AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR016935

Special Section:
The 50th Anniversary of Water Resources Research

Point rainfall statistics for ecohydrological analyses derived from satellite integrated rainfall measurements

Manuel del Jesus[1], Andrea Rinaldo[2,3], and Ignacio Rodríguez-Iturbe[4]

[1]Hydrology group, Universidad de Cantabria, Cantabria, Spain, [2]Laboratory of Ecohydrology, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [3]DICEA, University of Padova, Padova, Italy, [4]Environmental Engineering and Water Resources, Princeton University, Princeton, New Jersey, USA

Key Points:
- A satellite to point rainfall downscaling framework is presented
- Point rainfall statistics are predicted from satellite measurements
- The framework is tested with rain gage data from the Kalahari transect

Correspondence to:
I. Rodríguez-Iturbe,
irodrigu@princeton.edu

Citation:
del Jesus, M., A. Rinaldo, and I. Rodríguez-Iturbe (2015), Point rainfall statistics for ecohydrological analyses derived from satellite integrated rainfall measurements, *Water Resour. Res.*, *51*, 2974–2985, doi:10.1002/2015WR016935.

Received 15 JAN 2015
Accepted 29 MAR 2015
Accepted article online 3 APR 2015
Published online 30 APR 2015

Abstract Satellite rainfall measurements, nowadays commonly available, provide valuable information about the spatial structure of rainfall. In areas with low-density rain gage networks, or where these networks are nonexistent, satellite rainfall measurements can also provide useful estimates to be used as virtual rain gages. However, satellite and rain gage measurements are statistically different in nature and cannot be directly compared to one another. In the present paper, we develop a methodology to downscale satellite rainfall measurements to generate rain-gage-equivalent statistics. We apply the methodology to four locations along a strong rainfall gradient in the Kalahari transect, southern Africa, to validate the methodology. We show that the method allows the estimation of point rainfall statistics where only satellite measurements exist. Point rainfall statistics are key descriptors for ecohydrologic studies linking the response of vegetation to rainfall dynamics.

1. Introduction

Mathematical models of rainfall are idealized representations of the rainfall process that in hydrologic applications may serve different purposes. On the one hand, such models serve to mathematically reproduce the main statistical characteristics of the rainfall process by means of a reduced set of parameters. On the other hand, the models may provide a conceptual framework to explain rainfall as a physical process, simplifying the understanding and interrelation of the underlying dynamics that produce rain.

Many hydrological applications make use of the statistical structure of rainfall dynamics at a point. In ecohydrologic studies, such point rainfall characterization is needed for studying the response of vegetation to different climatic conditions [*Rodríguez-Iturbe and Porporato*, 2004; *Herrera et al*., 2001; *Wright et al*., 2012], but is also important in other types of dynamics like the prediction of waterborne disease reinfections, as rainfall (only available as remotely sensed data in most underdeveloped countries or in times of emergency) prompts washout of open-air defecation sites charging water reservoirs with pathogens [*Rinaldo et al*., 2012].

Rainfall characterization at a point is readily derived from rain gage data where this information is available. However, there are many regions in the world where rain gage data are scarce and thus a proper point rainfall characterization may be elusive.

In this paper, we will deal with point rainfall characterization derived from satellite observations. Satellite imagery is available with global coverage, so it may be used to characterize rainfall in remote, ungaged locations. Although we recognize the importance of bias and errors in satellite measurements, we will not deal with bias correction in this work, as our main objective is the downscaling of satellite data in regions where no rain gage data exist. For regions where rain gage information exists, *Müller and Thompson* [2013] present a methodology for bias correction using stochastic rainfall modeling. They also develop an upscaling methodology for the above situation. However, the pioneering work of *Müller and Thompson* [2013] cannot be applied in the absence of rain gage information and is thus outside the scopes of this work.

Satellite measurements are not direct measures of the point process of interest (rainfall) but result from the integration of the rainfall process throughout the observation window (given by the satellite spatial resolution). For this reason, satellite-derived rainfall statistics cannot be directly translated into point rainfall

statistics, and thus, satellite measurements cannot be directly used to derive the parameters of rainfall point models, which need to be implemented in the type of studies mentioned before.

In ecohydrologic studies, one needs to consider that vegetation responds to the local characteristics of rainfall and not to the integrated ones over areas of hundreds of square kilometers. In a statistically homogeneous region, i.e., a region small enough for synoptic variations of rainfall to be negligible and without strong topographical features, the mean value of the point rainfall and the integrated rainfall is obviously the same. Nonetheless, important second-order properties like variance and autocorrelation may easily differ by more than 50% for pixel areas of $20\times20\,km^2$. Moreover, other properties of key importance to study vegetation response to climate, such as the percentage of dry time, may be greatly different at the point level and at the pixel level.

This paper presents a framework to downscale satellite-derived rainfall statistics to rain-gage-equivalent statistics. A spatiotemporal rainfall point model [*Cox and Isham*, 1988] is used as the basis for the downscaling process. This model was used in previous studies [*Isham et al*., 2005; *Rodríguez-Iturbe et al*., 2006] to characterize rainfall over a savanna-like ecosystem in southern Italy. The equations relating the properties of the point rainfall and the integrated process (equivalent to the satellite observations) are derived in section 2. These equations are then used to estimate the parameters of the spatiotemporal rainfall model that reproduce, when properly aggregated, the satellite statistics. In section 3, the parameter fitting procedure is cast in the form of an optimization problem where the total root squared error of the model-derived satellite statistics is minimized. Particle Swarm Optimization (PSO) [*Kennedy*, 2010] is used to solve the optimization problem. The parameters of the spatiotemporal model are then linked to the rainfall dynamics at a point.

Once the mathematical framework has been developed, section 4 of the paper focuses on a practical application to the Kalahari transect, South Africa [*Shugart et al*., 2004], where only a very sparse rain gage network exists. The Kalahari transect was selected because of its lack of strong orographic effects which results in the appropriate conditions to assume rainfall statistical homogeneity at the satellite pixel scale. We apply the methodology to compute statistics of point rainfall within this region starting from processed satellite observations and then continue to compare the results with existing rain gage data. Finally, section 5 presents the conclusions of the study.

