Multivariate rainfall disaggregation at a fine time scale

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering

National Technical University, Athens

Heroon Polytechneiou 5, GR-157 80 Zographou, Greece

(dk@hydro.ntua.gr)

Christian Onof and Howard S. Wheater

Department of Civil and Environmental Engineering

Imperial College of Science, Technology and Medicine

London SW7 2BU, UK

Paper submitted to Water Resources Research

First submission: June 2000

Second submission: June 2002

Revision: February 2003

Abstract. A methodology for spatial-temporal disaggregation of rainfall is proposed. The methodology involves the combination of several univariate and multivariate rainfall models operating at different time scales, in a disaggregation framework that can appropriately modify outputs of finer time scale models so as to become consistent with given coarser time scale series. Potential hydrologic applications include enhancement of historical data series and generation of simulated data series. Specifically, the methodology can be applied to derive spatially consistent hourly rainfall series in raingages where only daily data are available. In addition, in a simulation framework, the methodology provides a way to take simulations of multivariate daily rainfall (incorporating spatial and temporal non-stationarity) and generate multivariate fields at fine temporal resolution. The methodology is tested via a case study dealing with the disaggregation of daily historical data of five raingages into hourly series. Comparisons show that the methodology results in good preservation of important properties of the hourly rainfall process such as marginal moments, temporal and spatial correlations, and proportions and lengths of dry intervals, and in addition, a good reproduction of the actual hyetographs.

1 Introduction

A common problem in hydrological studies is the limited availability of data at appropriately fine temporal and/or spatial resolution. In addition, in hydrologic simulation studies a model may provide as output a synthetic series of a process (such as rainfall and runoff) at a coarse scale while another model may require as input a series of the same process at a finer scale. Disaggregation techniques therefore have considerable appeal due to their ability to increase the time or space resolution of hydrologic processes while simultaneously providing a multiple scale preservation of the stochastic structure of hydrologic processes.

The linear disaggregation model introduced by *Valencia and Schaake* [1972, 1973] along with the contributions of *Mejia and Rousselle* [1976], *Tao and Delleur* [1976], *Hoshi and Burges* [1979], *Lane* [1979, 1982], *Salas et al*. [1980], *Todini* [1980], *Stedinger and Vogel* [1984], *Pereira et al*. [1984], *Stedinger et al.* [1985], *Oliveira et al.* [1988], *Grygier and Stedinger* [1988, 1990], *Lane and Frevert* [1990], *Santos and Salas* [1992], and *Salas* [1993, p. 19.34] has been the most important and widely used scheme for stochastic disaggregation problems in hydrological applications. Such linear modeling schemes have been used, among other applications, for univariate or multivariate disaggregation of annual to monthly rainfall. However, as first pointed out by *Valencia and Schaake* [1972], modeling schemes of this kind are not suitable for the disaggregation of rainfall for time scales finer than monthly, due to the skewed distributions and the intermittent nature of the rainfall process at fine time scales. Other disaggregation models have been proposed and used, particularly for the disaggregation of rainfall, but do not exhibit the generality of these linear schemes.

The particular problem of rainfall disaggregation at a fine time scale was first studied in a systematic manner by *Woolhiser and Osborn* [1985]. They presented a scheme for the disaggregation of individual storm depths into shorter periods, each corresponding to one tenth of the storm's duration; their scheme was based on a nondimensionalized Markov process, resulting from successive transformations of the actual rainfall process.

Since then, several studies (for example the works of *Marien and Vandewiele* [1986], *Hershenhorn and Woolhiser* [1987], *Koutsoyiannis and Xanthopoulos* [1990], *Koutsoyiannis and Foufoula-Georgiou* [1993], *Koutsoyiannis* [1994], *Glasbey et al.* [1995], *Olsson and Berdtsson* [1997], *Connolly et al.* [1998], *Gyasi-Agyei* [1999], and *Koutsoyiannis and Onof* [2000, 2001]) have been conducted on the problem of fine time scale rainfall disaggregation. Some of the methods aimed at the disaggregation of daily rainfall into hourly or finer time scale. A common characteristic of all is their univariate aspect, as they perform temporal disaggregation at one location only. *Koutsoyiannis* [1992] generalized the method by *Koutsoyiannis and Xanthopoulos* [1990] for multiple dimensions but this generalization was never used in fine-time-scale rainfall disaggregation.

The problem of multiple site rainfall disaggregation, as a means for simultaneous spatial and temporal disaggregation, has not been studied so far at a fine time scale. It presents significant differences from that of single-site disaggregation, including increased mathematical complexity. The spatial correlation (cross-correlation among different sites) must be maintained in the multivariate problem, whereas it does not appear at all in univariate problems.

