Conditioning Stochastic Rainfall Replicates on Remote Sensing Data

Citation: Wojcik, R. et al. "Conditioning Stochastic Rainfall Replicates on Remote Sensing Data." Geoscience and Remote Sensing, IEEE Transactions on 47.8 (2009): 2436-2449. © 2009, IEEE

As Published: http://dx.doi.org/10.1109/tgrs.2009.2016413

Publisher: Institute of Electrical and Electronics Engineers, IEEE Geoscience and Remote Sensing Society

Persistent URL: http://hdl.handle.net/1721.1/60256

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Conditioning Stochastic Rainfall Replicates on Remote Sensing Data

Rafał Wójcik, Dennis McLaughlin, Alexandra G. Konings, and Dara Entekhabi, *Senior Member, IEEE*

***Abstract*—Temporally and spatially variable rainfall replicates are frequently required in hydrologic applications of ensemble forecasting and data assimilation. Ensemble methods can be expected to work better when the rainfall replicates more closely resemble observed storms. In particular, the replicates should capture the intermittency and variability that are dominant features of rainfall events. In this paper, we present a new probabilistic procedure for generating realistic rainfall replicates that are constrained by (or conditioned on) remote sensing measurements. The procedure uses remotely sensed cloud top temperatures to identify potentially rainy regions. The cloud top temperatures are obtained from visible/infrared instruments in geostationary orbit. A multipoint geostatistical algorithm generates areas of nonzero rain (rain clusters) within each cloudy region. This algorithm relies on statistics derived from ground-based weather radar [National Operational Weather Radar (NOWRAD)] data. A truncated multiplicative cascade generates rain rates within each rain cluster. A computational experiment based on summer 2004 data from the Central U.S. indicates that the rainfall replicates simulated by the procedure are visually and statistically similar to individual NOWRAD images and to a large ensemble of NOWRAD images collected throughout the summer simulation period.**

***Index Terms*—Data assimilation, Geostationary Operational Environmental Satellites (GOES), multiple-point geostatistics, multiscale tree, National Operational Weather Radar (NOWRAD), precipitation, stochastic simulation.**

I. Introduction

ENSEMBLE (or Monte Carlo) approaches provide a flexible and convenient way to investigate the role of uncertainty in a wide range of geophysical applications. In meteorology and hydrology, ensemble methods have been widely used for both forecasting and data assimilation [1]–[4]. The basic concept is to simulate many possible responses of a system by varying uncertain inputs over a range of reasonable values. For example, variations in land surface fluxes can be investigated by using a hydrologic model to simulate responses to different precipitation inputs. Sample statistics derived from the simulated replicates provide useful information about the probability of extreme events and long-term trends. They also can be used to derive improved predictions.

In an analysis of hydrologic response, it is important to insure that the precipitation replicates that force the land surface system look as much as possible like real rain events. In fact, it is reasonable to ask that there be no obvious difference between the randomly generated replicates and observed events. In order to develop a systematic way to generate realistic rainfall replicates, we need to specify with some precision what constitutes realism. One of the most notable features of real rainfall is its spatial and temporal intermittency. Intermittency presents a significant challenge for ensemble methods that need to generate large numbers of rainfall replicates that properly represent uncertainty while remaining physically realistic.

Methods for simulating rainfall divide naturally into physics-based meteorological models and stochastic models. Physics-based models generate rainfall replicates by perturbing the initial and/or boundary conditions in primitive equations based on mass, momentum, and energy conservation [5]. The computational demands of this approach make it impractical for most ensemble applications. The alternative is to use a stochastic model that reproduces the observed space-time structure of rainfall without simulating the physical processes responsible for this structure. Typical examples of the stochastic approach include multifractal models and scaling laws [6], [7], multiplicative cascade models [8]–[10] which belong to a broader class of multiscale tree models [11], clustered point processes [12], and wavelet models [13]. Most stochastic models are unable to limit rainfall to specified spatial supports. Such a capability is required if rainfall data are to be conditioned on remote sensing data.

This paper describes a new approach for generating spatially and temporally intermittent rainfall replicates that are constrained by remote sensing observations. Our objective in this paper is to generate replicates that properly represent our knowledge about the current spatial distribution of rainfall intensity. We assume that this knowledge includes real-time visible/infrared remote sensing measurements obtained from a geostationary satellite (available at frequent intervals with nearly global coverage). Such measurements provide reasonably accurate information on the location of cloudy regions where rainfall may occur, but they are not able to reliably identify rain clusters within these regions. Ensembles of possible rainfall replicates conditioned only on geostationary visible/infrared measurement should display considerable variability, particularly regarding the location and intensity of rain within cloudy regions.

When uncertainty is large and the rainfall ensemble is highly variable, it may be advisable to introduce additional site-specific information to reduce uncertainty and obtain a better description of current conditions. Such information could include ground-based radar [14], disdrometer measurements [15], and low orbit satellite measurements [16], [17]. These data sources could be used to further constrain the set of possible

Manuscript received September 11, 2008; revised December 9, 2008. First published May 5, 2009; current version published July 23, 2009. This work was supported by National Science Foundation under Awards 0121182, 0530851, and 0540259.

The authors are with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: rwojcik@mit.edu; dennism@mit.edu; konings@mit.edu; darae@mit.edu).

Digital Object Identifier 10.1109/TGRS.2009.2016413

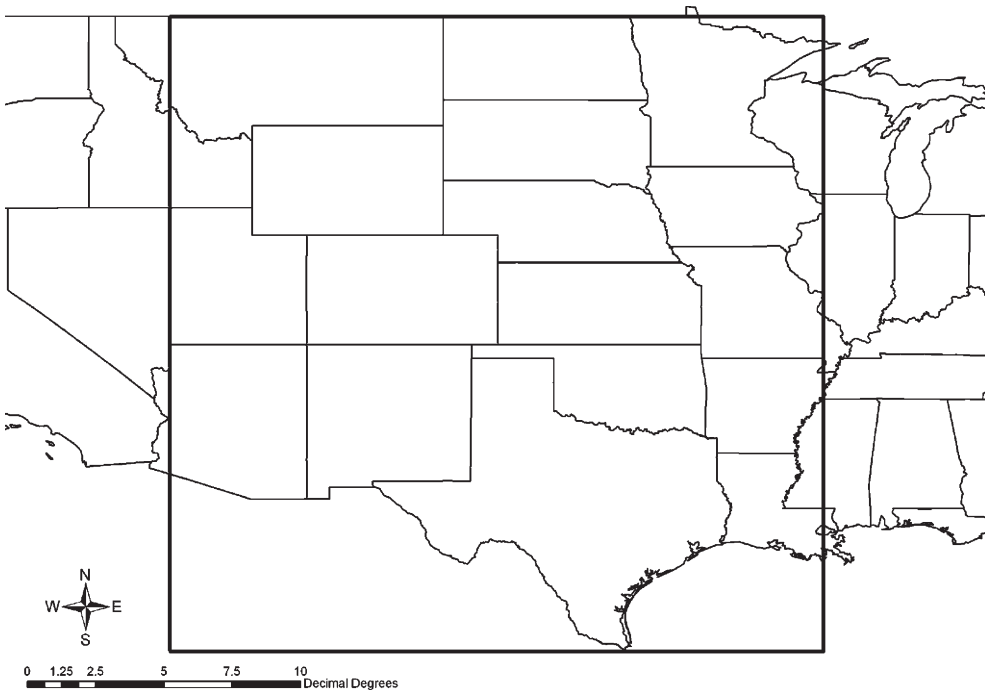


Fig. 1. Central U.S. study region is the region inside the black square.

replicates, using a Bayesian estimation procedure (such as an ensemble Kalman filter or a particle filter) that treats our rainfall ensemble as prior information. As more data are included, we expect the ensemble to become less variable and narrow in on the true rainfall field. When coupled with a predictive model, an improved ensemble could provide the basis for quantitative precipitation forecasts. This is, however, not the focus of this paper. Here, we are interested in generating a realistic rainfall ensemble that describes uncertainty about current conditions, conditioned only on geostationary data. This ensemble is intended to provide prior information for uncertainty analyses or for further Bayesian conditioning.

Our approach is best suited for situations in which clusters of short-term high-intensity rain cells are embedded within larger regions of low rain intensity that are characterized by longer characteristic time scales [10]. Clustered mesoscale rainfall systems are major contributors to summertime rainfall over the Central U.S. [18]. Here, we use the term "rain cluster" or just "cluster" to refer to a set of pixels with nonzero rain rate that is completely surrounded by pixels with zero rain rate. The "cluster support" is the set of pixels inside the cluster. The cluster is completely characterized by its support and by the rain rates at pixels in the support. Both of these attributes may vary over time. In our approach, rain cluster supports are obtained from a multipoint geostatistical algorithm that relies on training images constructed from weather radar measurements. Cluster rain rates are obtained from a multiplicative cascade model.

The operation and performance of our stochastic rainfall model are demonstrated here with data for summer 2004 from the Central U.S. region shown in Fig. 1. The Central U.S. test indicates that the model generates cluster geometries and rain rates similar to those observed in weather radar images.

II. Data Sets

The geostationary data used to condition our rainfall ensemble are obtained from the Geostationary Operational Environmental Satellites (GOES) data set produced by the U.S. National Oceanic and Atmospheric Administration. The measurements of particular interest are from the so-called "window channel" on the GOES infrared imager. The radio-brightness temperature for this channel is measured in a window centered on a wavelength of 10.7 μm, with a range from 10.23 to 11.24 μm. This measurement provides a reasonably accurate indication of cloud-top temperature (T_{top}) [19]. GOES data are available every 30 min, but we rely on hourly samples due to the large number of corrupt or missing data. The spatial resolution is 4 km. The GOES-12 instrument is our primary data source, with data from the GOES-10 satellite used to fill in times where hourly GOES-12 measurements are missing.

Our rainfall generation procedure uses training images to define the statistical properties of rain clusters within the GOES region. For the Central U.S. application, we obtain these images from the National Operational Weather Radar (NOWRAD) weather radar product. It is important to note that individual replicates generated by our procedure are not constrained by the weather radar data—these data are only used to determine rainfall statistics, as described below. Consequently, our procedure can be applied in regions or at times where weather radar data are unavailable, provided that the required statistics can be inferred from other areas and/or times.

NOWRAD is a Weather Services International (WSI) Corporation enhancement of the Next Generation Weather Radar (NEXRAD) data set [14], [20] obtained from the National Weather Service's ground-based WSR-88D radar network. WSI superimposes the assorted radars that are part of NEXRAD in a mosaic, removes some data artifacts (using both automated and manual intervention), and then uses an intensity–reflectivity

$(Z-R)$ relationship based on lookup tables to provide a final rain rate product [21]. NOWRAD data are available every 15 min, quantized in 16 levels of radar reflectivity (Z_e) factor data defined at 5-dB Z_e intervals [22]. Spatial resolution of the NOWRAD product is 2 km.