2. Model Description

A continuous spatiotemporal rainfall model [*Cox and Isham*, 1988] is used to characterize the structure of rainfall, which is assumed to be statistically homogeneous over extensive regions. The model is most appropriate for regions dominated by convective precipitations, although it can also capture the statistical characteristics of frontal precipitation. The mathematical scheme attempts to represents the rainfall structure at any point within this homogeneous region, incorporating the space-time correlation structure of the rainfall within the region.

Satellite observations will be interpreted as integrated measurements of the spatiotemporal rainfall model. The integration of the satellite observations is both in space and time, with the integration domain defined by the satellite spatial and temporal resolutions. The statistics of satellite observations can be related to the statistical structure of the integrated spatiotemporal model, and thus to the parameters of the latter. Statistics of point rainfall, e.g., rain gage, can then be estimated based on rainfall statistics derived from satellite imagery.

2.1. Spatiotemporal Rainfall Model

Rainfall is assumed to be generated by rain cells, which appear in the domain following a Poisson distribution in space and time of rate λ (cells/km^2/d). Rain cells are assumed to be circular in shape and static, i.e., they cover the same region from the moment they are born until they fade away. Every rain cell is characterized by means of three mutually independent random variables: (1) cell radius, (2) cell duration, and (3) cell rainfall intensity. Cell rainfall intensity is assumed to be constant throughout the duration of the rain cell. These three random variables are assumed to follow exponential distributions with parameters ρ (1/km), η (1/d), and β (d/mm), respectively. These parameters are the inverse of the mean values of their distributions.

2.2. Point Rainfall Characteristics

A brief description of the most useful model properties is presented here. For an in-depth derivation of these properties please refer to *Cox and Isham* [1988] and *Isham et al*. [2005]. We denote rainfall intensity at position **x** and time *t* by $r(\mathbf{x}, t)$. From *Isham et al*. [2005], the mean of $r(\mathbf{x}, t)$ is given by

$$E[r(\mathbf{x}, t)] = \mu_r = \frac{\lambda'}{\eta\beta} \tag{1}$$

where $\lambda' = 2\pi\lambda/\rho^2$ is the rate at which rain cells arrive to an arbitrary point in space. The variance of rainfall intensity at a point is given by:

$$\sigma_r^2 = \frac{2\lambda'}{\eta\beta^2} \tag{2}$$

The correlation, Υ, between rainfall intensity at location $\mathbf{x_A}$ at time t, and at location $\mathbf{x_B}$ at time $t+h$ is approximately given by [*Isham et al*., 2005]:

$$\Upsilon[r(\mathbf{x_A}, t), r(\mathbf{x_B}, t+h)] = \left(1 + \frac{\rho d}{4}\right) e^{-\rho d/2} e^{-\eta h} \tag{3}$$

where d is the distance between locations $\mathbf{x_A}$ and $\mathbf{x_B}$ and h is the time lag between both variables.

2.3. Aggregated Rainfall in Space and Time

Satellite measurements are average values of the variable of interest over fixed areas (pixels) and time intervals. The aggregation domain is fixed by the satellite spatial and temporal resolutions. Thus, satellite rainfall measurements can be interpreted as the average, during the satellite measurement interval, of the rainfall occurring within a pixel. Mathematically this can be expressed as:

$$R_D(x, y, t) = \frac{1}{\Delta x \Delta y \Delta t} \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \int_{y-\frac{\Delta y}{2}}^{y+\frac{\Delta y}{2}} \int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} r(x, y, t)\, dx\, dy\, dt \tag{4}$$

where $r(x, y, t)$ is the rainfall intensity of the spatiotemporal process at location (x, y) at time t, $R_D(x, y, t)$ is the rainfall intensity of the integrated process over domain D centered at the same location and time, and Δx, Δy, and Δt are the aggregation intervals, whose cartesian product defines the aggregation domain $D = \Delta x \times \Delta y \times \Delta t$. Δx and Δy are the sides of the satellite measurement pixel (its spatial resolution) and Δt is the length of the averaging period (temporal resolution).

Assuming that rainfall is a statistically homogeneous process over the measurement pixel, i.e., that the parameters of the underlying point process are constant over the pixel, the properties of the point process can be derived by using satellite measurements by means of the relation established in equation (4). To this end, the general results presented in *Vanmarcke* [2010] will now be used.

The mean values of the point and the integrated processes are equal. Thus, equation (1) applies also to the integrated process. However, second (and higher)-order properties of the point process and the integrated processes are quite different. They are related by the variance function, which measures the reduction of the point variance (σ^2) under the averaging process. The magnitude of this reduction is a function of the size and shape of the averaging domain.

The relation between the variances of the point and the integrated processes is expressed mathematically as follows:

$$\sigma_{X_\Omega}^2 = \gamma(\mathbf{\Omega})\, \sigma^2 \tag{5}$$

where $\sigma_{X_\Omega}^2$ is the variance of the integrated process, σ^2 is the variance of the point process, $\gamma(\mathbf{\Omega})$ is the variance function, and $\mathbf{\Omega}$ is the vector defining the averaging domain.

The variance function is related to the correlation function of the process [*Vanmarcke*, 2010, equations (5.1.6) and (6.1.6)]. In the one-dimensional case, this relation can be expressed as follows:

$$\gamma(\Omega)=\frac{1}{\Omega^2}\int_0^{\Omega}\int_0^{\Omega}\Upsilon(t_1-t_2)\,dt_1\,dt_2 \tag{6}$$

where $\gamma(\Omega)$ is the variance function, Υ is the correlation function, and Ω is the size of the averaging domain, which in the one-dimensional case is simply a distance.

