Stochastic Environmental Research and Risk Assessment 16 (2002) 188–206 © Springer-Verlag 2002
DOI 10.1007/s00477-002-0094-4

A two steps disaggregation method for highly seasonal monthly rainfall

L. Guenni, A. Bárdossy

Abstract. The need for high resolution rainfall data at temporal scales varying from daily to hourly or even minutes is a very important problem in hydrology. For many locations of the world, rainfall data quality is very poor and reliable measurements are only available at a coarse time resolution such as monthly. The purpose of this work is to apply a stochastic disaggregation method of monthly to daily precipitation in two steps: 1. Initialization of the daily rainfall series by using the truncated normal model as a reference distribution. 2. Restructuring of the series according to various time series statistics (autocorrelation function, scaling properties, seasonality) by using a Markov chain Monte Carlo based algorithm. The method was applied to a data set from a rainfall network of the central plains of Venezuela, in where rainfall is highly seasonal and data availability at a daily time scale or even higher temporal resolution is very limited. A detailed analysis was carried out to study the seasonal and spatial variability of many properties of the daily rainfall as scaling properties and autocorrelation function in order to incorporate the selected statistics and their annual cycle into an objective function to be minimized in the simulation procedure. Comparisons between the observed and simulated data suggest the adequacy of the technique in providing rainfall sequences with consistent statistical properties at a daily time scale given the monthly totals. The methodology, although highly computationally intensive, needs a moderate number of statistical properties of the daily rainfall. Regionalization of these statistical properties is an important next step for the application of this technique to regions in where daily data is not available.

L. Guenni (✉)
Centro de Estadística y Software Matemático
Universidad Simón Bolívar. Apartado 89.000
1080-A Caracas, Venezuela
e-mail: lbravo@cesma.usb.ve

A. Bárdossy
Institut für Wasserbau, Universität Stuttgart
D-70569, Germany

Substantial parts of this work were carried out while the first author was on sabbatical at the University of New Hampshire (USA) at the Complex Systems Research Center, Institute for the Study of the Earth, Oceans, and Space. We give thanks to Charles Vörösmarthy from the University of New Hampshire (USA) and to Michael Hutchinson from the Australian National University (Australia) for useful discussions. We also give thanks to an anonymous referee for constructive comments.

Keywords: Rainfall disaggregation, Seasonal rainfall, Simulated annealing, Truncated normal model, Rainfall models

1 Introduction

One of the main problems with hydrological times series is that measured values are usually given at a time scale coarser than the one needed. For example, daily rainfall values are usually measured at many rainfall recording stations. But hourly values are usually required for drainage systems design or many other applications in hydrology and environmental sciences.

For many parts of the world, rainfall data quality is poor and raingauge networks are very sparse, which poses a limitation in the estimation of spatially continuous rainfields, especially at a short time scale (Guenni et al., 1997). More reliable data is available, for example, at a monthly time step than at a daily time step. Also many of the global gridded data sets are available at a monthly scale (New et al., 2000).

Since a disaggregated series is a realization from the original aggregated time series, stochastic approaches are necessary in order to reproduce the right statistical characteristics of the data at the required time scale. According to Torfs, 1997, a disagregation technique must be a statistical simulation technique, since a deterministic approach is impossible in practical terms. The reason for this is that the original process which is normally unknown is reconstructed from the aggregated process.

A significant amount of contributions on disaggregation methods give support to the importance of this technique to solve this data limitation problem (Valencia and Schaake, 1972; Curry and Bras, 1978; Brass and Rodriguez-Iturbe, 1985; Grygier and Stedinger, 1988; Hershenhorn and Woolhiser, 1992; Santos and Salas, 1992). More recent contributions are presented by Glasbey et al. (1995), Maheepala and Perera (1996), Torfs (1997), Bürger (1997), Lebel et al. (1998), Tarboton et al. (1998), Connolly et al. (1998), Gyasi-Agyei (1999), Burian et al. (2000). In many of the techniques a distributional assumption about the disaggregated series is required which is usually the normality assumption. Since the most common hydrological variables, rainfall and streamflow, are usually non-normal, a normalizing transformation is required (Maheepala and Perera, 1996; Bürger, 1997). Tarboton et al. (1998) propose a non-parametric approach where the joint probability density function is estimated directly from the historical data using kernel density estimates. Burian et al. (2000) used artificial neural networks to disaggregate hourly rainfall at a single location. The required joint probability density function is $f(\mathbf{X}, Z)$, where $\mathbf{X}$ is the vector of disaggregated variables (e.g. monthly rainfall, hourly streamflows) and Z is the aggregated variable (annual rainfall, daily streamflows). Given the aggregated variable Z the problem can be posed as a sampling from the conditional probability density function:

$$f(\mathbf{X}|Z) = \frac{f(\mathbf{X}, Z)}{\int f(\mathbf{X}, Z)\mathrm{d}\mathbf{X}} \tag{1}$$

Most of the existing methodologies are used for the disaggregation of rainfall time series at a single location. Recently Lebel et al. (1998), presented a space-time disaggregation approach for areal storm depth which deals separately the spatial and temporal variabilities. The spatial disaggregation is based on the turning bands method (TBM) (Matheron, 1982) and the temporal disaggregation is achieved by imposing a standard hyetogram at each location of the spatial disaggregation domain. The methodology used as input data the mean areal rainfall depth as estimated from a dense recording raingauge network for the mesoscale convective complexes occurred in a Niger region. Gyasi-Agyei (1999), used a point process based model to demonstrate how the parameters of this model could be regionalized for hourly disaggregation.

Due to the intermitence characteristics of the rainfall process and the lack of gaussianity, rainfall still poses the most challenging problems in the disaggregation methods. To illustrate the disaggregation problem with rainfall data we present in Fig. 1 aggregated time series of rainfall data from one location at the central plains of Venezuela in Guárico State. In this figure, the daily, monthly, quarterly and annual rainfall intensities in units of mm/day are shown. It is clear from this figure that the rainfall process at a shorter time scale is by far more complex that the rainfall process at a much coarse time scale. Infinite realizations of the process at a daily time scale would result in the same process at a larger time scale.

If the monthly values are available for a particular location, the purpose of the disaggregation procedure is to reproduce a time series that "looks like" the original daily series. There are a number of desirable characteristics that should be preserved from the original time series. As a first step, the proportion of dry periods within the aggregated period should be reproduced. Serial dependence, simple scaling properties, probability distribution are some of the many characteristics that could be considered.

For non-stationary series with a strong seasonal component, many of the desirable characteristics that should be preserved are also non-stationary and usually seasonal. This is the case of the data set that will be analyzed in this work. The simulated annealing technique is applied such that an objective function with prescribed characteristics of the observed daily series is minimized in the simulation procedure. A similar approach was presented in Bárdossy (1998), where the technique was applied to selected stations in the Ruhr catchment, Germany, to simulate hourly and 5 min rainfall.

The paper is organized as follows: the data set is described in Sect. 2. In Sect. 3 the computational disaggregation methodology is explained. The results and validation of the methodology are presented in Sect. 4. Finally the discussion and conclusions indicating possible extensions of the technique are presented.

2
Data set

In order to apply the disaggregation method, we need to investigate the statistical properties of the daily rainfall series that will be preserved by the proposed technique. Daily rainfall data from fourteen stations of Guárico state, located at the Central Plains of Venezuela were used for this purpose. A diagram of the position of the selected locations is shown in Fig. 2. The stations are located in an area of $250 \times 150\,\mathrm{km}^2$ approximately and the southern limits are roughly in the same direction as the Orinoco river.