For our example, the GOES and NOWRAD data sets are both interpolated onto a common 0.05° (∼4 km) spatial grid covering the Central U.S. region ranging from 25.85°–49.01° N and 114.07°–90.12° W. This region is shown in Fig. 1. The period of study was summer 2004 (June 1–August 31). Further details on the methods used to identify GOES cloud regions and the properties of rainfall clusters within these regions are provided in Section III.

III. Rainfall Simulation Procedure

Our procedure for generating rainfall replicates consists of three steps.

1) Use GOES $T_{\rm top}$ data to identify cloudy regions where rainfall may occur. All members of the rainfall ensemble are conditioned on the same GOES data, and rain is permitted only within the GOES cloudy regions. Much of the area inside these cloudy regions may have no rainfall.
2) Use a multipoint geostatistical technique based on training images derived from NOWRAD data to generate the spatial support for rain clusters within each GOES region. Each replicate in the ensemble is characterized by a different set of rainy clusters. Rain rates are nonzero everywhere within each cluster.
3) Use a multiplicative cascade with parameters estimated from NOWRAD data to generate rain rates within each rain cluster for each replicate.

The GOES cloudy regions, training images, and rain supports are all described by indicator values (zeros and ones) at the pixels of the 0.05° computational grid previously mentioned. The entire ensemble generation process is repeated for each time in the period of interest (1 h for our example).

As time progresses, the changing GOES images account for the temporal evolution of the large-scale rainfall field. The rain cluster supports and the rain rates within clusters change over time but are temporally uncorrelated, reflecting the short characteristic time scales of convective rainfall described by the clustering process. Further details on each step of our procedure are described below.

A. Identification of Cloudy Regions Where Rainfall May Occur

This step of the procedure uses GOES cloud top temperatures to identify regions where rain may be occurring. These regions are defined to be the generally disconnected areas where $T_{\rm top} < T_0$, where T_0 is a specified threshold. The fractions of rainy areas occurring inside and outside the thresholded region depend on the threshold value selected. This dependence is shown in Fig. 2, using Central U.S. GOES and NOWRAD data from summer 2004. Fig. 2 shows the probability (or area fraction) that the observed NOWRAD rain rate is zero outside the GOES region versus the probability (or area fraction) that the NOWRAD rain rate is greater than zero inside the GOES region, for various GOES T_0 values. Ideally, we would like both probabilities to be 1.0, so that rainfall occurs everywhere

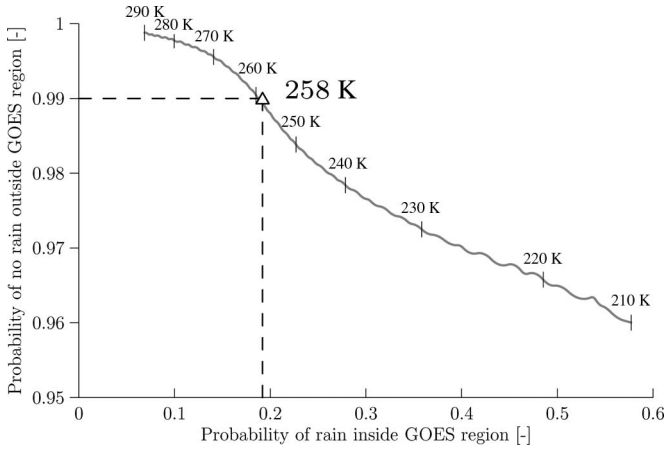


Fig. 2. Tradeoff curve for GOES $T_{\rm top}$ threshold selection. The threshold of $T_0 = 258$ K (shown as triangle) gives a low probability of 0.01 for rain outside the GOES region and a moderate probability of around 0.20 for rain inside the GOES region.

inside the GOES region and nowhere outside it. Unfortunately, these are conflicting objectives. Fig. 2 can be viewed as a multiobjective tradeoff curve that shows how much we must decrease one probability in order to increase the other. In the experiments described here, we use a threshold of $T_0 = 258$ K, which gives a low probability of 0.01 for rain outside the GOES region and a moderate probability of around 0.20 for rain inside the GOES region. This choice supports our subsequent assumption that rain does not occur outside the 258-K GOES region. As a result, we only generate rain clusters inside this region in the subsequent steps of the procedure. In other applications, it may be reasonable to choose a threshold that gives different probabilities. Note that the selected T_0 is clearly dependent on the lowest NOWRAD quantization level ($1.2\ {\rm mm \cdot h^{-1}}$). If this quantization level had been lower, the total rainy area in the NOWRAD image would have increased making the probability of rain inside the GOES region (for a given T_0) higher. Conversely, the probability of no rain outside the GOES region would have slightly decreased simply because the larger NOWRAD rain clusters could lead to larger mismatch between GOES and NOWRAD rain support.

B. Generation of Rain Support Within Each Cloudy Region

As indicated above, 2004 summer rainfall in the Central U.S. occurs in scattered clusters that cover only about 20% of the area inside the $T_0 = 258$ K GOES region selected for this paper. GOES data do not provide reliable information about the location of these clusters or the intensity distribution within each cluster. We use a multipoint geostatistical technique to generate realistic replicates of possible cluster configurations within the GOES cloud region. The technique is based on an algorithm that derives probabilities of particular rainfall patterns from NOWRAD training images [23], [24].

Fig. 3 shows some typical Central U.S. NOWRAD rainfall clusters at different times during summer 2004. The upper left panel in the figure is a blowup of one NOWRAD image with nonzero rain rates indicated by varying colors set in a dark blue background of zero rain rates. The upper right panel shows the rain cluster supports for this image (rain area is blue and no rain area is yellow).

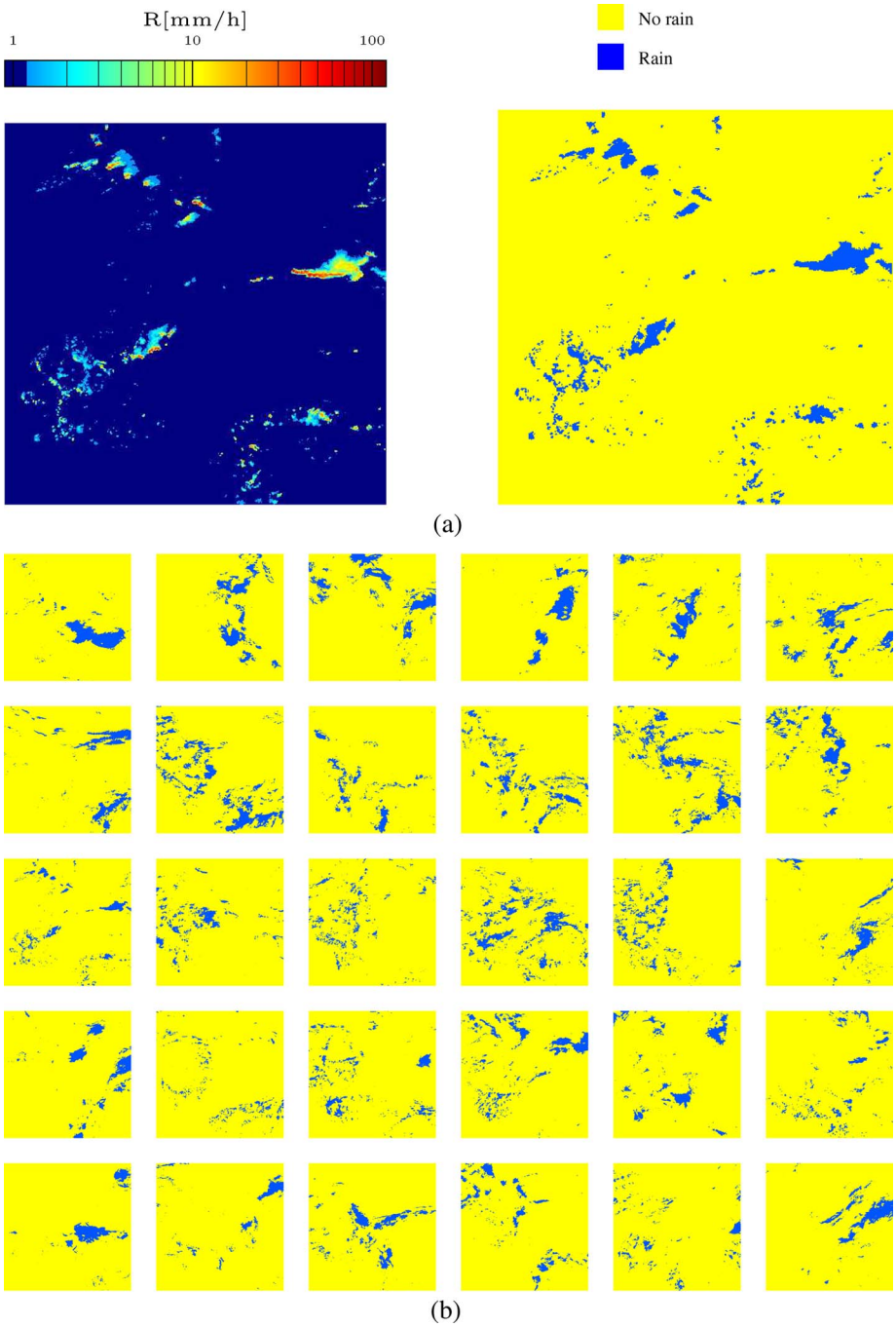


Fig. 3. (a) Typical NOWRAD signature of a summer storm that occurred on July 12, 2004, 00:00 UTC over (left panel) the Central U.S. region and its (right panel) binary version. The latter is an example of a training image which generates rain support replicates. (b) Entire ensemble of 30 training images used in this paper. The truncated logarithmic color scale in the left panel of (a) was constructed such that zero rain rates are represented as dark blue. Rain rates at NOWRAD quantization threshold (1.2 mm $\cdot$ h^{-1}) are represented with the light blue color. The maximum of the color scale corresponds to the maximum NOWRAD rain rate (120 mm $\cdot$ h^{-1}) observed in the study period.

In our experiment, there are 8796 different NOWRAD images for summer 2004. We sampled without replacement from this population to obtain a working set of 550 images. The working set sampling process was constrained to insure that the time separation between any two sampled images was at least 2 h. This was done to reduce the effect of temporal correlation between storms. Most summer rain events come from convective cells embedded in larger mesoscale features. These cells are induced by convective clouds, mainly of cumulonimbus type, with durations of about 2 h [25]–[28].