Note that sometimes it is convenient to work with the variance function of the local integral process, $\Delta(\Omega)$, rather than with the local average one, $\gamma(\Omega)$. The local integral process is generated by the integral of the point process over the aggregation domain, rather than by its average. It is directly obtained from the integral in equation (4), without dividing by the domain size. Both variance functions are related by the following relation:

$$\Delta(\Omega)=\Omega^2\,\gamma(\Omega) \tag{7}$$

Equation (6) implies that in the one-dimensional case in order to compute the variance function a double integral is needed. This pattern is kept for higher-dimensional cases, which implies that in our case a sextuple integral is required. However, by means of a change of variables $\tau = t_1 - t_2$, integrating over t_1 and realizing that $\Upsilon(\tau)$ is an even function, equation (6) can be rewritten as follows:

$$\gamma(\Omega)=\frac{2}{\Omega}\int_0^{\Omega}\left(1-\frac{\tau}{\Omega}\right)\Upsilon(\tau)\,d\tau \tag{8}$$

where the double integral has been simplified into a single one.

When the limits of integration of the double integral (or the sextuple integral in the three-dimensional case) are not equal, for instance when computing covariances, a domain decomposition must be carried out prior to the change of variable [*Vanmarcke*, 2010]. The domain decomposition increases the number of integrations, but the overall computation complexity is reduced as the dimensionality of the new integrals can be reduced. These manipulation simplify our sextuple integral into a sum of triple ones.

For the three-dimensional case, the covariance can be computed by using the following expression [*Vanmarcke*, 2010, p. 309]:

$$Cov[R_D,R_{D'}]=\frac{\sigma^2}{2^3DD'}\sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3}(-1)^i(-1)^j(-1)^k\Delta(\mathbf{\Omega_{1i}},\mathbf{\Omega_{2j}},\mathbf{\Omega_{3k}}) \tag{9}$$

where σ^2 is the variance of the point process, D and D' are the aggregation domains in space and time of R_D and $R_{D'}$, respectively, which in our case are equal, and $\Delta(\mathbf{\Omega_{1i}},\mathbf{\Omega_{2j}},\mathbf{\Omega_{3k}})$ is the variance function related to the three-dimensional vector $[\mathbf{\Omega_{1i}},\mathbf{\Omega_{2j}},\mathbf{\Omega_{3k}}]$ extracted from matrix $\mathbf{\Psi}$, described below, which defines the aggregation domains and their relative positions.

Equation (9) involves 64 addends (three summations, indexed by i, j, and k, with four elements per summation, as described below), every one with a potentially different value for the three-dimensional vector $[\mathbf{\Omega_{1i}},\mathbf{\Omega_{2j}},\mathbf{\Omega_{3k}}]$, as the indices of the summations are linked to the subindices defining the three-dimensional vector. For a particular term of the summation, characterized by three values of the indices (i_o, j_o, k_o), the vector $[\mathbf{\Omega_{1i_o}},\mathbf{\Omega_{2j_o}},\mathbf{\Omega_{3k_o}}]$ is constructed by taking from matrix Ψ the i_o column from the first row, the j_o column from the second row, and the k_o column from the third row. The 64 potentially different $[\mathbf{\Omega_{1i}},\mathbf{\Omega_{2j}},\mathbf{\Omega_{3k}}]$ vectors are constructed in this way.

In the particular case of this study where integration domains are of equal size and shape, nonoverlapping and located in a square grid, the matrix $\mathbf{\Psi}$ containing the complete information about averaging size and relative position is:

$$\mathbf{\Psi}=\begin{bmatrix}(I-1)L, & L, & (I+1)L, & L\\ (m-1)L, & L, & (m+1)L, & L\\ (n-1)T, & T, & (n+1)T, & T\end{bmatrix} \tag{10}$$

where I indicates the distance (in pixel-side lengths) along the X axis, m indicates the distance (in pixel-side lengths) along the Y axis, n indicates the distance (in satellite time resolution sizes) along the time axis, L is the length of a pixel side, and T is the time resolution of the satellite measurements.

To compute equation (9), σ^2 is substituted by its value given by equation (2), D and D' are substituted by their values, both equal to the volume of the averaging domain, L^2T, and the variance function of the local integral process, $\Delta(\mathbf{T_{1i}}, \mathbf{T_{2j}}, \mathbf{T_{3k}})$, is computed using the following expression [*Vanmarcke*, 2010]:

$$\Delta(\mathbf{\Omega_{1i}}, \mathbf{\Omega_{2j}}, \mathbf{\Omega_{3k}}) = 8 \int_0^{\Omega_{1i}} \int_0^{\Omega_{2j}} \int_0^{\Omega_{3k}} (\Omega_{1i} - |\tau_1|)(\Omega_{2j} - |\tau_2|)(\Omega_{3k} - |\tau_3|) \Upsilon(d, h) \mathrm{d}\tau_1 \, \mathrm{d}\tau_2 \, \mathrm{d}\tau_3 \tag{11}$$

where $\Upsilon(d, h)$ is the correlation function of the point process given in equation (3), with $d = \sqrt{\tau_1^2 + \tau_2^2}$ and $h = \tau_3$.

The variance of the pixel-integrated process is computed by fixing $l = 0$, $m = 0$, $n = 0$. The lag-1, time covariance is obtained by setting $l = 0$, $m = 0$, $n = 1$. Last, the lag-1 spatial covariance is obtained with $l = 1$, $m = 0$, $n = 0$ or $l = 0$, $m = 1$, $n = 0$, assuming, as it is the case here, that the process is homogeneous and isotropic in space.

2.4. Properties of the Integrated Model

As mentioned before, the process at a point and the integrated process have the same mean value. However, second-order statistics such as the variance, spatial covariance, and temporal covariance are different. They are related by equation (9).

A statistic of great ecohydrologic interest is the probability of a point (local scale) being dry during a fixed time period. Competition among different species of plants, as well as vegetation water stress, are closely related to this feature. The probability of a pixel being dry during a fixed period of time may be estimated directly from satellite information and will now be linked to the same probability at a point by using the spatiotemporal stochastic rainfall model described earlier in this section.