These locations were selected among a number of 120 stations available because of their better data quality and larger number of records. Data from 1967 to 1991 were used. Rainfall at these locations is strongly seasonal with a dry season from November to April, and a rainy season from May to October. A boxplot diagram for each month is shown in Fig. 3 for two selected locations, which clearly shows the seasonal pattern of the data.

The high seasonality in the data suggests that the probability distribution parameters and the serial and temporal scaling characteristics might change with

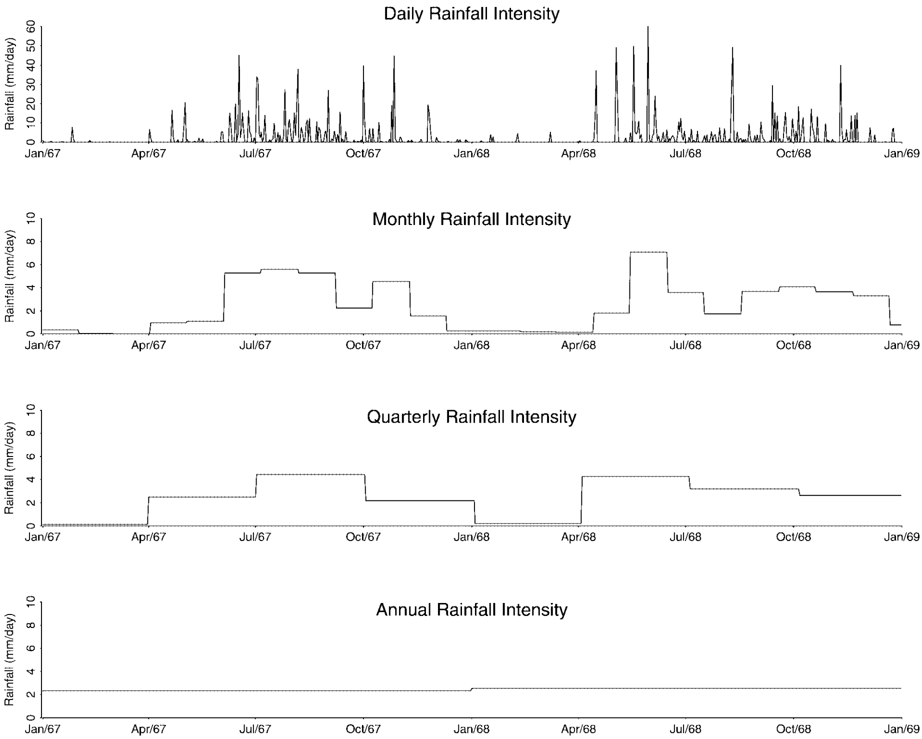

Fig. 1. Precipitation time series for different aggregation (annual, quarterly, monthly and daily) time periods in mm/day

time of the year. This assumption was investigated from the observed data and subsequently used in the disaggregation procedure.

3 Methodology

3.1 Computational disaggregation procedure

Let

$$Z_j^i = \sum_{k=1}^{n_j} z_j^i(k) \tag{2}$$

where Z_j^i is the rainfall amount on month j and year i; $z_j^i(k)$ is the daily amount on day k of month j and year i and n_j is the number of days of month j.

In general, it is reasonable to assume that if the number of wet days increases the monthly rainfall amount Z_j^i will increase. Similar attempts to relate the number of wet days with the rainfall amounts have been presented by other authors. For example, New et al. (2000) use a power function to relate the wet day frequency with the monthly amounts in order to produce global maps of this variable conditioned on the monthly rainfall values. A similar strategy is used here to estimate the proportion of wet days conditioned on the observed monthly amounts.

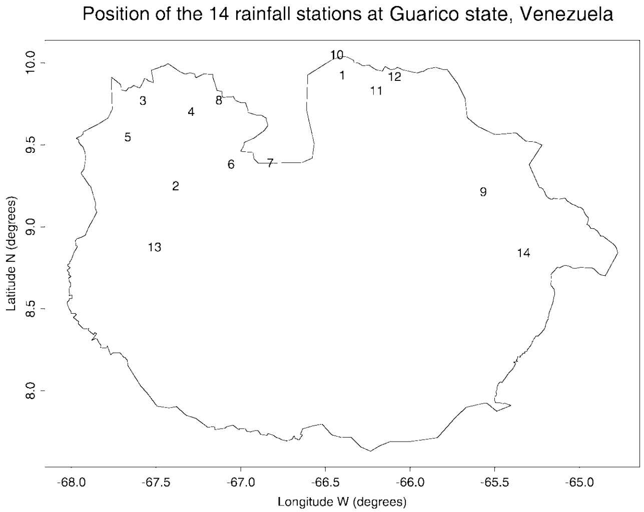

Fig. 2. Position for the 14 stations used in the analysis

The relationship between the probability of having a wet day p_w^j and the rainfall amount Z_j^i was investigated. Different nonlinear models, including the power function model proposed by New et al. (2000) and the exponential model proposed by Vörösmarty et al. (1998) were fitted to the observed data for separate seasons (wet and dry season) and different locations. The best fit was found by using a nonlinear model of the form:

$$p_w^j = \beta_0(1 - \exp(-\beta_1 * Z_j)) + \varepsilon$$

where the parameters β_0 and β_1 are estimated by conventional nonlinear least square and ε is assumed a normally distributed variable with zero mean and variance σ^2.

The observed values and fitted models are shown in Fig. 4 by combining all months of the dry season (November–April) and all months of the wet season (May–October) for locations 1 and 5.

The disaggregation procedure is then carried out in two steps:

- Generate an initial sequence of rainfall values with an initial estimate of the proportion of wet or dry sub-periods conditioned on the monthly value and an appropriate reference probability distribution.
- Re-order iteratively the arbitrary sequence of sub-period values until convergence is reached to a new series with similar serial dependence and similar temporal scaling characteristics to some prescribed values.

These two steps are described in more detail as follows:

1. Generating a daily rainfall sequence by using the truncated normal model as a reference distribution

Let $z_j(k)$ the daily rainfall amount for month j and day k. An initial assumption is that the $z_j(k)$ are independent and identically distributed random variables. It is assumed that each $z_j(k)$ is a sample of a random variable that has been raised to an appropiate positive power and truncated for values of the random variable below a given threshold t. Following Stidd (1973), Hutchinson et al. (1993) and Henze and Klar (1993), the truncated normal model (TNM) can be written as:

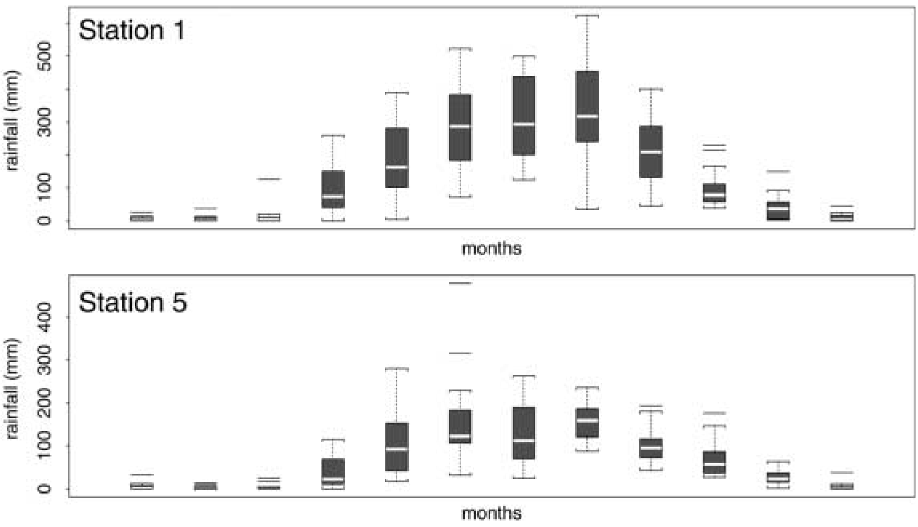

Fig. 3. Monthly boxplots of rainfall amounts at two locations

$$z_j(k) = \begin{cases} w_j^{\beta}(k) & \text{if } w_j(k) > t^{1/\beta} \\ 0 & \text{if } w_j(k) \le t^{1/\beta} \end{cases} \tag{3}$$