To obtain a set of training images, we selected a subgroup of 30 images from the NOWRAD working set and then identified the cluster supports for each of the 30 samples. The 30 training images, which are shown in Fig. 3(b), were selected to provide a representative cross section of storms that occurred during the summer 2004 study period.

The objective of the multipoint geostatistical procedure is to generate cluster support replicates that have the same spatial structure as the NOWRAD training images (e.g., that have the same general size, shape, and spatial density). The key to

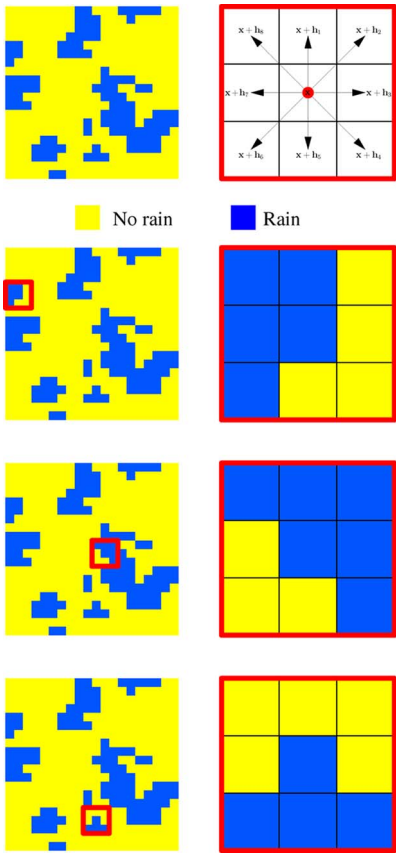


Fig. 4. Identifying template patterns in multipoint geostatistical simulation. (Upper left panel) An example of a training image. (Upper right panel) Nine-pixel template. (Panels in the lower three rows on the left) Moving the template (outlined in red) at random over the training image. (Panels in the lower three rows on the right) Blowups that show the pixel configuration within the randomly moving template.

generating realistic cluster supports is to describe the collection of possible cluster geometries (contiguous groups of rain pixels) in a concise mathematical form. This can be done by identifying a finite collection of possible pixel patterns and assigning a probability to each pattern, based on the number of times it is observed to occur in the training image. Multipoint geostatistical techniques consider patterns composed of a small number of pixels contained within a template that can be superimposed on the training image. The procedure is shown in Fig. 4.

The upper left panel of Fig. 4 is a small training image used to generate template patterns. The panels in the lower three rows on the left side of the figure show a small nine-pixel template (outlined with a thick line) that can accommodate $2^9 = 512$ possible patterns of rain pixels. The template is centered on a different location in each panel. The corresponding panels on the right side of the figure are blowups of the pixel configuration within the template for this location. As the template is moved over the training image in a specified path (usually random), a tally is kept of the number of times each of the possible 512 patterns is observed (including the particular three patterns identified in Fig. 4).

After the entire training image has been scanned, the sample probabilities can be used to generate replicates with rain pixel patterns that are similar but not identical to those observed in the NOWRAD training image. As might be expected, the generated patterns are more realistic when the template they depend upon is larger. However, the computational effort required grows dramatically as the template size is increased. This conflict can be mitigated somewhat if a multigrid template approach is used [29].

The method used to derive pattern probabilities from template data is based on Bayes rule. Suppose that we define a reference pixel in the center of the template. The center of the reference pixel is temporarily fixed at location $\mathbf{x}$ in the training image, as indicated by the red dot in the template blowup in the upper right panel of Fig. 4. Let $Z(\mathbf{x})$ be a categorical variable that can take on a discrete value k at $\mathbf{x}$ (for our example, $k = 0$ for no rain and $k = 1$ for rain). Let $\mathbf{h}_j$ be a translation vector that points from $\mathbf{x}$ to the center of one of the remaining J pixels in the template. The values of Z at the remaining pixels can be collected in a vector $\mathbf{D}(\mathbf{x}) = [Z(\mathbf{x}+\mathbf{h}_1), \ldots, Z(\mathbf{x}+\mathbf{h}_J)]^T$. The probability that the reference pixel value $Z(\mathbf{x})$ is in state k given a particular template pattern $\mathbf{d}(\mathbf{x})$ of remaining pixel values is

$$\Pr_{Z(\mathbf{x})|\mathbf{D}(\mathbf{x})}\left(z(\mathbf{x}) = k|\mathbf{d}(\mathbf{x})\right) = \frac{\Pr_{Z(\mathbf{x})\mathbf{D}(\mathbf{x})}\left(z(\mathbf{x}) = k, \mathbf{d}(\mathbf{x})\right)}{\Pr_{\mathbf{D}(\mathbf{x})}\left(\mathbf{d}(\mathbf{x})\right)}. \tag{1}$$

The multiple-point probabilities on the right-hand side of (1) are inferred from the training image as follows:

$$\Pr_{Z(\mathbf{x})\mathbf{D}(\mathbf{x})}(z(\mathbf{x}) = k, \mathbf{d}(\mathbf{x})) = c_k(\mathbf{d}(\mathbf{x}))/N_{\mathrm{TI}} \tag{2}$$

$$\Pr_{\mathbf{D}(\mathbf{x})}(\mathbf{d}(\mathbf{x})) = c^{\star}(\mathbf{d}(\mathbf{x}))/N_{\mathrm{TI}} \tag{3}$$

where N_{TI} is the total number of pixels in the training image, $c^{\star}(\mathbf{d}(\mathbf{x}))$ is the total number of occurrences of $\mathbf{d}(\mathbf{x})$ found in the training image, and $c_k(\mathbf{d}(\mathbf{x}))$ is the number of occurrences of $\mathbf{d}(\mathbf{x})$ that are associated with the event $z(\mathbf{x}) = k$ at the reference pixel. Various combinatorial algorithms are available for efficient computation of these probabilities [29], [30].

Once the pattern probabilities are known, simulation of rain support on a particular computational grid is performed sequentially. First, one has to assign a random path visiting all the grid nodes. Then, the template $T_J = [\mathbf{h}_1, \ldots, \mathbf{h}_J]^T$ is centered on a particular randomly selected node and the value of $z(\mathbf{x})$ at this node is drawn using (1). Note that in the beginning of the simulation, the probabilities will actually be unconditional since the are no conditioning events present yet. As the simulation progresses, the conditioning events are formed out of the previously simulated values of $z(\mathbf{x})$. Then, at another randomly selected node $\mathbf{x}$, the probability (1) is determined based on values of $\mathbf{D}(\mathbf{x})$ inside the template, and a realization $z(\mathbf{x})$ is drawn based on this probability. This process continues until all the grid nodes have been visited. If the multigrid approach is used, the procedure above is performed on a series of nested and increasingly finer grids $\mathcal{G}^{(l)}$ where $l = L-1, \ldots, 0$ denotes the grid aggregation level.[1] Simulation starts on the coarsest grid, and afterward, all the simulated nodes become part of the data set (conditioning events) for simulation on the next finer grid. This multigrid recursion continues for a number of grids L.

[1] $\mathcal{G}^{(l)}$ is obtained by down-sampling $\mathcal{G}^{(l-1)}$ by a factor of two in two coordinate directions, i.e., $\mathcal{G}^{(l)}$ is the subset of $\mathcal{G}^{(l-1)}$ obtained by retaining every other node of $\mathcal{G}^{(l-1)}$.

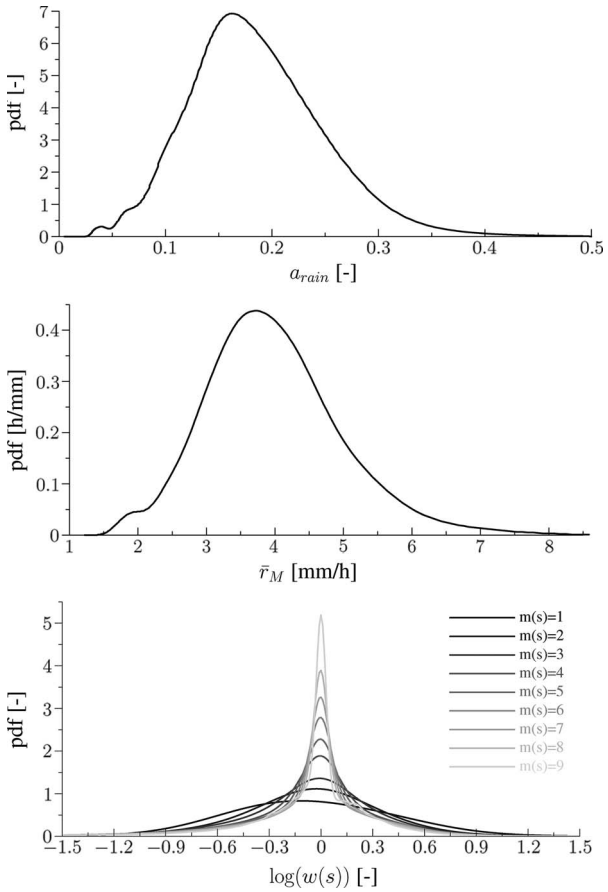


Fig. 5. Kernel representations of probability densities for: 1) (uppermost panel) fractional rainy area $A_{\rm rain}$; 2) (middle panel) mean rain rate $\bar{R}_M$ on NOWRAD rain support at 0.05° resolution; and 3) (lowermost panel) logarithm of breakdown coefficients $\log(W(s))$. The first density is used for rain support simulations using multipoint geostatistics; the second and the third densities are used for rain rate simulation with the multiscale truncated quad tree.