The probability of a pixel being dry during a given period of time is the product of: (1) the probability that no rain occurs in the pixel during the specified time interval (P_1) and (2) the probability of the pixel being initially dry, which equals the probability of the pixel being dry at any randomly chosen time (P_2).

We first calculate the former. No rain occurs in a pixel during an interval of time (T) if two conditions are met: (1) no cell center is born within the pixel during the interval (probability P_{11}) and (2) no cell center occurring outside the pixel possesses a cell radius larger than the distance from the cell center to the pixel (probability P_{12}).

The probability of no storm center occurring within the pixel during the interval T is easily computed. The number of cell centers occurring within a pixel of area A during a time interval T follows a Poisson distribution [*Cox and Isham*, 1988] of parameter $\lambda'' = \lambda \cdot A \cdot T$. Therefore, the probability that a cell center is born within the pixel during the interval T equals:

$$P_{11} = \exp(-\lambda A T) \tag{12}$$

The calculation of the probability that no rain occurs within the pixel from cell centers located outside the pixel requires that we assume a circular pixel to simplify computations. The real pixel (a square) and the idealized one will have the same area, and thus the square length (L) and the circle radius (R) will verify $L = \sqrt{\pi} \cdot R$.

Consider now an annulus of infinitesimal area located at a distance $R + r$ from the center of the pixel (see Figure 1). The number of cell centers born in the annulus during an interval of duration T follows a Poisson distribution with parameter, $\Lambda = 2\pi\lambda T(R+r)\mathrm{d}r$. Note that the parameter of the Poisson distribution increases as the annulus is located further away from the pixel because the area of the annulus increases linearly with r.

Multiplying Λ by the probability that a cell radius is larger than a specified threshold, r_c, given by $\exp(-\rho r_c)$, one obtains the parameter of the Poisson process of cell centers being born in the annulus during the interval T whose radii are larger than r_c, $\lambda_{sr>r_c}$:

$$\lambda_{sr>r_c} = 2\pi\lambda T(R+r) \exp(-\rho r_c) \, \mathrm{d}r \tag{13}$$

We now need to specify a threshold for the radius of a cell below which the cell does not contribute to rain inside the pixel. In fact, detecting rainy days via satellite measurements involves the use of thresholds below

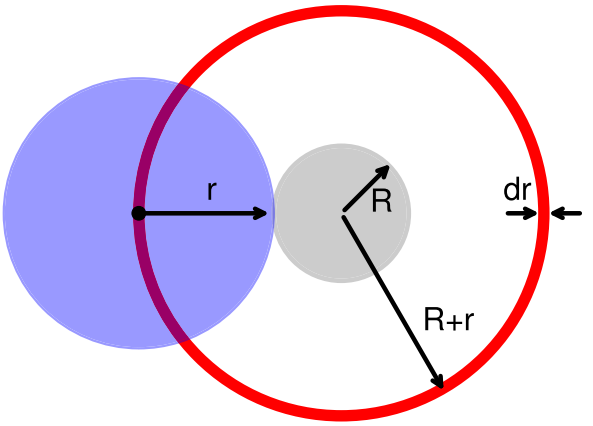

Figure 1. Sketch to compute the contribution to the probability of rain coming from outside the pixel of an annulus located at a distance $R + r$. The gray circle is the pixel of radius R under consideration. The red annulus is one infinitesimal area from which rainfall may be coming. The pale blue circle is a storm whose center is within the annulus. If the radius is smaller than r, its rainfall will not affect the pixel.

which rainfall is assumed not to occur. Although it may be raining, intensity may be so low that the satellite cannot distinguish it from background noise. When rain cells occur within the pixel it is likely that satellites will capture the event. However, when rainfall is coming from cell centers located outside the pixel, successful identification of the event will depend on the amount of rainfall produced inside the pixel which in our mathematical representation is related to the area of overlap between the storm cell and the pixel. This area of overlap can be approximated by

$$A_o = \pi R^2 \frac{r_\alpha - r}{2R} \text{ for } r \leq r_\alpha \leq r + 2R \tag{14}$$

where A_o is the area of overlap between the storm cell and the pixel, R is the radius of the pixel, r_α is the radius of the cell, and r is the minimum distance from the cell center to the edge of the pixel (see Figure 1). The threshold area of overlap, A_{crit}, is obtained when r_α is substituted by r_c.

Normalizing the area of overlap by the area of the pixel, the threshold percentage of overlap, α_c, is obtained. Rearranging terms in equation (14), the threshold cell radius required to ensure α_c is given by:

$$r_c = r + 2R\alpha_c \tag{15}$$

Using this as the critical radius in equation (13), the parameter of the Poisson process of cell centers being born during an interval T in the annulus of width dr located at a distance $R + r$ from the center of the pixel which contribute to rainfall inside the pixel can be written as:

$$\lambda_{sr>r_c} = 2\pi\lambda T(R+r)\exp\left(-\rho(r+2R\alpha_c)\right) dr \tag{16}$$

The parameter of the Poisson process of the number of cell centers producing rain within the pixel from its outside is, thus, the sum of the contributions from every annulus, that is, the integration of equation (16) from 0 to ∞, which results in:

$$\Lambda_{sr>r_c} = 2\pi\lambda T \frac{(1+R\rho)\exp\left(-2R\rho\alpha_c\right)}{\rho^2} \tag{17}$$

The probability of no rain coming from outside the pixel, P_{12}, equals the probability that no storm cell center with an associated radius larger than the distance defined in equation (15) is born during the interval of time considered. This probability is:

$$P_{12} = \exp\left(-2\pi\lambda T \frac{(1+R\rho)e^{-2R\rho\alpha_c}}{\rho^2}\right) \tag{18}$$

Therefore, the probability, P_1, that no rain will occur within the pixel during a time interval of duration T is the product of P_{11} (equation (12)) and P_{12} (equation (18)):

$$P_1 = \exp\left(-\lambda T\left(A + 2\pi \frac{(1+R\rho)e^{-2R\rho\alpha_c}}{\rho^2}\right)\right) \tag{19}$$

Note that the effect of considering the possibility of rainfall coming from outside the pixel corresponds to an increase of the effective area of the pixel.