such that $w_j(k)$ has a normal distribution $N(\mu_j, \sigma_j^2)$ with parameters μ_j and σ_j^2 and β which is a positive value to be estimated. The subscript j indicates that these parameters might change on a monthly basis. A typical value of β is 3 as suggested by Stidd (1973). Hutchinson et al. (1993), using ground based data across the contiguous US estimated a value of β around 1.82. He found that smooth surfaces as function of location could be fitted to the β values as estimated from point data by using maximum likelihood.

Typically the parameters μ, σ^2 and β are estimated by maximum likelihood (Henze and Klar, 1993; Hutchinson et al., 1993) or by using a Bayesian approach (Sansó and Guenni, 1998; Sansó and Guenni, 1999). Maximization of the the log likelihood function can be achieved by calculating the derivatives of the log likelihood function with respect to μ, σ^2 and β and by solving the resulting set of nonlinear equations. A recurrence relationship is obtained in terms of beta (Hutchinson et al., 1993) and 10–20 iterations are usually enough to reach convergence. The availability of daily data is essential to be able to estimate all the model parameters with this approach.

Envisioning the application of this methodology under data restrictive conditions, in where daily data might not be available, a different approach for the estimation of the model parameters was attempted. The number of wet days and the monthly amounts are more readily available for many locations than the daily values. By assuming a constant value of β the parameters μ and σ can be estimated by the method of moments with previous knowledge of the proportion of wet days in a given month and the monthly rainfall for the month.

From Eq. (3), the probability of rainfall falling below a threshold value t is given by:

$$Pdry = \Phi\left(\frac{t^{1/\beta} - \mu}{\sigma}\right)$$

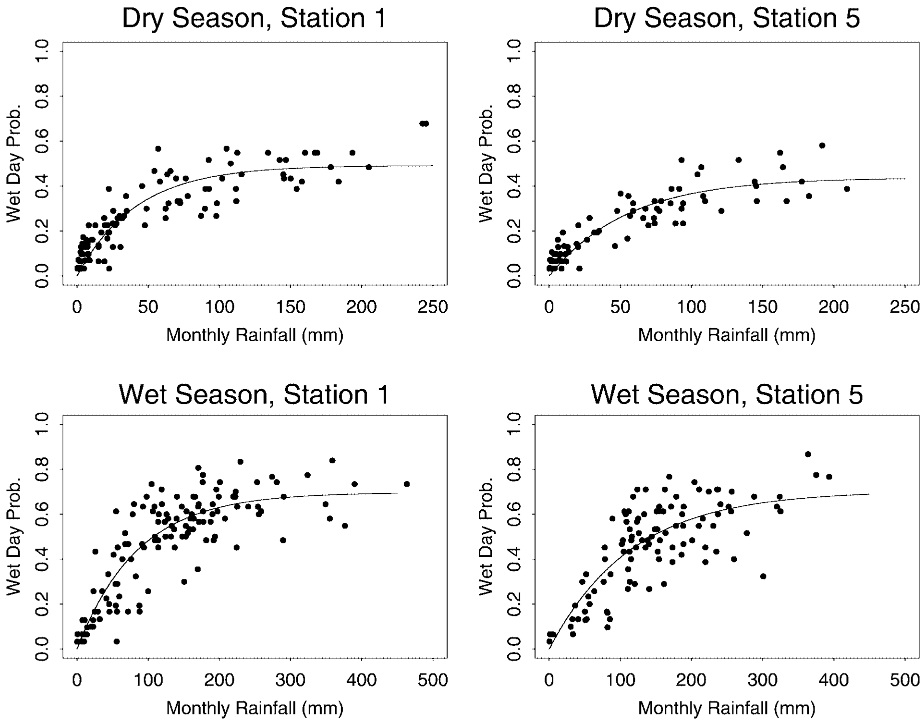

Fig. 4. Relationship between the probability of a wet day and the monthly rainfall amount for the dry season (November–April) and the wet season (May–October) at locations 1 and 5

where Φ is the standard normal cumulative distribution function. Let $\alpha = 1/\beta$. Assuming that α is known and $Pdry$ can be estimated from a previously known relationship between the aggregated rainfall amounts and the number of wet days ($Pdry = 1 - n_w/n$, where n_w is the number of wet days and n is the total number of days in the period or month), the quantile, $(t^\alpha - \mu)/\sigma$ corresponding to the value of $Pdry$ can be estimated as $\Phi^{-1}(Pdry)$.

Assuming the model described in Eq. (3) holds, the expected value of the daily rainfall z can be calculated as:

$$E(z) = \int_{t^\alpha}^{\infty} \frac{w^\beta}{\sqrt{2\pi}\sigma} \exp\left(-\left(\frac{w-\mu}{\sigma}\right)^2 \Big/ 2\right) \mathrm{d}w \tag{4}$$

This equation can be re-written as:

$$E(z) = \sigma^\beta \int_{\frac{t^\alpha - \mu}{\sigma}}^{\infty} (x + \mu/\sigma)^\beta \phi(x) \mathrm{d}x \tag{5}$$

where $x = (w - \mu)/\sigma$ and $\phi(x)$ is the standard normal density function.

By solving the integral of the previous equation by numerical integration and by equating the left hand side of this equation to the sample estimate of $E(z)$, it is possible to get an estimate of σ and therefore, an estimate of μ (Hutchinson, 2001).

By fitting the model with the above methodology, and by sequentially changing the values of β from 0.8 to 20, it was observed that increasing values of β resulted in increasing values of μ during the dry months, while increasing values of β resulted in decreasing values of μ during the wet months. These results are consistent with the results found by Hernandez (1998) about the correlation between μ and β. By using monthly data from the same locations and the maximum likelihood method, Hernandez (1998) also found that the values of β tend to be lower during the dry months than during the wet months. This means that the normalizing power $\alpha = 1/\beta$ tend to be higher during the dry season than during the wet season. This is consistent with the more clear observed departure from normality of the rainfall data during the dry season than during the wet season. Therefore it is expected that because of the significant correlation among all the parameters, the values of μ and σ will change accordingly with the value of β. A constant value of $\beta = 2.5$ was used in this analysis. Theoretical results suggests, that the probability density function of a random variable following a truncated normal model has Gumbel distribution which is independent of β. (Sansó and Hernandez, personal communication), therefore it is expected, from the point of view of the extremes, that the arbitrary choice of $\beta = 2.5$ will not affect the extreme behaviour of the resulting disaggregated series.

After μ and σ are estimated, a sequence of n_j values $z_j(1), z_j(2), \ldots, z_j(n_j)$ are sampled from the model described by Eq. (3) as the initial daily rainfall sequence.