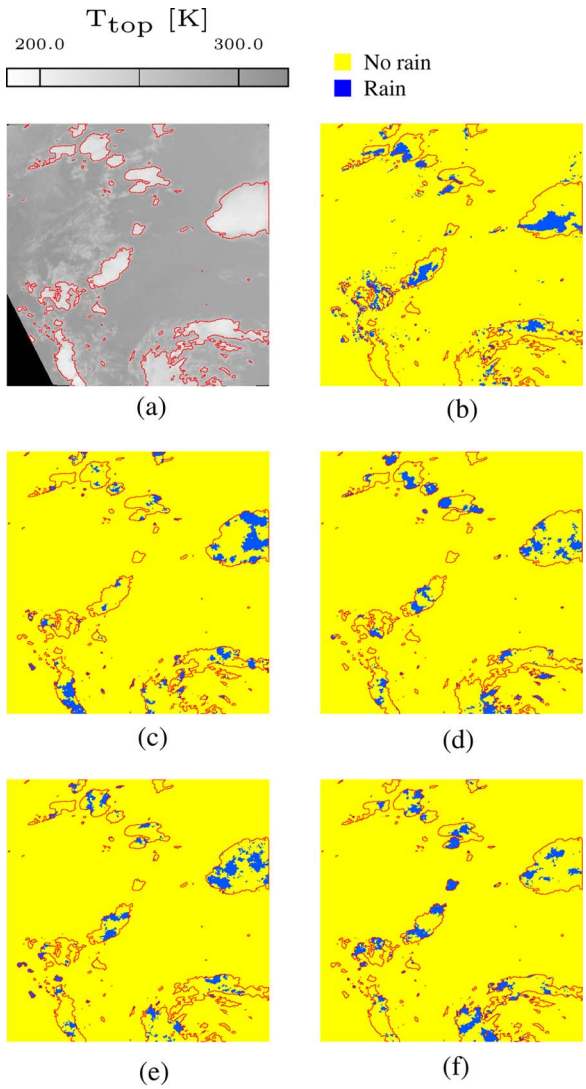


Fig. 6. (a) GOES signature of $T_{\rm top}$ on July 12, 2004, 00:00 UTC over the Central U.S. region. Solid red line delineates an irregular grid on which $T_{\rm top} <$ 258 K. (b) Corresponding NOWRAD rain support image which serves as a training image for multipoint geostatistical simulation. (c)–(f) Four examples of rain support replicates simulated on the irregular GOES grid using multipoint geostatistics and the NOWRAD training image in (b).

The template $T_J^{(l)}$ for the coarse grid $\mathcal{G}^{(l)}$ is simply a rescaled version of the finest grid's template $T_J^{(0)}$

$$T_J^{(l)} = 2^l T_J^{(0)}. \tag{4}$$

Note that the template size does not decrease with the grid aggregation level.

In the rainfall application, it is useful to further constrain the cluster support generated by the geostatistical algorithm to insure that the total support area approximates the total rainy area observed in NOWRAD images. This can be achieved by updating $\Pr_{Z(\mathbf{x})|\mathbf{D}(\mathbf{x})}(z(\mathbf{x}) = k|\mathbf{d}(\mathbf{x}))$ subject to the constraint that the marginal probability $\Pr_{Z(\mathbf{x})}(z = 1)$ is equal to a specified area fraction $A_{\rm rain}$ [31]. We treat $A_{\rm rain}$ as a random variable characterized by a probability density $f_{A_{\rm rain}}(a_{\rm rain})$ and infer its distribution from the working set of 550 NOWRAD images. The uppermost panel in Fig. 5 shows a kernel density estimate [32]–[34] of $f_{A_{\rm rain}}(a_{\rm rain})$. This density estimate uses a Gaussian kernel with Silverman's rule-of-thumb bandwidth [33].

The performance of the multipoint geostatistical algorithm is shown in Fig. 6. The upper left panel of this figure shows a typical GOES image with cloudy regions (regions with GOES brightness temperature $T_{\rm top}$ below the threshold of 258 K) outlined in red. The upper right panel shows the cluster supports identified from a NOWRAD image taken at the same time as the GOES image. The remaining four panels show the supports for four replicates generated with the geostatistical technique summarized above. The pattern probabilities used to generate these replicates were derived from the NOWRAD image in the upper right, which served as the training image. The template size was 21 × 21 pixels, and a four-level multigrid approach was used to identify the probabilities. We used these parameters because they gave results that are consistent with the training image ensemble. This is shown in Fig. 7. The left side of this figure shows one of the training images divided into rain cluster supports of different size classes (indicated by different colors). Fig. 7(b) and (c) show, respectively, the histogram of the logarithm of the cluster support size over 300 training images from the NOWRAD working set and the corresponding histogram from 300 simulated replicates. Comparison of the two histograms confirms that the procedure is producing clusters of the right size.

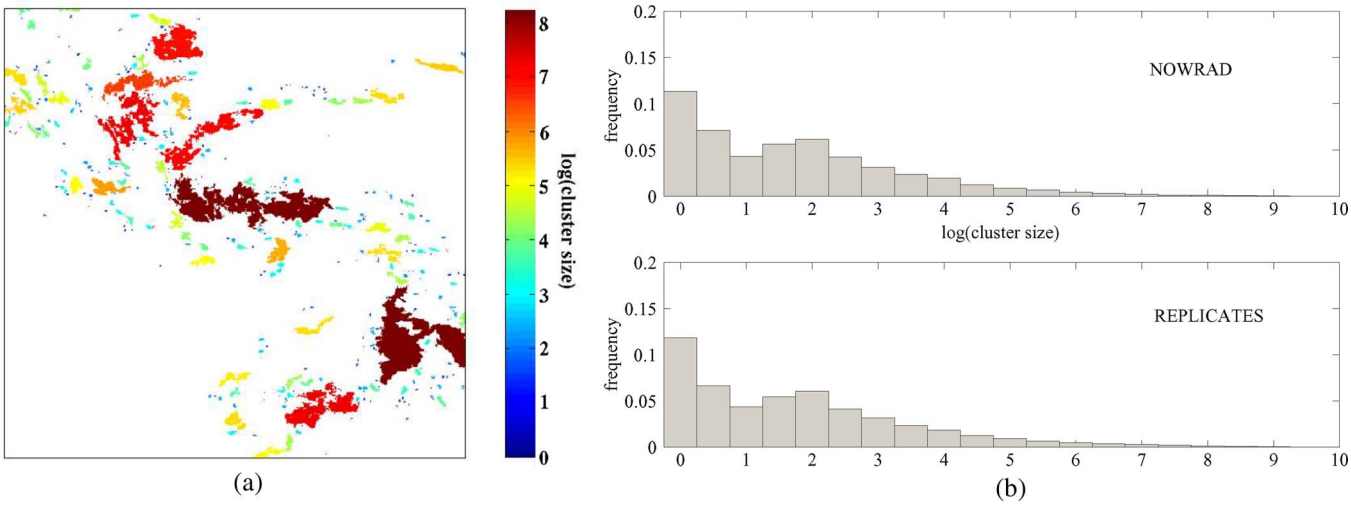


Fig. 7. (a) Example of partitioning a NOWRAD training image (July 1, 2004, 00:30 UTC) into disjoint rain support clusters of different sizes. (b) (Upper panel) Histogram of the logarithm of the cluster size for the ensemble of training images in Fig. 3(b). (Lower panel) Histogram of the logarithm of the cluster size for the simulated replicates.

The multipoint geostatistical approach is a flexible and convenient way to generate rainfall support replicates having complex shapes. Most importantly, it is able to constrain these shapes to lie within specified irregular regions such as the thresholded GOES cloud top temperature regions of interest in our application. To constrain the simulation to the estimated GOES clouds' boundaries, we first perform the simulation with a given A_{rain} over the entire study domain and then set to zero all the pixels outside GOES clouds. Due to stationarity, this is equivalent to performing geostatistical simulation on GOES cloudy regions.

C. Generation of Rain Rates Within Cluster Supports

The first two steps of our rainfall generation procedure together define the intermittent and irregular spatial support for the rain field. We still need to specify rain rates within this support. While there are many alternatives, we have found that multiscale truncated tree models provide a particularly convenient and flexible option.

To formulate a truncated tree model, we map an inverted tree onto all pixels in the specified simulation region (e.g., the rectangular Central U.S. region shown in Fig. 1). This is shown in Fig. 8 for a simple example with a single cluster contained within a 4×4 simulation region. The tree consists of nodes (rectangular areas) arranged into a series of connected scales. A particular node s is located at scale $m(s)$, with the scales indexed from $m(s) = 0$ (the coarsest or highest scale) to $m(s) = M$ (the finest or lowest scale).

Each node in the truncated quad tree used in our example is associated with up to four children $s\alpha_1$, $s\alpha_2$, $s\alpha_3$, $s\alpha_4$ (except at the finest scale) and one parent $s\gamma$ (except at the coarsest scale). Each of the nodes at the finest scale of the tree is matched to a particular rainy pixel in the cluster support. Active nodes are defined to have rainy finest scale descendants while inactive nodes do not. The active nodes are connected to their rainy descendants through a series of parent–child relationships indicated by thin lines. Inactive nodes are omitted from the tree (i.e., are truncated or pruned) and have no connections to other nodes. The example shown in Fig. 8 applies to a 4×4 simulation region containing a single rain cluster. In general, this region will contain many clusters covered by a single tree. The tree is truncated to retain only nodes with finest scale descendants in the clusters.

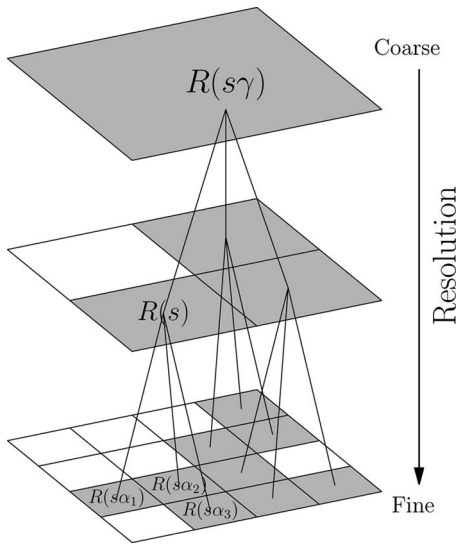


Fig. 8. Example of truncated quad tree model. Active (gray) nodes at the finest resolution correspond to rain cluster support.

Rain rates at the finest scale nodes of each tree (i.e., at pixels within each rain cluster) are generated with a multiscale recursion, starting at the coarsest scale and moving downward. The nodal rain rate (in LT^{-1}) at active tree node s is characterized by a random variable $R(s)$. In order to conserve the volume of water generated at the finest scale (in an ensemble sense), the expected value of the rain rate (taken over all replicates in the ensemble) at each tree node must equal the spatial average rain rate $\bar{R}_M$ taken over all pixels in the rain cluster (i.e., all nodes at the finest scale)

$$E[R(s)] = \bar{R}_M. \tag{5}$$

The multiscale recursion starts at the single coarsest scale node (the root node) by setting the rain rate $R(0)$ equal to $\bar{R}_M$. Since $\bar{R}_M$ is generally uncertain, it is sampled from a specified

probability distribution. The rain rates at the children in the next scale down are obtained by multiplying the root node rain rate by a nonnegative random variable $W(s)$, referred to here as a breakdown coefficient. The breakdown coefficient is independent of the root node rain rate and all other breakdown coefficients and has a specified scale-dependent probability distribution (discussed below). When repeated at each scale, the rain rate generation procedure forms a multiplicative cascade that can be represented by the following recursive equation at a particular node s

$$R(s) = R(s\gamma)W(s). \tag{6}$$

For computational reasons, it is convenient to put (6) in an equivalent additive form. This is accomplished by taking the logarithm of (6)

$$\log(R(s)) = \log\left(R(s\gamma)\right) + \log\left(W(s)\right). \tag{7}$$

The multiplicative cascade algorithm requires the probability density $f_{\bar{R}_M}(\bar{r}_M)$ of the average cluster rain rate and the breakdown coefficient densities $f_{W(s)}(w(s))$ for $s = 1, \ldots, M$. For our Central U.S. application, samples of the average cluster rain rate were obtained from the population of spatial average rain rates for the NOWRAD working set. The $f_{\bar{R}_M}(\bar{r}_M)$ kernel density estimate derived from this population is shown in the middle panel in Fig. 5.