The probability of a pixel being dry at a randomly chosen moment, P_2, must be computed next. P_2 is also made up of two factors. The first factor is the probability (P_{21}) that no cell center born in the past within the pixel is still producing rainfall in the pixel at the instant of interest, that is, the duration of any cell center

having occurred in the past within the pixel must be smaller than the time lapse existing between the moment the cell center was born and the instant of interest. The second factor is the probability (P_{22}) that no cell center born in the past outside the pixel is still producing rainfall in the pixel at the instant of interest. This probability involves not only the duration of the cell center but also its radius.

Let us assume that the instant of interest is $t = 0$, and let us consider a pixel of area A located at time $-t$, that is, a distance t into the past. The number of cell centers born in the pixel during an infinitesimal time interval, dt, around that time instant, $-t$, follows a Poisson distribution with parameter, $\lambda A dt$. The parameter of the Poisson distribution does not depend on the past time instant in which the pixel is considered and remains constant in time.

Multiplying this parameter, $\lambda A dt$, by the probability that a cell being born at $-t$ is still producing rain at $t = 0$, one obtains the average number of cell centers born in a pixel during an infinitesimal time interval in the past that still produce rain at the instant of interest. A cell born at $-t$ will produce rain at $t = 0$ if its duration is larger than t. As cell durations are assumed to follow an exponential distribution, the probability of a duration larger than t is $\exp(-\eta t)$, and therefore the average number of cell centers born in a pixel during an infinitesimal time interval in the past that still produce rain at the instant of interest is:

$$\lambda_{d>t} = \lambda A \exp(-\eta t) \mathrm{d}t \tag{20}$$

To obtain the parameter of the Poisson distribution of the number of storm cells having occurred in the past within the pixel that still produce rain at the instant of interest, one needs to sum all the contributions from past instants, from the most recent ones ($t = 0$) to the most remote ones ($t = \infty$). Integrating equation (20) between these limits, the aforementioned parameter is obtained:

$$\Lambda_{d>\tau} = \frac{\lambda A}{\eta} \tag{21}$$

The probability of the pixel being dry from cell centers born inside the pixel is then:

$$P_{21} = \exp\left(-\frac{\lambda A}{\eta}\right) \tag{22}$$

Finally, the probability that no rain will be produced from cell centers born in the past outside the pixel of interest needs to be computed. In this case, an infinitesimal annulus covering an infinitesimal time interval in the past will be considered. Similarly to the derivation of equation (21), the parameter of the Poisson distribution of cell centers born in an interval dt in the past in an annulus outside the pixel whose duration and size are such that produces rainfall in the pixel is:

$$\lambda_{NRds} = 2\pi\lambda(R+r)e^{-\rho(r+2R\alpha_c)}e^{-\eta t}\,dr\,dt \tag{23}$$

which after integration yields the parameter of the distribution of all cell centers from the past and outside the pixel that produce rain at the considered time instant:

$$\Lambda_{NRds} = \frac{2\pi\lambda}{\eta}\frac{(1+R\rho)e^{-2R\rho\alpha_c}}{\rho^2} \tag{24}$$

The probability that none of these cell centers contributes to rain inside the circular pixel is then:

$$P_{22} = \exp\left(-\frac{2\pi\lambda}{\eta}\frac{(1+R\rho)e^{-2R\rho\alpha_c}}{\rho^2}\right) \tag{25}$$

Therefore, the probability of a pixel of area A being dry during a time interval of duration T is the product of equations (19), (22), and (25), which results in:

$$P_{dry} = \exp\left(-\lambda\left(T+\frac{1}{\eta}\right)\left(A+\frac{2\pi}{\rho^2}\left((1+R\rho)e^{-2R\rho\alpha_c}\right)\right)\right) \tag{26}$$

Equation (26) can be used to relate the parameters of the space-time rainfall model to the probability of dry intervals at a pixel. This will be used in the next section toward the characterization of the statistical structure of rainfall at a point which is commonly needed in ecohydrological studies.

3. Methodology

The goal is to use satellite rainfall data toward the statistical characterization of the rainfall process at a point. This will allow the estimation of the parameters of point models of rainfall, commonly required for many ecohydrological purposes [*Rodríguez-Iturbe and Porporato*, 2004]. For every location of interest, the satellite pixel containing a specific rain gage site is selected. From the time series of daily average rainfall over the pixel, obtained from satellite information, the mean, standard deviation, lag-1 temporal autocovariance, and percentage of dry days are then estimated. The spatiotemporal rainfall model described in section 2 is then fitted to the satellite observations using its integrated characteristics.

The model parameters thus obtained allow us to statistically describe the rainfall process at a point. The goodness of the methodology, which is based on a simple space-time rainfall model, is evaluated through a comparison of the point statistics thus derived with those obtained from rain gages existing inside the pixels under consideration and whose data have not been used at all for the estimation based on satellite integrated measurements.

This study uses rainfall data derived from TRMM using the 3B42 algorithm [*Huffman et al*., 2009]. TRMM-3B42 rainfall estimates provide 3 hourly time series of rainfall intensity on a global grid with a resolution of $0.25^{\circ} \times 0.25^{\circ}$. For the purposes of this paper, rainfall data are aggregated to daily values which are the most commonly used in ecohydrological studies. Rainfall statistics were computed for four locations in the Kalahari transect during the growing seasons of the period 2000–2013. The growing season, assumed statistically homogeneous, runs from October to April in the Kalahari transect [*Porporato et al*., 2003]. Different seasons in a region may be dominated by different types of precipitation, i.e., convective or frontal. The parameters of the model will be different and their estimation should be carried out separately for each season.