2. Re-ordering of the daily rainfall amounts within a month by simulating annealing

At this stage we have found a sequence of daily rainfall values $z_j(1), z_j(2), \ldots, z_j(n_j)$ for a given month j which are not following any particular time structure according to the autocorrelation properties of the original time series. There are other charateristics of the daily time series as the probability distributions of dry and wet runs and the simple scaling properties as described by Burlando and Rosso (1996), that will not be preserved by the unstructured series.

A computational intensive procedure based on the Metropolis-Hastings algorithm, a powerful markov chain Monte Carlo (MCMC) simulation method, is used to generate a sequence of irreducible Markov chains with a stationary distribution $\phi(.)$. This stationary distribution is precisely the multivariate probability distribution of interest $\pi(.)$. A more detailed description of the Metropolis-Hastings algorithm can be found in Chib and Greenberg (1995). Given an initial time series ζ_1, a new candidate series ζ_2 is selected from a proposed distribution $q(./\zeta_1)$.

The new candidate is accepted with probability $\alpha(\zeta_1, \zeta_2)$, where:

$$\alpha(\zeta_i, \zeta_j) = \min\left(1, \frac{\pi(\zeta_j)q(\zeta_i/\zeta_j)}{\pi(\zeta_i)q(\zeta_j/\zeta_i)}\right)$$

In the Metropolis algorithm, $q(./\zeta_1)$ is selected such that $q(\zeta_i/\zeta_j) = q(\zeta_j/\zeta_i)$. Therefore, $\alpha(\zeta_i, \zeta_j)$ only depends on the ratio $\pi(\zeta_j)/\pi(\zeta_i)$.

The limiting distribution $\pi(.)$ is defined with the help of an objective function O. This approach uses the "annealing theorem" combined with the Metropolis-Hastings algorithm to produce time series corresponding to the minima of O sampled from the limiting distribution:

$$\pi(\zeta) = K \exp(-O(\zeta)) \qquad (6)$$

where $O(\zeta)$ is the objective function for a given time series ζ and K is a constant to ensure that $\pi(\zeta)$ is a probability distribution function. Since the algorithm only requires the ratio $\pi(\zeta_j)/\pi(\zeta_i)$, the value of K does not need to be calculated.

A similar description of this approach is found in Bárdossy (1998). Smaller values of the objective function $O(\zeta)$ are attained when the time series ζ fulfills the prescribed properties of the observed precipitation series. In order to avoid local minima when searching for a candidate for a global optimum, a new parameter called the "temperature" T is introduced such that the limiting distribution is now a function of T:

$$\pi_T(\zeta) = K(T) \exp\left(-\frac{O(\zeta)}{T}\right) \tag{7}$$

T is a control parameter which guarantees convergence of the objective function to the global minima and it is reduced gradually (generally $T \to 0$) according to a user-specified schedule. Note that for realizations ζ with $O(\zeta) > 0$ the probabilities decrease with decreasing T and for realizations ζ with $O(\zeta) = 0$ the probabilities increase with decreasing T.

The following procedure is described to reorder the data. Giving an initial value of the temperature control parameter τ_l for $l = 0$, repeat M times the following steps:

1. Select the precipitation amount for month j, $Z_j > 0$, at random.
2. Select two indices i_1 and i_2 from the set $1, 2, \ldots, n_j$ at random such that $z_j(i_1)$ and $z_j(i_2)$ are not both equal to zero.
3. Calculate the value of an objective function for the case the values of $z_j(i_1)$ and $z_j(i_2)$ are swaped (O_s) and for the case the values of $z_j(i_1)$ and $z_j(i_2)$ are unswaped (O_u).
4. If $O_s \leq O_u$ the values of $z_j(i_1)$ and $z_j(i_2)$ are swaped since the objective function decreases. If $O_s > O_u$ the values of $z_j(i_1)$ and $z_j(i_2)$ are swaped with probability:

$$P = \exp\left(\frac{O_u - O_s}{\tau_l}\right) \tag{8}$$

The process is repeated from step 2. The lower the value of τ_l, the less likely is the probability of performing a swap that does not decrease the objective function.

After M iterations, index l is increased by 1 and we set $\tau_{l+1} < \tau_l$. Steps 2–4 are again repeated M times. The whole process is repeated until $\tau_j < \tau_{crit}$.

3.2
Definition of the objective function

The limiting distribution of the disaggregated series will converge asymptotically to the distribution of the original series which is described in terms of an objective function that we want to minimize. The objective function can be formulated by including different properties of the rainfall sequences that one wants to preserve, as for example, autocorrelation function, scaling properties, wet and dry duration distributions, etc.

For example, for the autocorrelation function, an objective function can be formulated as the squared difference between a prescribed autocorrelation function $\rho^*(k)$ and a simulated autocorrelation function $\rho(k)$:

$$O_1 = \sum_{k=1}^{K} (\rho^*(k) - \rho(k))^2 \tag{9}$$

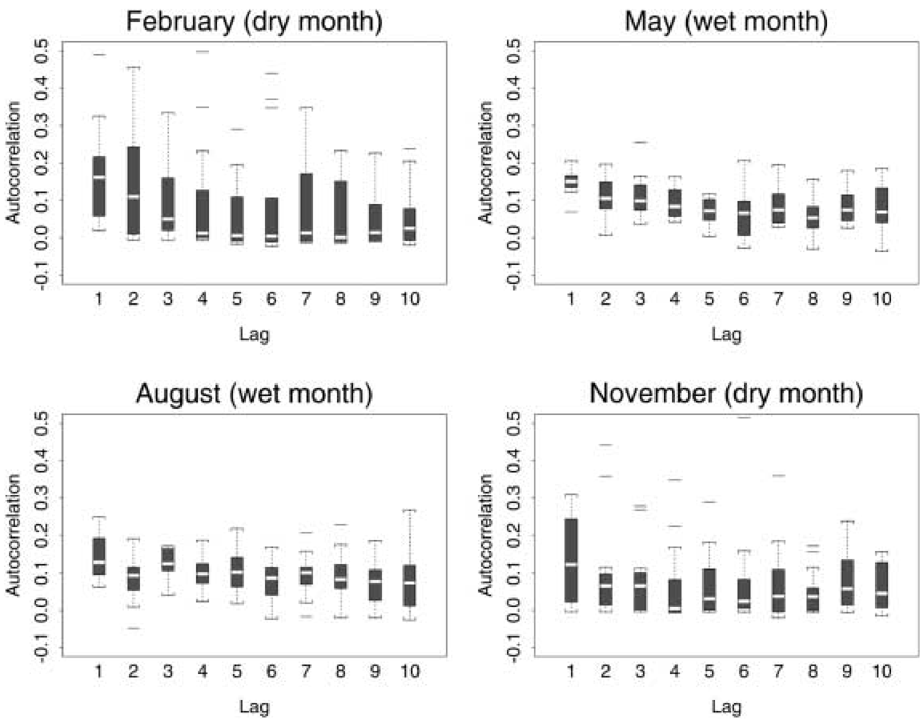

Fig. 5. Autocorrelation function for different months and all locations

where K is the number of lags used in the calculations.

If the serial dependence of the rainfall occurrences wants to be preserved, an objective function could be defined as:

$$O_2 = \sum_{k=1}^{K} (\rho_I^*(k) - \rho_I(k))^2 \tag{10}$$

$\rho_I(k)$ stands for the "indicator autocorrelation function" which is applied to the transformed rainfall data through the indicator function:

$$I(z(t)) = \begin{cases} 1 & \text{if } z(t) > \delta \\ 0 & \text{if } z(t) \le \delta \end{cases} \tag{11}$$

where δ is a rainfall threshold representing a small rainfall amount.