Sample densities for the breakdown coefficients can be estimated indirectly from upscaled NOWRAD rain rates. The upscaled rates are computed with a recursion, starting with the finest scale pixel-based values obtained from NOWRAD and moving up the tree

$$\hat{R}(s) = \frac{\bar{R}_M}{N_c(s)\bar{R}_{m(s)+1}} \sum_{i=1}^{N_c(s)} \hat{R}(s\alpha_i), \qquad s = M-1, \ldots, 0 \tag{8}$$

where $\hat{R}(s)$ is the upscaled rain rate at node s, $\bar{R}_M$ is the spatial average rain rate over all finest scale nodes (i.e., over all pixels in the rain cluster), $N_c(s) \leq 4$ is the number of active children of node s, and $\bar{R}_{m(s)+1}$ is the average over $\hat{R}(s)$ at *all* active nodes at scale $m(s)+1$ (not just those that are children of s). This upscaling relationship satisfies the ensemble mass balance requirement imposed in (5). Sample estimates of the breakdown coefficients between particular scales may be computed from the ratio of the upscaled rain rates

$$\hat{W}(s) = \hat{R}(s)/\hat{R}(s\gamma). \tag{9}$$

The breakdown coefficient estimates at a given scale form a population that can be sampled by bootstrapping without replacement during the multiscale rain rate generation process. The $f_{\log(W(s))}(\log(w(s)))$ kernel density estimates for all scales used in our Central U.S. example are plotted in the lowermost panel of Fig. 5. It is clear from the figure that the variability of $\log(w(s))$ decreases as the scale gets finer.

Once the tree structure is derived from the cluster geometry and the required probability densities are specified, the simulation of rain rates is straightforward. First, the user sets the root rain rate equal to a random sample of $\bar{R}_M$ from the probability density $f_{\bar{R}_M}(\bar{r}_M)$. Then, at each remaining scale, from the coarsest to the finest scale, independent samples of $W(s)$ are drawn at random from the specified density $f_{W(s)}(w(s))$. These random samples are inserted into the recursion of (6), which gives the nodal rainfall values at the next scale down. The process is repeated over all scales until the desired rain rates are obtained at the finest scale tree nodes (i.e., at the pixels inside the rain cluster). Rainfall replicates are generated in this way at each simulation time, using a new GOES image, a new set of rain cluster supports, and a new set of multiscale trees.

IV. Analysis of Generated Replicates

The availability of extensive NOWRAD data for the Central U.S. region makes it possible for us to compare the statistical properties of simulated replicates and NOWRAD images. There are various ways to do this. One option is to compare univariate rain rate probability distributions, computed over all pixels in a given region. The resulting rain rate cumulative distribution functions (cdfs) may be interpreted as area-intensity curves, with the function value giving the fraction of total area with a rain rate less than the argument.

Assessments of spatial structure, including relationships between rain rates at different pixels, must rely on other metrics. For example, the cluster size histogram analysis shown in Fig. 7 provides a useful way to compare the sizes of observed and simulated rain cluster supports. Spatial correlation plots can also be informative. In this section, we describe several performance measures for quantifying differences between simulated and observed rainfall images and then evaluate these measures for the summer 2004 Central U.S. example.

A. Probabilistic Measures of Simulation Performance

There are a number of probabilistic measures that can be used to quantify differences between two probability distributions. Here, we use three different performance measures to compare univariate simulated and observed rain rate probability densities, indicated by $f_s(R)$ and $f_o(R)$, respectively.

Probabilistic distance measures can be conveniently interpreted if we view the ordinates of two different density functions as points in a plane. If the densities are identical, these points form a straight line with a slope of one. More generally, the points will be scattered around this line. Different distance measures use different ways to define the scatter in these points.

1) The Hellinger distance is a generalized measure of dispersion in the density ordinates

$$d_H(f_s, f_o) = \frac{1}{\sqrt{2}} \left[\int \left(\sqrt{f_s(\xi)} - \sqrt{f_o(\xi)} \right)^2 d\xi \right]^{1/2}. \tag{10}$$

The Hellinger distance is always finite and assumes values in the interval [0, 1]. In particular, $d_H = 0$ when $f_s = f_o$ and $d_H = 1$ when f_s and f_o are Dirac delta functions at different rain rates. This distance belongs to a broader class of so-called α-divergences [35] that contains many measures commonly used to compare two distributions, including the chi-squared divergence and the information-theoretic Kullback–Leibler divergence [36]. In fact, the square of (10) can be viewed as an approximation to a symmetrized Kullback–Leibler divergence (also referred to as the Jensen–Shannon divergence) and is the lower bound on the Kullback–Leibler divergence, as shown in [36].

2) The L^2 correlation distance is a generalized measure of correlation between the density ordinates

$$d_{L^2_{\rm corr}}(f_s, f_o) = 1 - \frac{\int f_s(\xi) f_o(\xi) d\xi}{\sqrt{\int f_s(\xi)^2 d\xi \int f_o(\xi)^2 d\xi}}. \quad (11)$$

The correlation distance is another measure defined over the interval [0, 1]. If f_s and f_o are similar the inner product $\int f_s(\xi)^2 d\xi \int f_o(\xi)^2 d\xi$ is larger than if f_s and f_o are dissimilar [37].

3) Zero rain rate difference. The measures given in (10) and (11) compare entire probability densities without making any distinction between values at zero versus other values. The rain rate over an arbitrary spatial region is generally a mixed density characterized by a nonzero probability of no rain and a probability of rain that is less than one [38]. That is, the rain rate probability density is a combination of a discrete atom (or Dirac delta function) at zero and a continuous density function for rain rates above zero. The difference between the atoms for two mixed densities is a useful measure of the realism of the simulated rain rate. This difference is given by

$$d_{\rm DA}(f_s, f_o) = p_s - p_o \quad (12)$$

where p_s is the probability of no rain for the simulated population and p_o is the probability of no rain for the observed population. In the context presented here, these probabilities are the fractions of total area with no rain for the two populations, taken over all simulated or observed NOWRAD images.

As mentioned in Section II, NOWRAD rain rates are quantized, with a minimum nonzero value of 1.2 mm $\cdot$ h^{-1}. The rain rates generated by our stochastic model are not quantized and may generally take values smaller than the NOWRAD threshold. In order to be able to compare NOWRAD data with simulated rain rates, we assume that both quantities are described by mixed probability density functions with atoms equal to the probability that the rain rate is smaller than or equal to 1.2 mm $\cdot$ h^{-1}. With this definition, we can use (10) and (11) to compare the continuous (positive rain rate) parts of the probability densities and (12) to compare their atoms.

In order to compute the integrals appearing in (10) and (11), we need discretized numerical approximations for f_s and f_o. The kernel density estimator technique mentioned in Section III-B accomplishes this but is computationally intensive. It is much more efficient to approximate the densities of interest with an average shifted histogram (ASH) [39]. This density estimator maintains the computational simplicity of the histogram while providing flexibility comparable to kernel density estimators. To insure that the estimators of f_s and f_o give bounded rain rates, the ASH is fit to the log-transformed data. After fitting, the log transformed ASH estimates are back-transformed to give the rain rate probability densities required in the integrals of the Hellinger and $L^2_{\rm corr}$ metrics.

B. Comparison of Generated Rainfall Ensembles With NOWRAD

When assessing the performance of a rainfall generator, it is useful to consider comparisons for individual storms as well as a larger population of similar storms. We begin by examining two typical summer storms that occurred on June 24, 2004 at 08:00 UTC and on August 19, 2004 at 06:00 UTC. Fig. 9 compares the NOWRAD image for each storm with some representative simulated replicates. All replicates for a given storm use the associated NOWRAD image for training, with rainfall constrained to occur only within the cloudy region identified from the corresponding GOES image. The root node rain rate in the multiscale tree model is set equal to the observed spatial average rain rate from the NOWRAD image. The experimental procedure for this storm-specific assessment is summarized in the pseudocode provided in Algorithm 1 of the Appendix.

Fig. 9 shows that the simulated storm images are qualitatively similar to the corresponding NOWRAD images with respect to criteria such as the typical size, shape, and density of clusters, and the magnitude of high-rainfall regions. The blockiness occasionally observed in the simulated rain storms is an artifact due to the finite number of scales and children used in the tree.

Fig. 10(a) and (b) shows a more quantitative assessment of the performance of the rainfall generator by comparing the NOWRAD and generated rain rate cdfs for each of the two storms. The red line in each figure is obtained from the quantized NOWRAD data while the continuous dark gray lines are obtained from 300 simulated replicates. In both cases, there is reasonable correspondence between the NOWRAD and simulated cdfs. The comparisons suggest that the rainfall generator slightly underestimates rain rates in the interval [1.2; 7] mm $\cdot$ h^{-1}. Possible reasons for this underestimation are the following: 1) the breakdown coefficients used in the multiscale tree are uncorrelated (correlated breakdown coefficients may give a better representation of spatial structure) and 2) there is no effort to distinguish between stratiform and convective rainfall within GOES cloud regions. The multiplicative cascade is better able to reproduce the high rain rates associated with convective storms than the low rates associated with stratiform rainfall.

The lower right panels of Fig. 10(a) and (b) compare the 75th, 90th, and 100th rainfall percentiles (p75, p90, and max, respectively) for the selected NOWRAD image with ensemble probability densities of the same percentiles for the ensemble of 300 simulated replicates. The densities are presented in the form of violin plots which include a marker for the median of the data (white dot), a black box indicating the interquartile range, and whiskers which extend to 1.5 times the interquartile range from the box as in standard box-and-whiskers plots. Overlaid on this plot is a kernel estimate of the probability density. The violin plots reveal that reproduction of high percentiles is very good. In particular, the medians of the high percentiles' distributions for simulated replicates are close to the point estimates of high percentiles extracted from NOWRAD data.