The rainfall model described in section 2 incorporates four parameters: λ, ρ, η, and β. We thus need a minimum of four equations to estimate these parameters. These four equations result from equating four statistics of the integrated model in space and time with the statistics of satellite observations. The statistics chosen are mean daily rainfall, variance of daily rainfall, lag-1 temporal autocovariance of daily rainfall, and percentage of dry days over the pixel under analysis. The integration process in space and time forces the four equations to be solved simultaneously.

The parameter α_c, described in the previous section and related to the minimum overlap between a storm cell and the pixel, could, in principle, vary from one region to another, although one would expect little change of its value within topographically homogeneous regions, like the ones studied in this paper. Thus, α_c is calibrated here for only one of the locations under study (although it could have been estimated independently for each region). The estimation procedure for the model parameters λ, ρ, η, and β is repeated for several values of the parameter α_c at the pixel located at Mongu, retaining the value which best reproduces the satellite statistics related to that location. That value of α_c is used for the other locations without further calibration.

The mean of the integrated model is computed by means of equation (1) and equated to the average observed by the satellite. The variance of the integrated process is computed from equation (9), where $l = 0$, $m = 0$, and $n = 0$. Lag-1 autocorrelation is also computed from equation (9) with $l = 0$, $m = 0$, and $n = 1$. In all cases, daily rainfall data integrated over the pixel are used in the calibration. The percentage of dry days of the integrated model is calculated making use of equation (26).

The nonlinear nature of the equations preclude direct solutions. However, it is easy to compute the properties of the integrated process given the parameters of the spatiotemporal process. For this reason, the parameter estimation problem is cast as an optimization problem.

The function to be optimized is the total relative root squared error, ϵ, measured over the properties of the integrated process that match with satellite measurements:

$$\epsilon = \sqrt{\left(\frac{\bar{r}_{int} - \bar{r}_{sat}}{\bar{r}_{sat}}\right)^2 + \left(\frac{\sigma^2_{int} - \sigma^2_{sat}}{\sigma^2_{sat}}\right)^2 + \left(\frac{\tau^{(1)}_{int} - \tau^{(1)}_{sat}}{\tau^{(1)}_{sat}}\right)^2 + \left(\frac{p^{dry}_{int} - p^{dry}_{sat}}{p^{dry}_{sat}}\right)^2} \tag{27}$$

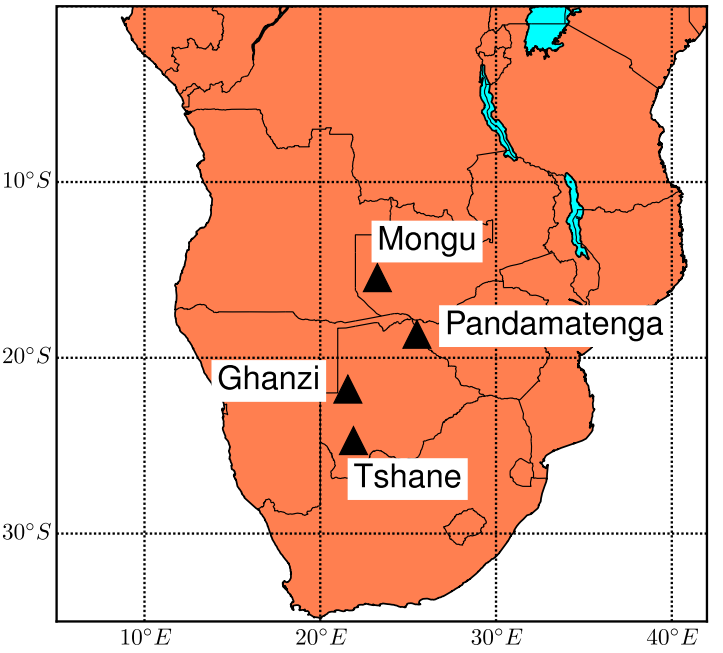

Figure 2. Location of the four rain gage stations in the Kalahari transect, southern Africa.

In equation (27), the "*int*" subscript denotes an integrated model magnitude and the "*sat*" subscript denotes a satellite-derived magnitude. $\bar{r}$ refers to the average of daily rainfall, σ^2 refers to the daily rainfall variance, $\tau^{(1)}$ refers to the daily lag-1 autocovariance, and p^{dry} refers to the percentage of dry days.

The optimization process is carried out using the Particle Swarm Optimization (PSO) technique [*Kennedy*, 2010]. PSO is a computational method to solve optimization problems iteratively. The method uses a population (swarm) of candidate solutions (particles). The particles keep track of the best position that they ever visited, and also keep track of the best position ever visited by any particle in the whole swarm. Optimization is achieved by moving the particles through the search space. The velocity of every particle is constructed as a combination of three vectors: (1) its previous velocity, (2) the vector joining the particle with the swarm best visited position, and (3) the vector joining the particle with its own best visited position. The information exchange between particles allows them to share the optimum position and tends to drive the solution to the overall optimum. This study uses a swarm of 1000 particles.

4. Study Site Description

For the present study, four locations (shown in Figure 2) have been selected in southern Africa to carry out the satellite to point rainfall downscaling. These four locations, identified in Table 1, were selected in the Kalahari transect, a region where many ecological and hydrological studies have been carried out [*Shugart et al.*, 2004; *Scanlon et al.*, 2007; *Caylor et al.*, 2003], allowing a comparison of results with those of previous studies.

Table 2 lists rainfall statistics, derived from satellite observations, for every location that has been analyzed in the present study [*National Climatic Data Center*, 2014]. A strong north-south precipitation gradient can be observed along the transect between Mongu and Tshane. The reduction of precipitation is accompanied also by a reduction of the frequency of rainfall events, therefore increasing the percentage of dry days at the different locations.

5. Results

The parameter estimation procedure described in the previous section is used to compute the parameters of the spatiotemporal rainfall model for the locations analyzed in this study. The value of the parameter α_c, representing the minimum overlap required in the model between a cell and a pixel for the satellite to identify it as a rainfall event, is 0.6 for all the four locations. This value was estimated as the one providing the best reproduction of satellite statistics for the Mongu location.