The "wide sense multiple scaling" property described by Burlando and Rosso (1996) usually holds for the raw moments such that:

$$E[z^l(\lambda h)] = \lambda^{\psi_l} E[z^l(h)] \tag{12}$$

where l is the order of the moment, $E[.]$ is the expected value, $z(h)$ is the accumulated rainfall at duration h and $z(\lambda h)$ is the accumulated rainfall at duration λh and ψ_l is the exponent of the scaling relationship.

An objective function could be formulated in order to preserve the scaling properties of rainfall as follows:

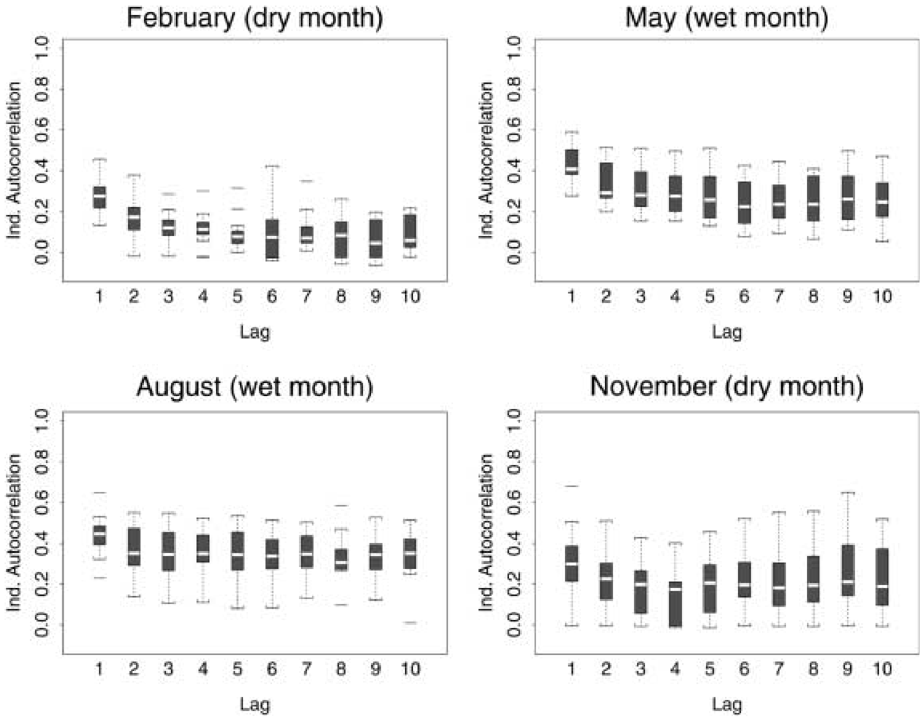

Fig. 6. Indicator autocorrelation function for different months and all locations

$$O_3 = \sum_{\lambda\in\Lambda} \sum_{l=1}^{L} (\psi_l^* - \psi_l)^2 \tag{13}$$

where L is the number of moment orders considered and ψ_l^* and ψ_l are the prescribed and simulated scaling exponents respectively.

All these objectives functions could be combined into a single objective function as:

$$O = \gamma_1 O_1 + \gamma_2 O_2 + \gamma_3 O_3 \tag{14}$$

where the γ_i's are positive weights. They can be arbitrary (e.g. $\gamma_i = 1/3$ for $i = 1, 2, 3$) since different realizations will be generated such that O converges to zero. Other aspects of the precipitation series could be incorporated into the objective function as the distribution of the durations of wet and dry periods, different distributions for varying atmospheric conditions as El Niño or La Niña effects or different circulation patterns as considered by Bárdossy and Plate, 1992.

4
Application of the methodology

The different daily rainfall properties used in the annealing algorithm were calculated for each month and each location in order to investigate their seasonal and spatial variability. Twenty five years of daily data (1967 until 1991) were used to estimate these properties. Boxplots of the autocorrelation and indicator autocorrelation functions for four different months and all locations are shown in Figs. 5 and 6. The horizontal white line within each box represents the median value for all locations at each lag, and the vertical extent of each box gives a measure of the dispersion of the function values among the different locations.

It is evident from the above figures that differences among locations and months are visible. The autocorrelation function is more variable during the dry months than during the wet months. One can expect, for example, a stronger serial dependence in the rainfall occurrence process for a wet month than for a dry month. In fact, the lag-one indicator autocorrelation has a mean value of 0.4 in August (wet month) and a mean value of 0.2 in February (dry month) (see Fig. 8). A higher autocorrelation in the occurrences than in the amounts it is observed from these plots. A lot of scatter is however present in both functions. It is important to mention that rainfall in the region is mainly of convective type and rainfall amounts do not tend to be highly autocorrelated at a daily time scale.

The scaling exponents of Eq. (12) are shown in Fig. 7, where two characteristics can be observed: Firstly, the exponents vary with time of the year being higher during the rainy season than during the dry season. Secondly, a higher dispersion is observed during the dry season, which is attributed to the large number of zeroes in the dry months.

In Fig. 8 the relationships between different aggregation time periods and their corresponding moments in a log-log scale are shown for March and August at station 1. The relationships are presented for the first three moments and the slopes of the linear regressions are fitted from the data. These slopes are the scaling exponents of Eq. (12).

The previous characteristics of autocorrelation and scaling properties are used as parameters to the disaggregation model and the annealing process is carried out by specifying the objective function described in 14 for each month and each location.

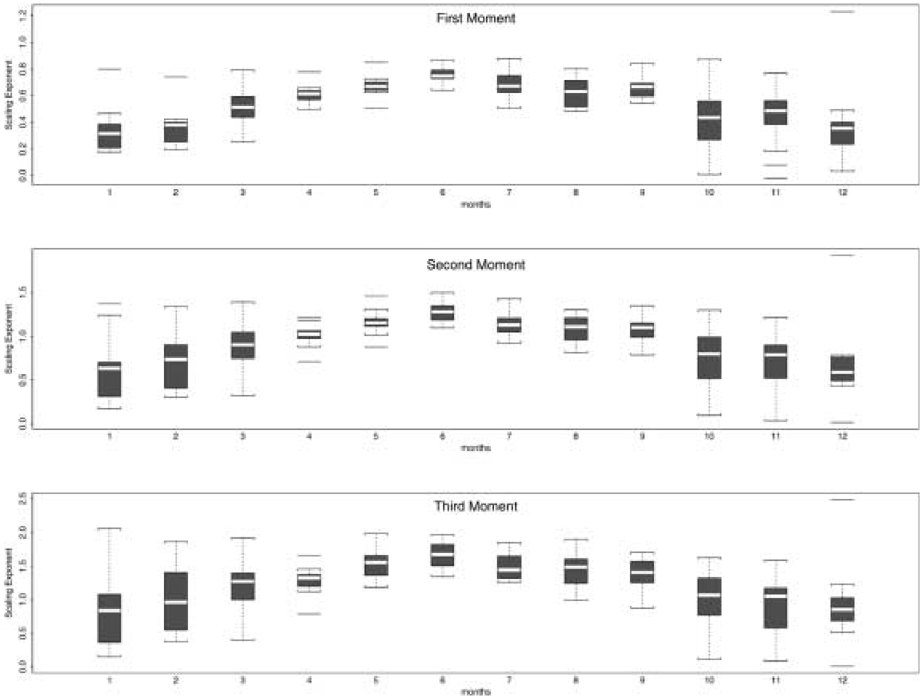

Fig. 7. Scaling exponents for the first, second and third moments for all locations