The upper left panels of Fig. 10(a) and (b) show violin plots for the three performance measures defined in the previous section. The distribution of the zero rain distance $d_{\rm DA}$ in (12) confirms the underestimation of low rain rates at the smallest NOWRAD quantization level (1.2 mm $\cdot$ h^{-1}). The narrow range of variability in the $d_{L^2{\rm corr}}$ correlation in (11) shows a very good agreement between the shapes of the NOWRAD and replicate density functions. The Hellinger distance has a slightly higher median value and broader interquartile range.

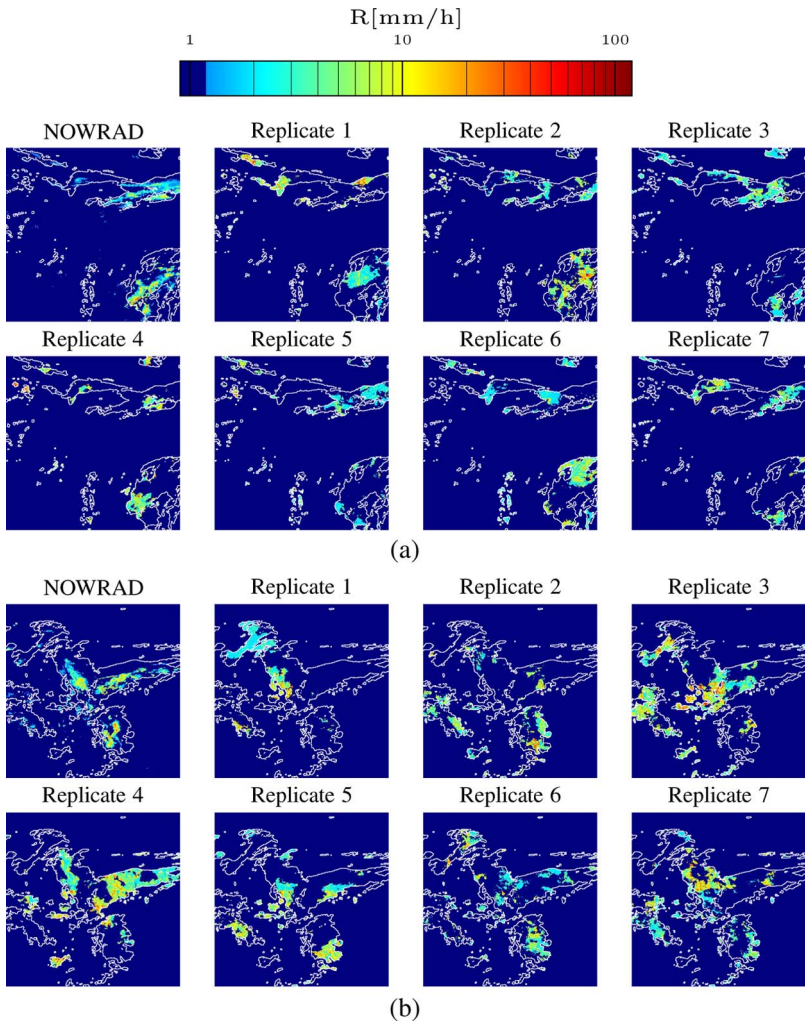


Fig. 9. (a) NOWRAD signature of a summer storm that occurred on June 24, 2004, 08:00 UTC over (upper left panel) the Central U.S. region and seven examples of simulated replicates. (b) NOWRAD signature of a summer storm that occurred on August 19, 2004, 06:00 UTC over (upper left panel) the Central U.S. region and seven examples of simulated replicates. The white contour in each image (both (a) and (b) part) delineates the region where GOES $T_{\mathrm{top}} < 258$ K. To make both simulated and NOWRAD rain support visible and in order to avoid overemphasis on very small rain rates in the replicates, the truncated logarithmic color scale was constructed such that zero rain rates are represented as dark blue. For the replicates, the rain rates below or equal to NOWRAD quantization threshold (1.2 mm $\cdot$ h^{-1}) are represented using the same (light blue) color as the rain rates in NOWRAD images equal to that threshold. The maximum of the color scale corresponds to the maximum NOWRAD rain rate (120 mm $\cdot$ h^{-1}) observed in the study period.

Comparisons of generated replicates and NOWRAD images for individual storms are useful, particularly to give a qualitative sense of the generator's ability to produce realistic looking rain replicates. However, comparisons that use storm-specific NOWRAD images for training do not provide a realistic test for situations where NOWRAD data are not available. In such cases, the generator must rely on weather radar images from other sites or times for its statistical inputs. Then, performance is better measured by comparing observed and simulated cdfs over a range of different storms, using generic rather than storm-specific training images.

This is done in Fig. 10(c), which compares cumulative distributions for 300 NOWRAD images (red) to 300 simulated replicates (dark gray). The NOWRAD images are drawn without replacement from the summer 2004 working set. One simulated replicate is generated for each storm (i.e., for each NOWRAD image). The GOES image for each storm defines the cloudy region containing rain clusters, but the statistical inputs used to generate the simulated replicate are obtained from the generic set of 30 training images shown in Fig. 3(b). The root node rain rate required by the multiscale tree model is drawn at random, for each replicate, from the NOWRAD working set. The experimental procedure for this summer-long assessment is summarized in the pseudocode provided in Algorithm 2 of the Appendix.

Fig. 10(c) shows that the simulated and NOWRAD images compare favorably. The NOWRAD percentiles considered in the violin plots in the lower right portion of the figure are now described as probability densities rather than point values. The percentile densities for the model replicates display somewhat more variability than in Fig. 10(a) and (b), reflecting the greater diversity of storms and training images used in Fig. 10(c). Fig. 10(c) shows a reasonable indication of the performance that might be expected over a large set of storms that share general features conveyed by a limited number of generic training images.

The univariate probability distribution comparisons shown in Fig. 10 show that the rainfall generator is producing the

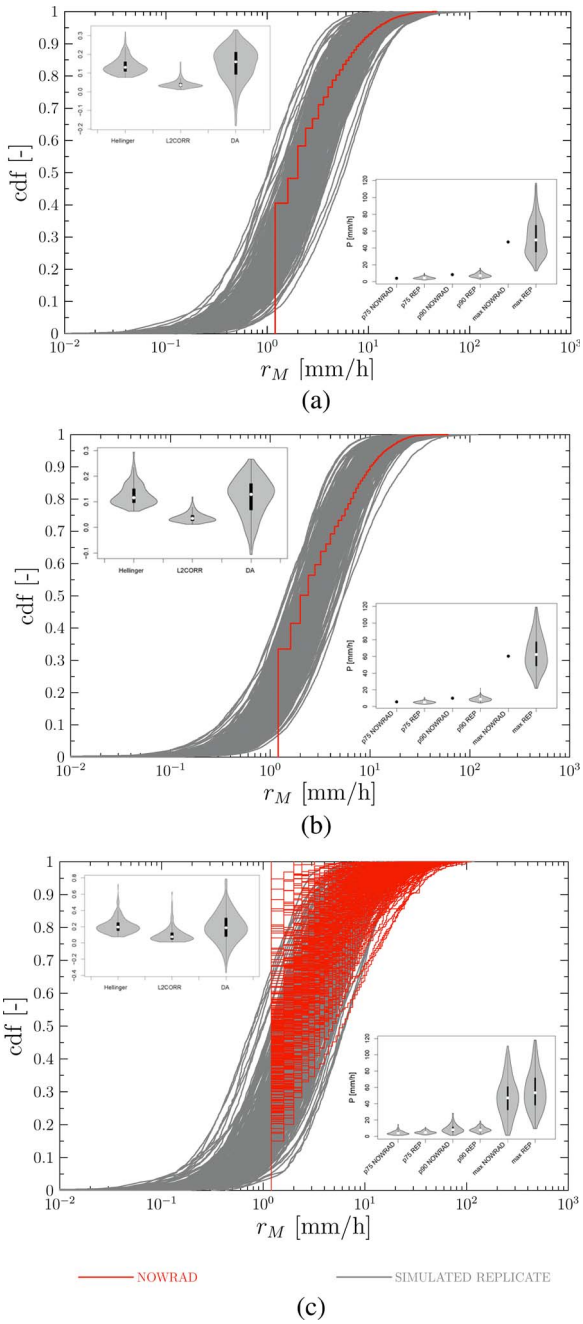


Fig. 10. CDFs for (gray lines) simulated replicates and (red lines) NOWRAD, respectively. (a) and (b) Ensemble simulation of 300 replicates with single NOWRAD signature for a summer storms that occurred on (a) June 24, 2004, 08:00 UTC and (b) August 19, 2004, 06:00 UTC. These simulations were performed using Algorithm 1. (c) Ensemble simulation of 300 replicates with ensemble of 300 NOWRAD images selected at random from the study period of summer 2004 (June 1–August 31). This simulation was performed using Algorithm 2. The upper left panel in each figure shows violin plots (see text) for performance statistics defined in (10)–(12), whereas the lower right panel shows violin plots for 75th, 90th, and 100th percentiles (p75, p90, and max, respectively) for simulated replicates (REPS) and NOWRAD, respectively.

correct range of rainfall values over the simulated clusters. The histograms shown in Fig. 7 show that the sizes of simulated rain clusters are similar to the sizes of NOWRAD clusters. In addition to these comparisons, it is useful to compare the spatial structure of simulated and NOWRAD rain rates within clusters. A commonly used quantity for describing the structure of random fields is the spatial correlation function, which quantifies

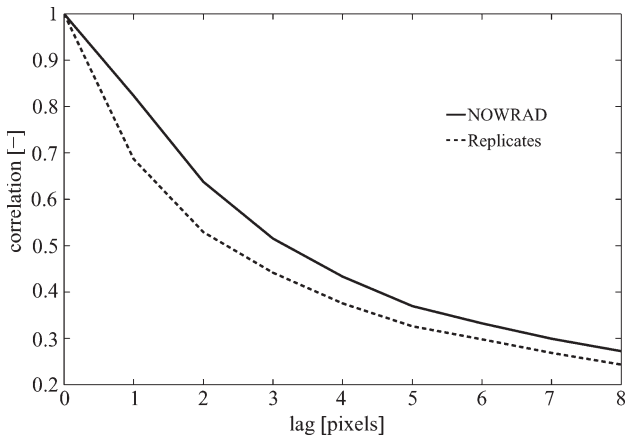


Fig. 11. Radially averaged spatial correlation function for NOWRAD and simulated replicates computed from the 300 storms considered in Fig. 10.

the tendency for values of the field at two pixels to deviate in the same direction above or below the spatial mean. If the field of interest is smooth and statistically homogeneous, the spatial correlation can be used to estimate the second moment of the bivariate probability distribution for the field values at any two pixels. Gaussian random fields are completely described by their mean and this second moment.