Table 1. Name and Location of the Rain Gage Stations Used for the Study From *Caylor et al.* [2006]

Location	Latitude (°)	Longitude (°)
Mongu	−15.44	23.25
Pandamatenga	−18.66	25.50
Ghanzi	−21.78	21.57
Tshane	−24.17	21.89

Parameter values are transformed to storm characteristics in order to simplify their

Table 2. Satellite Rainfall Statistics for the Pixel Containing the Rain Gage Locations Listed in Table 1

Location	Average (mm/d)	Std. Dev.(mm/d)	Lag-1 Temp. (autocorr.)	Dry Days (%)
Mongu	4.81	8.43	0.18	37
Pandamatenga	2.97	6.61	0.22	57
Ghanzi	2.21	6.02	0.22	59
Tshane	1.70	4.66	0.16	70

Table 3. Parameters of the Spatiotemporal Rainfall Model (See Section 2) for the Rain Gage Locations Listed in Table 1, Expressed as Storm Characteristics

	λ (cells/d/km^2)	$1/\rho$ (km)	$24/\eta$ (h)	$1/\beta$ (mm/d)
Mongu	8.52×10^{-4}	10.92	6.72	26.95
Pandamatenga	6.73×10^{-4}	9.43	7.92	23.75
Ghanzi	6.30×10^{-4}	7.63	8.00	28.90
Tshane	4.65×10^{-4}	9.71	6.02	24.51

interpretation. They are listed in Table 3. The inverse of each parameter is computed with the exception of the frequency of arrivals which remains unchanged. It can be observed that the most important factor controlling the rainfall gradient is the reduction in the frequency of storms from north to south along the Kalahari transect. This same feature has been captured in the previous studies [*Porporato et al*., 2003; *Caylor et al*., 2006]. Table 4 shows the comparison of satellite rainfall statistics and the integrated model statistics.

It should be mentioned that the direct interpretation of parameter values as the physical characteristics of storms should be carefully weighted by the fact that the model is a very simplistic representation of very complex atmospheric dynamics.

The parameters ρ, η, and β, which describe the physical characteristics of rain cells, do not show a clear trend along the transect. This is also compatible with data reported in previous studies [*Porporato et al*., 2003; *Caylor et al*., 2006] where the average amount of rain per rainy day remains relatively stable along the transect gages.

A major advantage arising from the characterization of rainfall through a spatiotemporal model lies on its possible use to describe precipitation patterns at different levels of aggregation on space and time. For ecohydrologic analyses, the local characterization of daily rainfall at a point is of key importance. We now proceed to estimate the daily rainfall statistics at a point for each of the four locations in Table 1, based on the results from the integrated space-time rainfall model calibrated on the basis of satellite measurements. The results will then be compared with the statistics resulting from daily rainfall data collected by rain gages located inside the pixel used for the calibration of the space-time rainfall model.

The comparison proceeds as follows. The probability of a dry day at a point, e.g., at a rain gage, corresponds to the product of two probabilities: (1) the probability of a random space-time point being dry given by $\exp\left(\frac{-2\pi\lambda}{\eta\rho^2}\right)$ [*Cox and Isham*, 1988] and (2) the probability of no rain occurring during a day at the point given by $\exp\left(\frac{-2\pi\lambda}{\rho^2}\right)$. Multiplying both probabilities one obtains the probability of a dry day at a given point derived from the spatiotemporal model:

$$P_{dry} = \exp\left\{\frac{-2\pi\lambda}{\eta\rho^2}(1+\eta)\right\} \tag{28}$$

The expressions for the daily rainfall variance at a point and daily lag-1 autocorrelation at a point corresponding to the space-time rainfall model are given by *Rodríguez-Iturbe et al*. [2006]:

Table 4. Comparison of Satellite Statistics (Obs.) and Integrated Model Statistics (Mod.) for the Four Study Locations Listed in Table 1

	Mongu		Pandamatenga		Ghanzi		Tshane	
Statistic	Obs.	Mod.	Obs.	Mod.	Obs.	Mod.	Obs.	Mod.
Average (mm/d)	4.81	4.81	2.97	2.97	2.21	2.21	1.70	1.70
Standard deviation (mm/d)	8.43	8.71	6.61	6.61	6.02	6.02	4.66	4.65
Lag-1 autocorrelation	0.18	0.18	0.22	0.22	0.22	0.22	0.16	0.16
% time dry	37	37	57	48	59	55	69	60

Table 5. Comparison of Rain Gage and Model Derived Point Rainfall Statistics for the Four Study Locations Listed in Table 1

	Mongu		Pandamatenga		Ghanzi		Tshane	
Statistic	Obs.	Mod.	Obs.	Mod.	Obs.	Mod.	Obs.	Mod.
Average (mm/d)	4.21	4.81	2.72	2.97	1.95	2.21	1.57	1.70
Standard deviation (mm/d)	10.17	10.28	8.26	7.99	6.71	7.63	6.09	5.61
Lag-1 autocorrelation	0.14	0.18	0.21	0.22	0.16	0.22	0.19	0.16
% time dry	56	44	68	60	74	74	79	71

$$\sigma_T^2 = \frac{8\pi\lambda}{\eta^3\beta^2\rho^2}\left(e^{-\eta}+\eta-1\right) \tag{29}$$

$$\gamma_T = \frac{e^{-\eta}\left(e^{-\eta}-2+e^{\eta}\right)}{2\left(e^{-\eta}+\eta-1\right)} \tag{30}$$

Comparisons of the statistics resulting from rain gage observations at the different locations and model derived statistics are shown in Table 5. The point rainfall statistics obtained from the spatiotemporal model after the appropriate transformations are judged to be an adequate representation of the rain gage information.

6. Conclusions

Because of the integrated character of rainfall satellite measurements, their statistics cannot be directly translated to point, i.e., rain gage statistics. Satellite rainfall estimates are becoming widely available throughout the world and, in regions with very sparse or nonexistent rain gage coverage, translation of pixel integrated statistics to point rainfall properties is necessary for many types of hydrologic and ecohydrologic studies where vegetation responds to local rainfall conditions rather than to spatial averages.