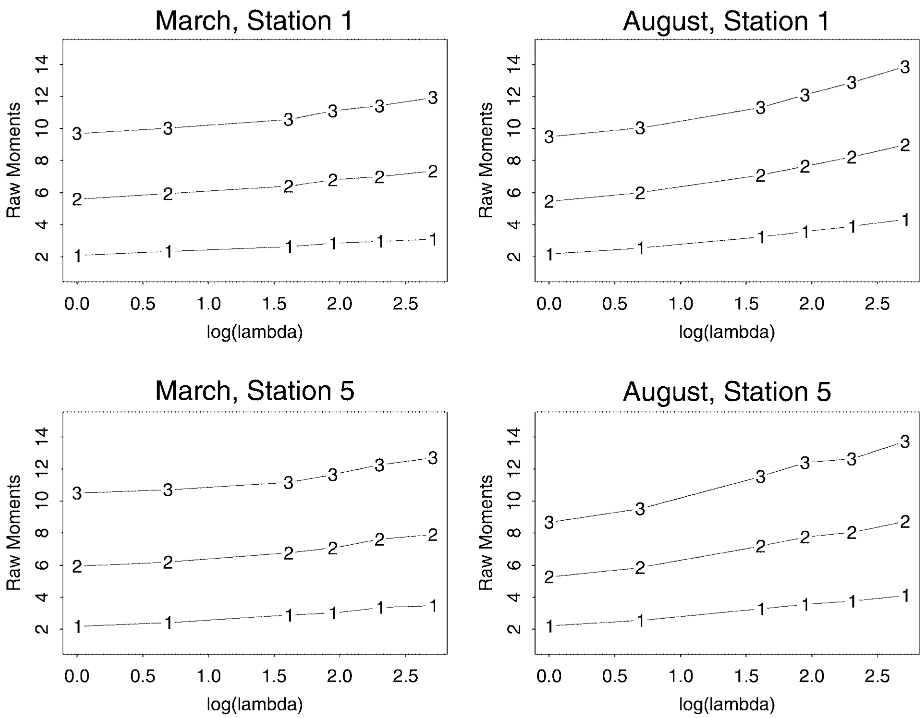

Fig. 8. Log–log relationships between different aggregation periods and the rainfall moments for March and August at locations 1 and 5. The numbers represent the order of the moment

For a given temperature value, a maximum of 5000 swaps of the rainfall amounts are intended in order to decrease the objective function. To save some computational time, each evaluation is only performed for the part of the objective function which is affected by the swaped values. Then the temperature is decreased until a critical value is reached. As an example, in Fig. 9 it is observed how the objective function and the number of swaps decrease with temperature when we disaggregate 22 years of August rainfall at location 1. Similar behaviour was observed for different monthly values and different locations, which is an indication of the convergence of the methodology.

An infinite number of realizations with different number of wet days can be generated by sampling from the corresponding relationship between the probability of wet days and the monthly rainfall amount as the ones presented in Fig. 4. This would allow an estimation of, for example, the 5%, 50% and 95% quantiles of all possible realizations which provides a measurement of the uncertainty in the simulated series.

In Figs. 10 and 11 a comparison between the observed daily values and one possible realization of the disaggregated rainfall series is presented for locations 1 and 5. In all cases the years with missing values have been eliminated and consequently at each location, a different number of annual replications for each month have been disaggregated.

In Fig. 12 a quantile-quantile plot is presented to compare the probability distributions between the observed values and the simulated values for a dry month (March) and a wet month (August) at location 1.

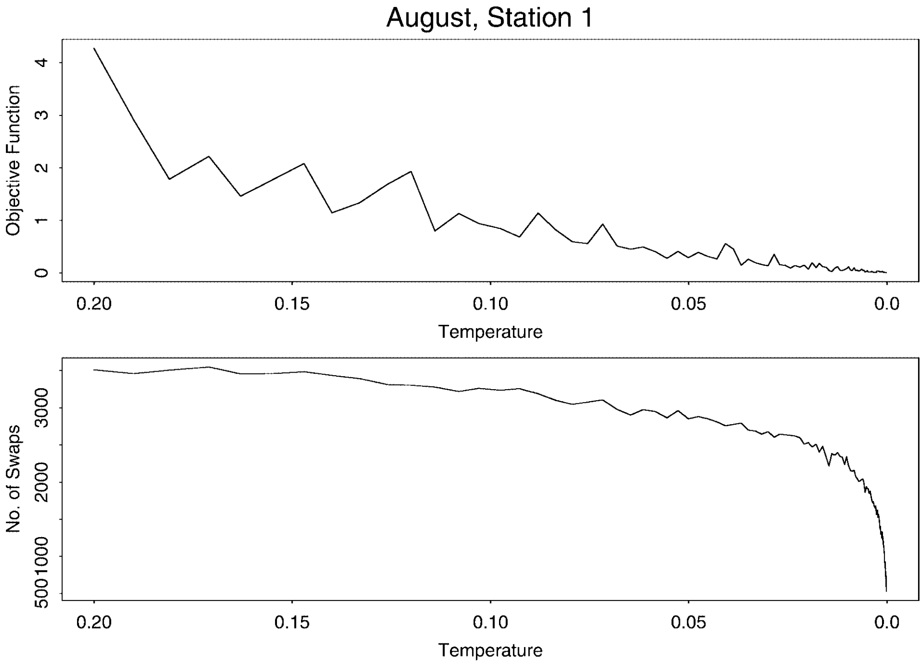

Fig. 9. Objective function vs. temperature and no. of swaps vs. temperature for August at location 1

The quantiles for the simulated and observed series show a good agreement in general. However, other characteristics of the rainfall series might be of interest, as for example, the cumulative probability distribution of consecutive wet and dry day durations. Such distributions were estimated and compared for the observed and simulated values for location 1 in August. From Fig. 13 it is observed that dry durations probabilities tend to be overestimated by the simulated values. Since this rainfall property was not included in the optimization procedure, it will not necessarily be preserved by the simulated data.

5
Discussion and Conclusions

The proposed methodology has certain advantages and limitations, both of which are enumerated as follows:

1. The probability density function used as a reference distribution of the disaggregated series is very simple and its parameters can be initialized from readily available observations as the number of wet days and the monthly rainfall amounts. This is an advantage over other disaggregation models in where more complex distributional assumptions have been made, since these two quantities are already available as interpolated monthly fields from the period 1990 until 1995 at a global scale (New et al., 2000).

2. Recent methodologies for rainfall disaggregation deal with the problem of daily, hourly and sub-hourly time scales (Connolly et al., 1998; Burian et al., 2000; Gyasi-Agyei, 1999). The proposed methodology could also be used at shorter time scales, since the truncated normal model can be easily calibrated across different scales and the general method of using simulated annealing to preserve prescribed characteristics of the rainfall data could also be extended to other scales.

