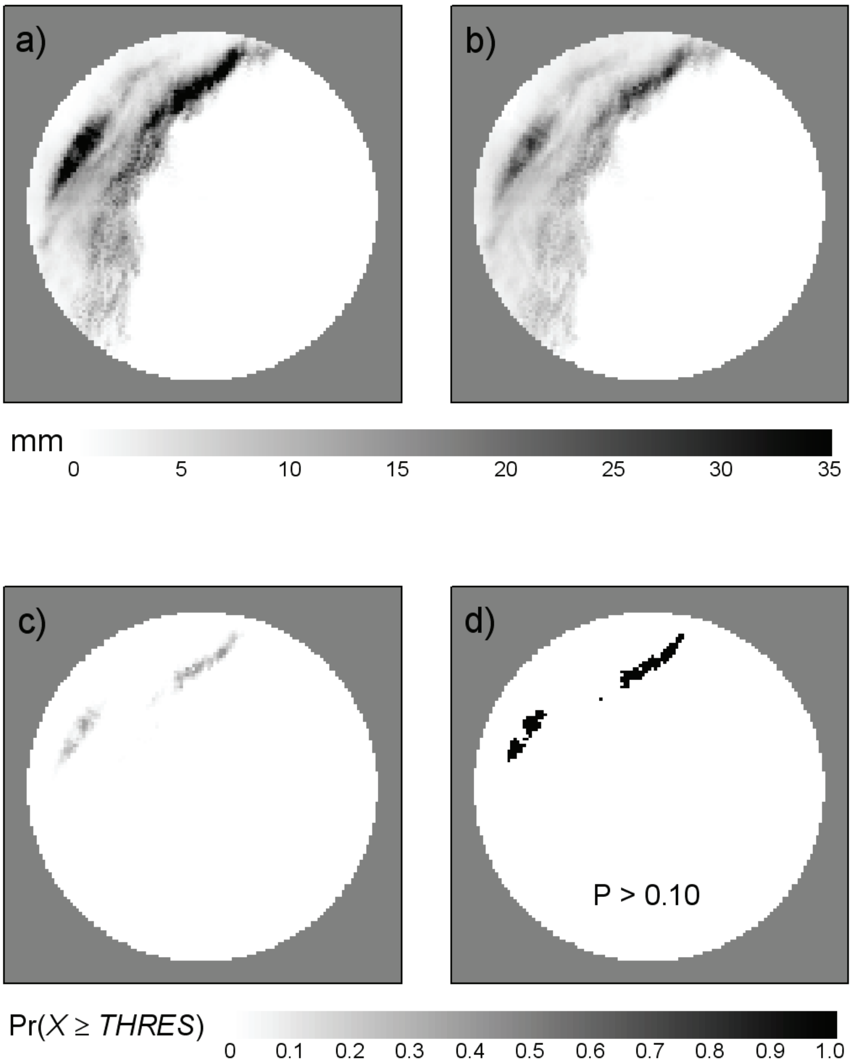

Fig. 2 Images of the original DPA map (a), of the deterministic distortion function (b), of the probability of exceedence of a threshold of 35 mm (c), and map of $P(X \geq THRES) \geq 0.10$ (d).

DISCUSSIONS AND CONCLUSIONS

The accuracy of streamflow predictions is largely dependent on measurement of precipitation, a phenomenon characterized by high spatio-temporal variability. Even though raingauges can accurately measure rainfall with high temporal resolution, in the vast majority of the cases the existing networks do not have adequate density to capture the spatial variability of the precipitation system. Radar has this capability but radar-rainfall estimates are affected by several sources of uncertainty that need to be accounted for in order to improve streamflow predictions. Ciach *et al.* (2007) proposed a model able to describe the relationship between true and radar-rainfall. In this paper, based on their results, we have developed a generator of radar-rainfall fields consistent with the error model's characteristics. An ensemble of the generated fields could be used in all those models for which radar-rainfall is used as input, such as hydrologic models, or for which RR estimates are used as initial condition (e.g. precipitation forecast), improving our understanding and interpretation of the obtained results. As far as the static estimation is concerned, given a RR field, it allows the construction of maps detailing the probability that the true rainfall exceeds an arbitrary threshold over an arbitrary domain within the radar umbrella. Such maps could be very useful to the flash-flood forecasting problem (Georgakakos, 2006; Ntelekos *et al*., 2006).

Acknowledgements This work was supported by the NSF grant EAR-0309644 and the contract under the PQPE initiative by the Office of Hydrologic Development of the National Weather Service. We acknowledge many useful discussions with Richard Fulton, David Kitzmiller, and members of the PQPE Advisory Team. The opinions expressed in this work are those of the authors and do not necessarily reflect those of NSF or NOAA or their sub-agencies.

REFERENCES

Carpenter, T. & Georgakakos, K. P. (2004) Impact of parametric and radar rainfall uncertainties on the ensemble streamflow simulations of a distributed hydrologic model. *J. Hydrol.* **298**, 202–221.

Ciach, G. J., Krajewski, W. F. & Villarini, G. (2007) Product-error-driven uncertainty model for probabilistic quantitavie precipitation estimation with NEXRAD data. *J. Hydromet.* **(**in press**)**.

Cressie, N. A. C. (1993) *Statistics for Spatial Data.* Wiley & Sons, Chichester, UK.

Crum, T. D., Alberty, R. L. & Burgess, W. (1993) Recording, archiving, and using WSR-88D data. *Bull. Am. Met. Soc.* **74**, 645–653.

Fulton, R. A., Breidenbach, J. P., Seo, D. –J. & Miller, D. A. (1998) WSR-88D rainfall algorithm. *Weather and Forecasting* **13**, 377–395.

Georgakakos, K. P. (2006) Analytical results for operational flash flood guidance. *J. Hydrol.* **317**, 81–103.

Gneiting, T., Sevcikova, H., Percival, D. B., Schlather, M. & Yindeng, J. (2005) Fast and exact simulation of large Gaussian lattice systems in R2: Exploring the limits. Report no. 477. Department of Statistics, Univ. Washington, USA.

Journel, A. G. & Huijbregts, Ch. J. (1978) *Mining Geostatistics*. Academic Press, New York, USA.

Krajewski, W. F. & Ciach, G. J. (2003) Towards probabilistic quantitative precipitation WSR-88D algorithms: Preliminary studies and problem formulation. Report, NWS Office of Hydrologic Development, Silver Spring, Maryland, USA.

Krajewski, W. F. & Georgakakos, K. P. (1985) Synthesis of radar-rainfall data. *Water Resour. Res.* **21**(5), 764–768.

Michaud, J. D. & Sorooshian, S. (1994) Effects of rainfall-sampling errors on simulations of desert flash floods. *Water Resour. Res.* **30**, 2765–2775.

National Academy Press (1998) *Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP): A review of progress and opportunities*. Washington, DC, USA.

NRC (2002) *Report of a Workshop on Predictability and Limits-to-Prediction in Hydrologic Systems*. National Academy Press, Washington, DC, USA.

Ntelekos, A. A., Georgakakos, K. P. & Krajewski, W. F. (2006) On the uncertainties of flash flood guidance: toward probabilistic forecasting of flash floods. *J. Hydromet.* **7**, 896–915.

Winchell, M., Gupta, H. V. & Sorooshian, S. (1998) On the simulation of infiltration- and saturation-excess runoff using radar-based rainfall estimates: effects of algorithm uncertainty and pixel aggregation. *Water Resour. Res.* **34**(10), 2655–2670.