This paper presents a downscaling methodology to derive rainfall statistics at a point from integrated satellite measurements. Moreover, the inverse problem, that of estimating areal characteristics of rainfall from point measurements, is also approachable using the analytical framework presented here.

The downscaling procedure is flexible enough to accommodate transformations among different variables measured with different degrees of aggregation. We focused in how to transform statistics from pixel integrated daily rainfall satellite observations into rain-gage-equivalent characteristics at the daily scale, but different spatial and temporal resolutions could have been used.

The procedure based on a very simple space-time rainfall model has been shown to yield satisfactory results for four different locations along a strong rainfall gradient in the Kalahari transect. The results reproduce general rainfall patterns observed in previous studies, such as the reduction in storm frequency along the transect, and the relative homogeneity of average rainfall depth of rainy days along the transect. Most importantly, it also satisfactorily reproduces crucial rainfall characteristics for ecohydrological applications, such as the variance, temporal autocorrelation, and the percentage of dry days at this local scale.

A most important assumption of the methodology is the one related to the statistically homogeneous structure of the rainfall process over the area in consideration. Its extension to mountainous regions would require a more complex space-time rainfall model with spatially varying parameters throughout the region under analysis.

Acknowledgments
The data used for this study are publicly available. Satellite rainfall measurements have been obtained from NASA's STORM data access interface (https://storm-pps.gsfc.nasa.gov/storm/). Data set: TRMM-3B42. Rain gage information has been obtained from NOAA's National Climatic Data Center (https://www.ncdc.noaa.gov/cdo-web/datasets).

References

Caylor, K. K., H. H. Shugart, P. R. Dowty, and T. M. Smith (2003), Tree spacing along the Kalahari transect in southern Africa, *J. Arid Environ.*, *54*(2), 281–296, doi:10.1006/jare.2002.1090.

Caylor, K. K., P. D'Odorico, and I. Rodríguez-Iturbe (2006), On the ecohydrology of structurally heterogeneous semiarid landscapes, *Water Resour. Res.*, *42*, W07424, doi:10.1029/2005WR004683.

Cox, D. R., and V. Isham (1988), A simple spatial-temporal model of rainfall, *Proc. R. Soc. London, Ser. A*, *415*(1849), 317–328, doi:10.1098/rspa.1988.0016.

Herrera, R. G., D. G. Puyol, E. H. Martín, and L. G. Presa (2001), Influence of the North Atlantic oscillation on the Canary Islands precipitation, *J. Clim.*, *14*(19), 3889–3903, doi:10.1175/1520-0442(2001)014<3889:IOTNAO>2.0.CO;2.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2009), The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.* *8*(1), 38–55, doi:10.1175/JHM560.1.

Isham, V., D. R. Cox, I. Rodríguez-Iturbe, A. Porporato, and S. Manfreda (2005), Representation of space-time variability of soil moisture, *Proc. R. Soc. A*, *461*(2064), 4035–4055, doi:10.1098/rspa.2005.1568.

Kennedy, J. (2010), Particle swarm optimization, in Claude Sammut, Geoffrey I. Webb editors *Encyclopedia of Machine Learning*, pp. 760–766, Springer, US, doi:10.1007/978-0-387–30164-8_630.

Müller, M. F., and S. E., Thompson (2013), Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal, *Adv. Water Resour.*, *60*, 121–134, doi:10.1016/j.advwatres.2013.08.004.

National Climatic Data Center (2014), *Climate Data Online*. National Climatic Data Center Federal Building 151 Patton Avenue Asheville, NC 28801-5001 USA. [Available at https://www.ncdc.noaa.gov/cdo-web/.].

Porporato, A., F. Laio, L. Ridolfi, K. K. Caylor, and I. Rodríguez-Iturbe (2003), Soil moisture and plant stress dynamics along the Kalahari precipitation gradient, *J. Geophys. Res.*, *108*(D3), 4127, doi:10.1029/2002JD002448.

Rinaldo, A., E. Bertuzzo, L. Mari, L. Righetto, M. Blokesch, M. Gatto, R. Casagrandi, M. Murray, S. M. Vesenbeckh, and I. Rodríguez-Iturbe (2012), Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections, *Proc. Natl. Acad. Sci. U. S. A.*, *109*(17), 6602–6607, doi:10.1073/pnas.1203333109.

Rodríguez-Iturbe, I., and A. Porporato (2004), *Ecohydrology of Water-Controlled Ecosystems: Soil Moisture and Plant Dynamics*, Cambridge Univ. Press, Cambridge, U. K.

Rodríguez-Iturbe, I., V. Isham, D. R. Cox, S. Manfreda, and A. Porporato (2006), Space-time modeling of soil moisture: Stochastic rainfall forcing with heterogeneous vegetation, *Water Resour. Res.*, *42*, W06D05, doi:10.1029/2005WR004497.

Scanlon, T. M., K. K. Caylor, S. A. Levin, and I. Rodríguez-Iturbe (2007), Positive feedbacks promote power-law clustering of Kalahari vegetation, *Nature*, *449*(7159), 209–212, doi:10.1038/nature06060.

Shugart, H. H., S. A. Macko, P. Lesolle, T. A. Szuba, M. M. Mukelabai, P. Dowty, and R. J. Swap (2004), The SAFARI 2000—Kalahari Transect Wet Season Campaign of year 2000, *Global Change Biol.*, *10*(3), 273–280, doi:10.1111/j.1365–2486.2004.00732.x.

Vanmarcke, E. (2010), *Random Fields: Analysis and Synthesis*, World Scientific Publishing Co. Pte. Ltd. 5 Toh Tuck Link, Singapore 596224.

Wright, D. B., J. A. Smith, G. Villarini, and M. L. Baeck (2012), Hydroclimatology of flash flooding in Atlanta, *Water Resour. Res.*, *48*, W04524, doi:10.1029/2011WR011371.