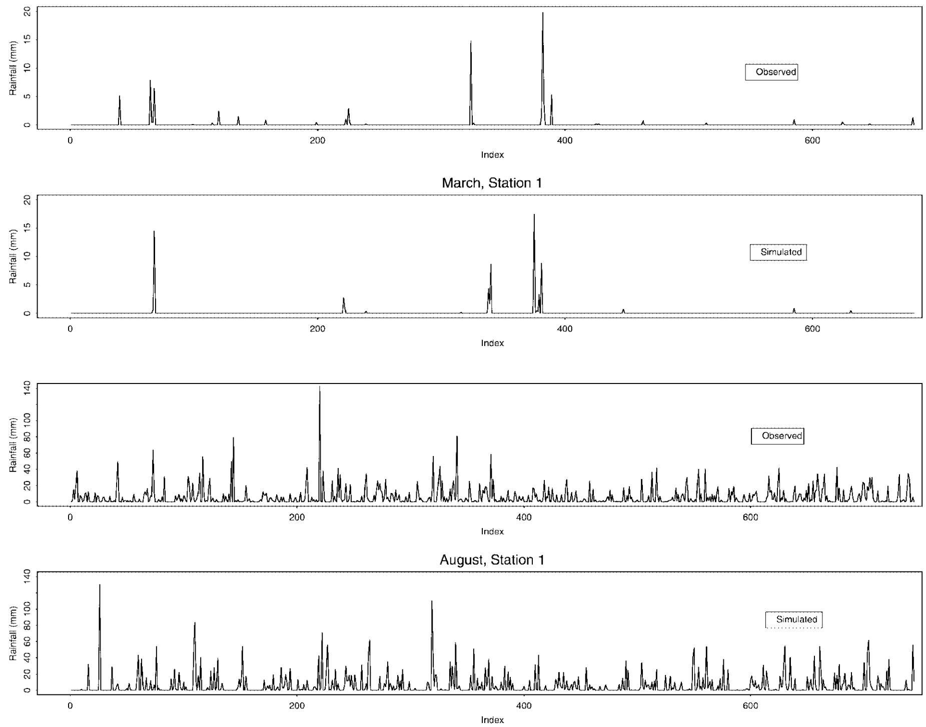

Fig. 10. Observed and disaggregated values for March and August at location 1

3. Different rainfall attributes and occurrence rainfall dependence on the aggregated amounts can be incorporated into the process in a parametric or non-parametric form. In the parametric case the distributional characteristics of the rainfall properties could be used in the definition of the objective function to consider the inherent uncertainty in the estimated values. Another possibility is by using a re-sampling procedure by which the rainfall properties being considered in the objective function are estimated from a sub-sample of the observations at each simulation step.

4. A relatively small number of attributes of the disaggregated series needs to be used to simulate a time series at a finer time scale. This is an advantage specially when considering the application of this methodology to multiple sites. However, important properties of the observed time series as the probability distributions of consecutive dry days, were not accurately reproduced by the disaggregation model. One way to overcome this problem is by including this property as part of the objective function to be minimized in the simulated annealing algorithm, at the expense of more computational complexity and a larger number of parameters.

5. The proposed methodology is computationally very intensive and care has been taken in the re-calculation of the objective function after each swap in the daily values, by modifying only the part of the function affected by the changed values. By increasing the number of attributes we would increase considerably the number of parameter estimates for the disaggregation procedure which would make the algorithm difficult to apply in regions where point daily information is not available. However, the flexibility of the methodology relies precisely in the

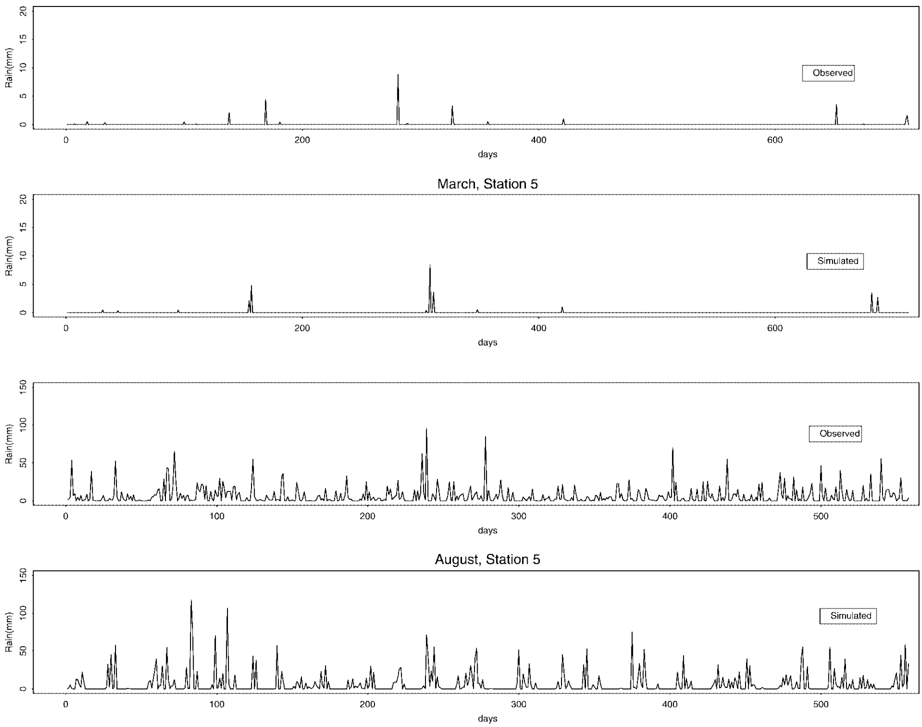

Fig. 11. Observed and disaggregated values for March and August at location 5

simplicity about the rainfall distribution assumption and the different choices about the rainfall properties that wished to be preserved in the simulated series.

6. Most of the analyzed properties show an annual cycle which is consistent with the seasonal rainfall behaviour. This cycle is taken into account into the modelling process by changing the rainfall properties with time of the year. It is clear for example, that the scaling exponents show a strong seasonality. Instead of using different monthly values, the annual cycle can be described in a parametric or non-parametric form by using a Fourier or a spline representation respectively. The parameter uncertainty can be incorporated with the methodologies mentioned above.

7. With a larger number of locations it is possible to study the spatial variability of the selected properties in more detail. This would require a more systematic analysis of the suggested rainfall properties. The analysed region does not present strong contrasts in precipitation regimes and can be considered as a fairly homogeneous region. However differences are observed from location to location and they could be modeled by regionalizing the model parameters with standard interpolation techniques. This approach would be very difficult to implement within the context of other methodologies such as the disaggregation techniques proposed by Burian et al. (2000) in where artificial neural networks are used as a disaggregation tool. A final goal would be to use this methodology at a global scale, by appliying this technique to the already existing global monthly data sets as the one provided by New et al. (2000).

8. For the analysed daily data set, it was found that accumulated rainfall over few days can usually be present at any day within the month, due to observational problems in the rainfall network and the impossibility of recording rainfall

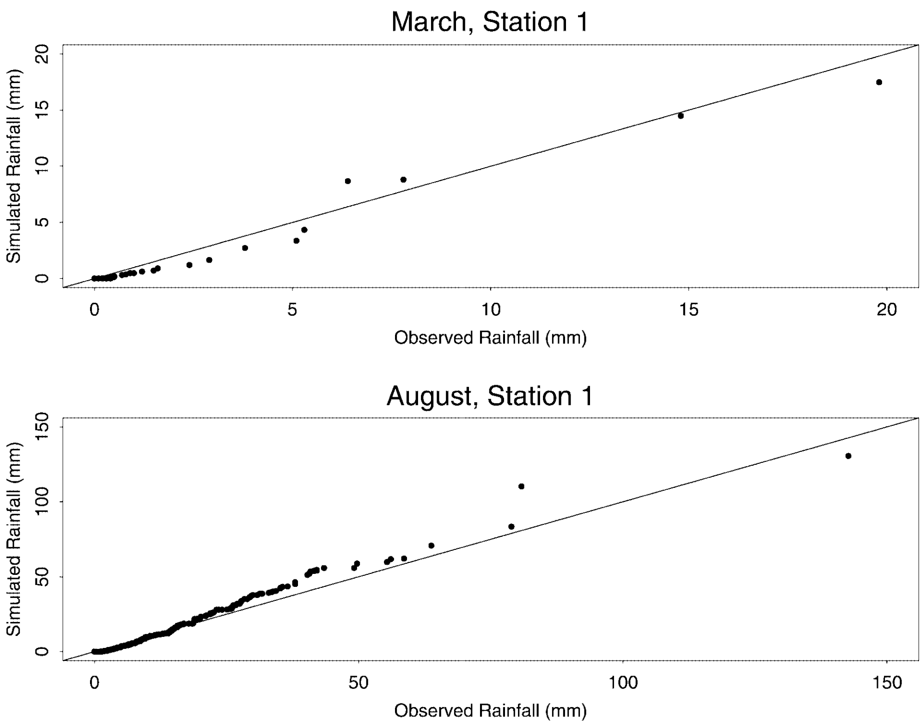

Fig. 12. Quantile–quantile plots comparing observed and disaggregated probability distributions for March and August at location 1