Caution must be taken when using spatial correlation functions to support probabilistic conclusions about intermittent nonstationary non-Gaussian fields such as the rain rate fields considered here. In such cases, spatial moments are not necessarily the same as ensemble moments. Nevertheless, it is useful to compare the spatial correlations obtained from NOWRAD and the rainfall generator, in order to get a feeling for the degree of spatial persistence exhibited in each case.

When computing spatial correlation functions for intermittent fields, it is necessary to decide on a support. Three possible alternatives are to: 1) compute correlations over the entire study area; 2) compute correlations over GOES regions (which are the same for NOWRAD and simulated images for any given storm); and 3) compute correlations only over rainy areas within GOES images. Moreover, correlations can be derived for single NOWRAD or simulated images (assuming spatial homogeneity) or they can be averaged over many replicates and/or over many storms. Comparisons are complicated by differences in simulated rain cluster supports within GOES cloudy regions. In order to keep the comparison straightforward and unambiguous, the correlation functions presented here are computed over the entire study region, including areas with no rain.

The results of the spatial correlation comparison are shown in Fig. 11, which plots correlation coefficient versus distance. For our application, the correlations were nearly isotropic, so the plots obtained in different directions were nearly identical. The correlations are computed from the 300 storms considered in Fig. 10. The NOWRAD and simulated correlation functions have similar shapes. The NOWRAD field reveals somewhat higher spatial correlations than the simulated field, suggesting that the NOWRAD rain rates have more spatial persistence, and the simulated rain rates vary more over shorter distances. Note, however, that the fact that the NOWRAD product is quantized has the impact on spatial correlation. For example, the lowest quantization level ($1.2\ \mathrm{mm} \cdot \mathrm{h}^{-1}$) determines the number of zero rainfall events in a particular NOWRAD

image. Generally, increasing the lowest quantization level would increase the number of zeros, decrease size of rain clusters, and, as a consequence, decrease the spatial correlation. Further insight about spatial structure could probably be obtained by examining conditional correlations (e.g., correlations in rainy areas or correlations within clusters of a given size). For present purposes, it is sufficient to note that the spatial correlation plots shown in Fig. 11 confirm the qualitative comparisons of Fig. 9.

V. Conclusion and Future Research

This paper presents a new procedure for generating realistic rainfall replicates conditioned on geostationary satellite (GOES) infrared measurements. Conditioning enables the procedure to reproduce the spatial and temporal intermittency observed in real rainfall events. Spatial intermittency is achieved by constraining rainfall to occur only in clusters generated within GOES cloudy regions. Temporal intermittency is a direct result of the temporal evolution of the GOES regions, which continually form, move, and dissipate over time. When taken as a whole, groups of simulated clusters inside GOES regions tend to move with the corresponding GOES images. However, the supports and rain rates of individual clusters within the GOES cloudy regions are not correlated over time, reflecting the transient nature of the convective rainfall emphasized in our summer 2004 example.

Our rainfall generation procedure relies on rainfall training images to obtain statistical information about the spatial structure of rain clusters and the average rain rates within cloudy regions. When weather radar measurements (e.g., NOWRAD) are available, they may be used to construct these images. The training images are scanned by a multipoint geostatistical procedure that computes the probabilities of all possible rainfall patterns within a specified spatial template. The procedure uses these properties to generate cluster support replicates that are statistically consistent with the training images. Replicates of rain rates within the cluster supports are obtained from a truncated multiplicative cascade model. The statistical inputs needed to construct the cascade model may be derived from weather radar data, when they are available. It is important to emphasize that weather radar provides a valuable but not essential source of input information for our rainfall generation procedure. If necessary, other rainfall data sources may be used. In this paper, uncertainties in weather radar were neglected, and the NOWRAD product was assumed to provide ground truth for verification of ensemble statistics. In reality, weather radar data are uncertain and sometimes even misleading [40]. It would be useful in future research to consider the implications of weather radar uncertainties and to incorporate quantitative descriptions of these uncertainties into the ensemble generation procedure.

Computational experiments for the Central U.S. during summer 2004 demonstrate that our procedure is able to produce spatially and temporally intermittent rainfall that is clustered in patterns comparable to those observed in NOWRAD images. The simulated rainfall support is limited to regions with GOES temperatures below a 258-K threshold. Simulated rain rates match observed NOWRAD weather radar values very well, particularly for rain rates higher than 7 mm $\cdot$ h^{-1}. This is confirmed by visual comparisons, as well as quantitative comparisons of univariate rain rate cdfs (or area intensity curves), rain cluster size distributions, and rain rate spatial correlation functions.

A number of issues raised by this paper merit further investigation. Currently, the breakdown coefficient probability densities depend on scale but not on the number of active children. It is possible that performance could be improved if distinctions were made between breakdown densities for nodes with one, two, three, and four children. Moreover, it may be useful to rely on conditional breakdown coefficient probability densities that depend on the geometrical arrangement of the active children associated with a given parent node. This conditioning could lead to better performance of our algorithm for lower rain rates.

Another useful enhancement would be to distinguish convective and stratiform rainfall based on additional information available from remote sensing sources (e.g., CAPE [41]). Such information could be incorporated into the geostatistical technique used to generate rain support so that different supports could be used for the two types of rainfall. Since stratiform events generally produce lower rain rates over larger areas, they are less spatially variable. For this reason, it is possible that stratiform rainfall does not need to be simulated with a multiscale tree but can be adequately represented with a simpler model, such as a Gauss–Markov random field. Disaggregation of breakdown coefficient probability densities combined with a distinction between convective and stratiform rainfall could potentially enhance the performance of the generation procedure at low rain rates.

It is possible that GOES-derived rapid-scan wind field vectors [42] could be incorporated into the multipoint geostatistical procedure to align simulated rain cluster supports along observed wind directions. This could give a more realistic description of rain support within GOES cloud regions. Wind information may also prove useful for enhancements that account for spatial advection of clusters that persist longer than the 1-h GOES update interval.

The rainfall generation procedure described in this paper addresses the need for realistic rainfall replicates to support ensemble prediction and data assimilation studies in hydrology, meteorology, and related disciplines. The realism offered by this procedure is largely due to its reliance on GOES imagery, which provides a frequent near-global picture of cloudy regions, and the NOWRAD weather radar product, which provides statistical inputs used by the procedure.

One of the most attractive features of our rainfall generation procedure is its ability to reproduce the general patterns of observed rainfall intensity, in both space and time. The replicates simulated by the procedure look very similar to NOWRAD images observed at the same times and locations. This suggests that the generator goes a long way toward fulfilling our objective of producing rainfall replicates that are visually and statistically indistinguishable from observations.

Appendix
Algorithm

Algorithm 1 ENSEMBLE SIMULATION WITH ENSEMBLE OF NOWRAD IMAGES

Select a NOWRAD image and the corresponding GOES image

$ii = 1$

loop over replicates
for $i = 1$ to 300 **do**
Multipoint geostatistics:
generate $a_{\mathrm{r}} \sim f_{A_{\mathrm{r}}}(a_{\mathrm{r}})$
if $\mathrm{mod}(ii, 10) = 1$ **then**
select iith training image form the training image ensemble
$ii = ii + 1$
end if
simulate rain support in GOES region where $T_{\mathrm{top}} < 258$ K
Multiscale tree:
set $\bar{r}_M = \bar{r}_{\mathrm{NOWRAD}}$
loop over scales
for $m(s) = 1$ to M **do**
coarse grain simulated rain support to scale $m(s)$
loop over the # of pixels in simulated rain support at scale $m(s)$
for $s = 1$ to # pixels **do**
$\log(w(s)) \sim f_{\log(W(s))}(\log(w(s)))$
Use (7) to obtain $\log(r(s))$
exponentiate to obtain $r(s)$
end for
end for
make sure that at scale M maximum simulated rain rate does not exceed maximum NOWRAD rain rate in the study period
while $\max(\{r(s)\}_{s=1}^{\#\,\mathrm{pixels}}) > 120\ \mathrm{mm \cdot h^{-1}}$ **do**
Multiscale tree
end while
end for

Algorithm 2 ENSEMBLE SIMULATION WITH ENSEMBLE OF NOWRAD IMAGES
$ii = 1$
loop over replicates
for $i = 1$ to 300 **do**
Select a NOWRAD image and the corresponding GOES image
Multipoint geostatistics:
generate $a_{\mathrm{r}} \sim f_{A_{\mathrm{r}}}(a_{\mathrm{r}})$
if $\mathrm{mod}(ii, 10) = 1$ **then**
select iith training image form the training image ensemble
$ii = ii + 1$
end if
simulate rain support in GOES region where $T_{\mathrm{top}} < 258$ K
Multiscale tree:
generate $\bar{r}_M \sim f_{\bar{R}_M}(\bar{r}_M)$
loop over scales
for $m(s) = 1$ to M **do**
coarse grain simulated rain support to scale $m(s)$
loop over the # of pixels in simulated rain support at scale $m(s)$
for $s = 1$ to # pixels **do**
$\log(w(s)) \sim f_{\log(W(s))}(\log(w(s)))$
Use (7) to obtain $\log(r(s))$
exponentiate to obtain $r(s)$
end for
end for
make sure that at scale M maximum simulated rain rate does not exceed maximum NOWRAD rain rate in the study period
while $\max(\{r(s)\}_{s=1}^{\#\,\mathrm{pixels}}) > 120\ \mathrm{mm \cdot h^{-1}}$ **do**
Multiscale tree
end while
end for

ACKNOWLEDGMENT

The authors would like to thank the help and insight provided by S. Friedman, V. Chatdarong, B. Jafapour, S. Lovejoy, and D. Schertzer. The NOWRAD rainfall data used in this paper were produced by WSI Corporation.