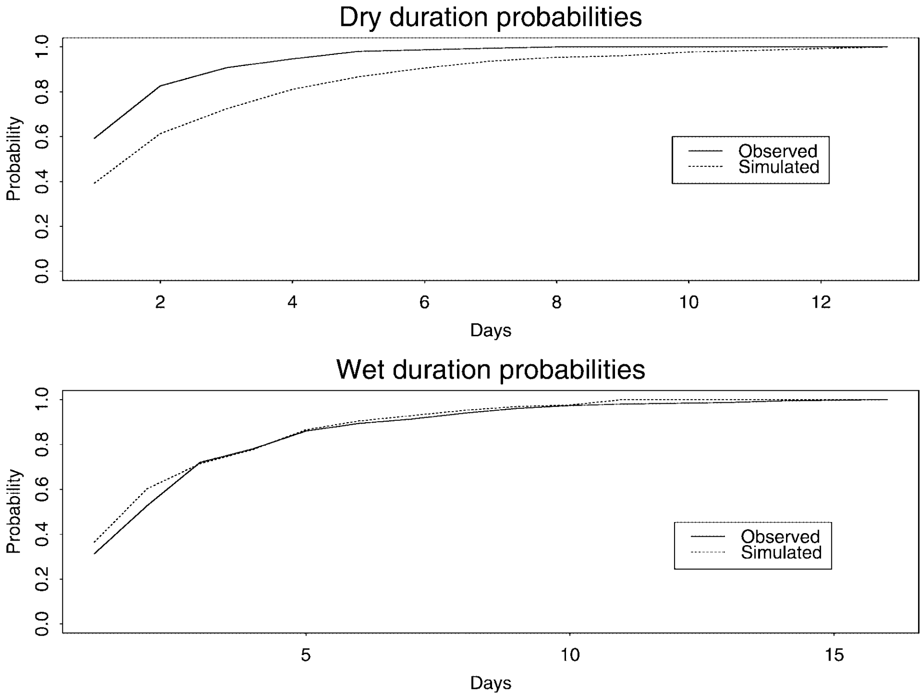

Fig. 13. Wet and dry duration probability distributions for August at location 1

amounts over successive days. Although the methodology has not yet been tested for this particular problem, it is feasible to apply this method for an aggregation period shorter than a month with arbitrary duration. This extension of the problem would provide a complete series of daily values making a extremely useful tool to deal with this kind of data quality problems.

References

Bárdossy A (1998) Generating precipitation time series using simulation annealing. Water Resour. Res. 34: 1737–1744

Bárdossy A, Plate EJ (1992) Space–time model for daily rainfall using atmospheric circulation patterns. Water Resour. Res. 28: 1247–1259

Brass RL, Rodriguez-Iturbe I (1985) Random Function in Hydrology. Addison-Wesley, Reading, Mass

Bürger G (1997) On the disaggregation of climatological means and anomalies. Climate Res. 8: 183–194

Burian SJ, Durrans SR, Tomić S, Pimmel RL, Wai CN (2000) Rainfall disaggregation using artificial neural networks. J. Hydrologic Eng. 5(3): 299–307

Burlando P, Rosso R (1996) Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. J. Hydrol. Eng. 187: 45–64

Chib S, Greenberg E (1995) Understanding the Metropoils-Hastings algorithm. The American Statistician 49: 327–335

Connolly R, Schirmer J, Dunn P (1998) A daily rainfall disaggregation model. J. Agric. For. Meteorol. 92: 105–117

Curry K, Bras RL (1978) Theory and application of the multivariate broken line, disaggregation and monthly autoregressive streamflow generators for the nile river. Technical Report 78-5, Ralph M. Parsons Lab., Mass. Inst. of Technol., Cambridge

Glasbey CA, Cooper G, McGechan MB (1995) Disaggregation of daily rainfall by condition simulation. J. Hydrol. 165: 1–9

Grygier JC, Stedinger JR (1988) Condensed disaggregation procedures and conservation corrections for stochastic hydrology. Water Resour. Res. 24: 1574–1584

Guenni L, Bárdossy A, Haberlandt U (1997) Spatial interpolation of rainfall data at different time scale. Technical Report 97-6, Universidad Simón Bolívar, CESMa

Gyasi-Agyei Y (1999) Identification of regional parameters of a stochastic model for rainfall disaggregation. J. Hydrol. 223: 148–163

Henze N, Klar B (1993) Goodness-of-fit testing for a space–time model for daily rainfall. Technical report, University of Karlsruhe. Germany

Hernandez A (1998) Parametrización, Estimación e Inferencia del Modelo Normal Truncado. Universidad Simón Bolivar, Master Thesis, Caracas, Venezuela

Hershenborn J, Woolhiser DA (1992) Disaggregation of daily rainfall. J. Hydrol. 95: 299–322

Hutchinson MF (2001) Calibration of the truncated power of normal distribution for precipitation data. In preparation

Hutchinson MF, Richardson CW, Dyke PT (1993) Normalization of rainfall across different time steps. Management of Irrigation and Drainage Systems, pp. 432–439

Lebel T, Braud I, Creutin JD (1998) A space–time rainfall disaggregation model adapted to sahelian mesoscale convective complexes. Water Resour. Res. 34: 1711–1726

Maheepala S, Perera B (1996) Monthly hydrologic data generation by disaggregation. J. Hydrol. 178: 277–291

Matheron G (1982) The intrinsic random functions and their applications. Adv. Appl. Probab. 5: 1379–1394

New M, Humes M, Jones PD (2000) Representing twentieth century space–time climate variability. ii: development of a 1901–1996 monthly terrestrial climate fields. J. Climate 13: 2217–2238

Sansó B, Guenni L (1998) A stochastic model for tropical rainfall at a single location. J. Hydrol. 214: 64–73

Sansó B, Guenni L (1999) Venezuelan rainfall data analysed by using a Bayesian space–time model. Appl. Statist. 48(3): 345–362

Santos EG, Salas JD (1992) Stepwise disaggregation scheme for synthetic hydrology. J. Hydraul. Eng. 118: 765–1784

Stidd CK (1973) Estimating the precipitation climate. Water Resour. Res. 9: 1235–1241
Tarboton DG, Sharma A, Lall U (1998) Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. Water Resour. Res. 34: 107–119
Trofs P (1997) Disaggregation in hydrology. Technical report, Agriculture University Wageningen
Valencia DR, Schaake JL (1972) A disaggregation model for time series analysis and synthesis. Technical Report 149, Ralph M. Parsons Lab., Mass. Inst. of Technol., Cambridge
Vörösmarty C, Federer C, Schloss A (1998) Potential evaporation functions compared on us watersdes: possible implications for global-scale water balance and terrestrial ecosystem. J. Hydrol. 207: 147–169