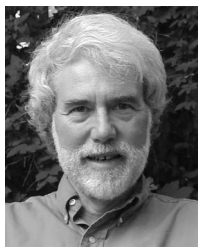
REFERENCES

[1] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast," *J. Geophys. Res.*, vol. 99, no. C5, pp. 10 143–10 162, May 1994.
[2] J. Anderson and S. Anderson, "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," *Mon. Weather Rev.*, vol. 127, no. 12, pp. 2741–2748, Dec. 1999.
[3] D. McLaughlin, Y. Zhou, and D. Entekhabi, "Computational issues for large-scale land surface data assimilation problems," *J. Hydrometeorol.*, vol. 7, no. 3, pp. 494–510, Jun. 2006.
[4] M. Pan, E. F. Wood, R. Wójcik, and M. McCabe, "Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation," *Remote Sens. Environ.*, vol. 112, no. 4, pp. 1282–1294, Apr. 2008.
[5] W. C. Skamarock, J. B. Klem, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, "A description of the advanced research WRF Version 2c," NCAR, Boulder, CO, Tech. Rep. NCAR/TN-468+STR, 2007.
[6] S. Lovejoy and D. Schertzer, "Multifractals, universality classes, satellite and radar measurements of clouds and rain," *J. Geophys. Res.*, vol. 95, no. D3, pp. 2021–2034, 1990.
[7] D. Marsan, D. Schertzer, and S. Lovejoy, "Casual space-time multifractal processes. Predictability and forecasting of rainfall fields," *J. Geophys. Res.*, vol. 26, pp. 333–346, 1996.
[8] V. Gupta and E. Waymire, "A statistical analysis of mesoscale rainfall as a random cascade," *J. Appl. Meteorol.*, vol. 32, no. 2, pp. 251–267, Feb. 1993.
[9] V. Gupta and T. Over, "A space-time theory of mesoscale rainfall using random cascades," *J. Geophys. Res.*, vol. 101, no. D21, pp. 26 319–26 331, 1997.
[10] T. Over and V. Gupta, "Statistical analysis of mesoscale rainfall: Dependence of cascade generator on mesoscale forcing," *J. Geophys. Res.*, vol. 101, pp. 319–332, 1997.
[11] I. P. Gorenburg, D. McLaughlin, and D. Entekhabi, "Scale-recursive assimilation of precipitation data," *Adv. Water Resour.*, vol. 24, no. 9/10, pp. 941–953, Nov./Dec. 2001.
[12] I. Rodriguez-Iturbe, D. Cox, and P. Eagleson, "Spatial modeling of total storm rainfall," *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. 403, no. 1824, pp. 27–50, Jan. 1986.
[13] P. Kumar and E. Foufoula-Georgiou, "A multicomponent decomposition of spatial rainfall fields: 1. Segregation of large and small scale features using wavelet transforms," *Water Resour. Res.*, vol. 29, no. 8, pp. 2515–2532, Aug. 1993.
[14] T. Crum, R. L. Alberty, and D. Burgess, "Recording, archiving and using WSR-88D data," *Bull. Amer. Meteorol. Soc.*, vol. 74, no. 4, pp. 645–653, Apr. 1993.
[15] M. Montopoli, F. Marzano, G. Vulpiani, M. Anagnostou, and E. Anagnostou, "Statistical characterization and modeling of raindrop spectra time series for different climatological regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2778–2787, Oct. 2008.

[16] F. Marzano and J. Weinman, "Inversion of spaceborne X-Band synthetic aperture radar measurements for precipitation remote sensing over land," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3472–3487, Nov. 2008.
[17] D. Sarma, M. Konwar, S. Sharma, S. Pal, J. Das, U. De, and G. Viswanathan, "An artificial-neural-network-based integrated regional model for rain retrieval over land and ocean," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1689–1696, Jun. 2008.
[18] R. Houze and A. Betts, "Convection in GATE," *Rev. Geophys. Space Phys.*, vol. 19, pp. 541–576, 1981.
[19] J. Mecikalski, K. Bedka, S. Paech, and L. Litten, "A statistical evaluation of GOES cloud-top properties for nowcasting convective initiation," *Mon. Weather Rev.*, vol. 136, no. 12, pp. 4899–4914, Dec. 2008. DOI: 10.1175/2008MWR2352.12.
[20] T. Crum, "An update on WSR-88D level II data availability," *Bull. Amer. Meteorol. Soc.*, vol. 76, pp. 2485–2487, 1995.
[21] C. Grassotti, R. Hoffman, E. R. Vivoni, and D. Entekhabi, "Multiple-timescale intercomparison of two radar products and rain gauge observations over the Arkansas Red river basin," *Weather Forecast.*, vol. 18, no. 6, pp. 1207–1229, Dec. 2003.
[22] R. E. Carbone, J. D. Tuttle, D. A. Ahijevych, and S. B. Trier, "Inferences of predictability associated with warm season precipitation episodes," *J. Atmos. Sci.*, vol. 59, no. 13, pp. 2033–2056, Jul. 2002.
[23] F. Guardiano and R. Srivastava, "Multivariate geostatistics: Beyond bivariate moments," in *Geostatistics Troia*, vol. 1, A. Soares, Ed. Dordrecht, The Netherlands: Kluwer, 1993, pp. 133–144.
[24] S. Strabelle and N. Remy, "Post-processing of multiple-point geostatistical models," in *Geostatistics Banff 2004*, vol. 2, O. Leuangthong and C. Deutsch, Eds. Dordrecht, The Netherlands: Springer-Verlag, 2004, pp. 979–988.
[25] Y. Shusse and K. Tsuboki, "Dimension characteristics and precipitation efficiency of cumulonimbus clouds in the region far south from the Mei-Yu front over the eastern Asian continent," *Mon. Weather Rev.*, vol. 134, no. 7, pp. 1942–1953, Jul. 2006.
[26] A. Dennis, A. Koscielski, D. E. Cain, J. H. Hirsch, and P. L. Smith, "Analysis of radar observations of a randomized cloud seeding experiment," *J. Appl. Meteorol.*, vol. 14, no. 5, pp. 897–908, Aug. 1975.
[27] R. Lopez, "Internal structure and development processes of C-scale aggregates of cumulus clouds," *Mon. Weather Rev.*, vol. 106, no. 10, pp. 1488–1494, Oct. 1978.
[28] A. Gagin, D. Rosenfeld, and R. E. Lopez, "The relationship between height and precipitation characteristics of summertime convective cells in south Florida," *J. Atmos. Sci.*, vol. 42, no. 1, pp. 84–94, Jan. 1985.
[29] S. Strabelle, "Conditional simulation of complex geological structures using multiple point statistics," *Math. Geol.*, vol. 34, no. 1, pp. 1–21, Jan. 2002.
[30] T. Tran, "Improving variogram reproduction on dense simulation grids," *Comput. Geosci.*, vol. 20, no. 7/8, pp. 1161–1168, Aug.–Oct. 1994.
[31] N. Remy, *Geostatistical Earth Modelling Software—User's Manual*, 2004. [Online]. Available: http://sgems.sourceforge.net/doc/sgems_manual.pdf
[32] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, 1962.
[33] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Chapman & Hall, 1986.
[34] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, U.K.: Chapman & Hall, 1995.
[35] S. Ali and S. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. R. Stat. Soc.*, vol. 28, pp. 131–143, 2000.
[36] F. Topsoe, "Some inequalities for information divergence and related measures of discrimination," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1602–1609, Jul. 2000.
[37] D. Scott and W. Szewczyk, "From kernels to mixtures," *Technometrics*, vol. 43, no. 3, pp. 323–335, Aug. 2001.
[38] B. Kedem and H. Pavlopoulos, "On the threshold method for rainfall estimation: Choosing the optimal threshold level," *J. Amer. Stat. Assoc.*, vol. 86, no. 415, pp. 626–633, 1991.
[39] D. Scott, "Averaged shifted histogram: Effective nonparametric density estimator in several dimensions," *Ann. Stat.*, vol. 13, no. 3, pp. 1024–1040, 1985.
[40] B. R. Nelson, W. F. Krajewski, A. Kruger, J. A. Smith, and M. L. Baeck, "Archival precipitation data set for the Mississippi River Basin: Algorithm development," *J. Geophys. Res.*, vol. 108, no. D22, p. 8857, 2003. DOI:10.1029/2002JD003158.
[41] T. Schmit, G. Wade, W. Feltz, J. Jung, J. P. Nelson, III, W. P. Menzel, A. P. Noel, and J. N. Heil, "Validation and use of GOES sounder moisture information," *Weather Forecast.*, vol. 17, no. 1, pp. 139–154, Feb. 2002.
[42] C. Velden, J. Daniels, D. Stettner, D. Santek, J. Key, J. Dunion, K. Holmlund, G. Dengel, W. Bresky, and P. Menzel, "Recent innovations in deriving tropospheric winds from meteorological satellites," *Bull Amer. Meteorol. Soc.*, vol. 86, no. 2, pp. 205–223, Feb. 2005.

Rafał Wójcik received the Ph.D. degree from Warsaw Agricultural University, Warsaw, Poland, in 2000.

From 2000 to 2002, he was a Researcher with the Climate Analysis Department, Royal Netherlands Meteorological Institute, De Bilt, The Netherlands. From 2002 to 2003, he was with the Quantitative Water Management Group, Department of Environmental Science, Wageningen University, The Netherlands. From 2004 to 2006, he was a Member of Research Staff with the School of Engineering and Applied Science, Princeton University, Princeton, NJ. He is currently a Postdoctoral Research Associate with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge. His research interests include geophysical data assimilation methods, remote sensing of the environment, nonlinear dynamics and stochastic processes, extreme value theory, geostatistics, machine learning, range predictability and risk analysis of natural hazards, probabilistic climate change scenarios, visualization of high-dimensional data, and applications of nonparametric probability density estimators.



Dennis McLaughlin received the Ph.D. degree from Princeton University, Princeton, NJ, in 1985.

He is currently the H. M. King Bhumibol Professor of Water Resources Management with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge. His research interests include environmental data assimilation, real-time control, and management of water resources in semiarid regions.

Dr. McLaughlin is a Fellow of the American Geophysical Union and a member of the American Meteorological Society.

Alexandra G. Konings is currently working toward the B.S. degree in environmental engineering at the Massachusetts Institute of Technology, Cambridge.

In the summer of 2008, she was with the Jet Propulsion Laboratory, Pasadena, CA, working on science support for the Soil Moisture Active Passive mission. Her research interests include hydrometeorology and data assimilation.

Dara Entekhabi (M'04–SM'04) received the B.A. and two M.A. degrees in geography from Clark University, Worcester, MA, in 1983, 1984, and 1987, respectively, and the Ph.D. degree in civil engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1990.

He is currently a Professor with the Department of Civil and Environmental Engineering and the Department of Earth, Atmospheric, and Planetary Sciences, MIT. He is the Leader of the Science Definition Team for NASA's Soil Moisture Active-Passive mission. He is the Director of the Ralph M. Parsons Laboratory. His research activities are in terrestrial remote sensing, data assimilation, and coupled land–atmosphere systems behavior.

Prof. Entekhabi is a Fellow of the American Meteorological Society and the American Geophysical Union.