The multivariate approach to rainfall disaggregation is of significant practical interest even in problems that are traditionally regarded as univariate. Let us consider, for instance, the disaggregation of historical daily raingage data into hourly rainfall. This is a common situation since detailed hydrological models often require inputs at the hourly time scale. However, historical hourly records are not as widely available as daily records. An appropriate univariate disaggregation model would generate a synthetic hourly series, fully consistent with the known daily series and, simultaneously, statistically consistent with the actual hourly rainfall series. Obviously, however, a synthetic series obtained by such a disaggregation model could not coincide with the actual one, but would be only a likely realization. Now, let us assume that there exist hourly rainfall data at a neighboring raingage. If this is the case and, in addition, the cross-correlation among the two raingages is significant (a case met very frequently in practice), then we could utilize the available hourly rainfall information at the neighboring station to generate spatially and temporally consistent hourly rainfall series at the raingage of interest. In other words, the spatial correlation is turned to advantage since, in combination with the available single-site hourly rainfall information, it enables more realistic generation of the synthesized hyetographs. Thus, for example, the location of a rainfall event within a day and the maximum intensity would not be arbitrary, as in the case of univariate disaggregation, but resemble their actual values.

This can be considered as a particular case of a general multivariate spatial-temporal rainfall disaggregation problem. This general problem, the simultaneous rainfall disaggregation at several sites, is the subject of this paper. The problem involves the combination of several univariate and multivariate rainfall models operating at different time scales. The problem is formulated in section 2 and the modeling approach is described in section 3. A detailed case study is given in section 4 and conclusions are drawn in section 5. Some mathematical derivations and additional information on a spatial-temporal cluster based rainfall model that is used here to infer parameters for the proposed modeling approach are contained in two appendices.

2 Problem formulation

We standardize the problem studied in this paper in the following manner. We assume that we are given:

1. an hourly point rainfall series at point 1, as a result of either:
   - measurement by an autographic device (pluviograph) or digital sensor,

- simulation with a fine time scale point rainfall model such as a point process model, or
- simulation with a temporal point rainfall disaggregation model applied to a series of known daily rainfall; and

2. several daily point rainfall series at neighboring points 2, 3, 4, 5, … as a result of either:
   - measurement by conventional raingages (pluviometers with daily observations), or
   - simulation with a multivariate daily rainfall model.

We wish to produce series of hourly rainfall at points 2, 3, 4, 5, …, so that:

1. their daily totals equal the given daily values; and
2. their stochastic structure resembles that implied by the available historical data.

We emphasize that in this problem formulation we always have an hourly rainfall series at one location, which guides the generation of hourly rainfall series at other locations. If this hourly series is not available from measurements, it can be generated using appropriate univariate simulation models (see section 3.1). If hourly rainfall is available at several (more than one) locations, the same modeling strategy described below can be used without any difficulty with some generalizations of the computational algorithm. In fact, having more than one point with known hourly information would be advantageous for two reasons. First, it allows a more accurate estimation of the spatial correlation of hourly rainfall depths (see discussion below) or their transformations (see section 3.6). Second, it might reduce the residual variance of the rainfall process at each site, thus allowing for generated hyetographs closer to the real ones.

The essential statistics that we wish to preserve in the generated hourly series are:

1. the means, variances and coefficients of skewness;
2. the temporal correlation structure (autocorrelations);
3. the spatial correlation structure (lag zero cross-correlations); and

4. the proportions of dry intervals.

If the hourly data set at location 1 is available from measurement, then all these statistics apart from the cross-correlation coefficients can be estimated at the hourly time scale using this hourly record. To transfer these parameters to other locations, spatial stationarity of the process can be assumed. The stationarity hypothesis may seem an oversimplification at first glance. However, it is not a problem in practice since possible spatial nonstationarities manifest themselves in the available daily series; thus the final hourly series, which are forced to respect the observed daily totals, will reflect these nonstationarities.

If more than one rainfall series are available at the hourly level, at least one cross-correlation coefficient of hourly rainfall can be estimated directly from these series. Then, by making plausible assumptions about the spatial dependence of the rainfall field an expression of the cross-correlation versus distance could be established. This would then be used to estimate cross-correlation between all pairs of raingages.

Otherwise, i.e., in those cases for which only one or no historical data set exists at the locations of interest, some or all the above categories of statistics must be inferred indirectly using a spatial-temporal stochastic model of the rainfall structure fitted on daily rainfall series, as described later (section 3.1). Obviously, in the latter case the accuracy of statistics at hourly level will be expected to be much poorer.

3 Modeling approach

3.1 Models involved

Several separate models are involved in the proposed disaggregation framework. These fall into three categories outlined below.

a. Models for the generation of multivariate fine-scale outputs. The first category includes two models that form the core of this framework in the sense that they provide the required output (the hourly series).

The first model is a simplified multivariate model of hourly rainfall that can preserve the statistics of the multivariate rainfall process and, simultaneously, incorporate the available hourly information at site 1, without any reference to the known daily totals at the other sites. The statistics considered here are the means, variances and coefficients of skewness, the lag-one autocorrelation coefficients and the lag-zero cross-correlation coefficients. All these represent statistical moments of the multivariate process. The proportion of dry intervals, although considered as one of the parameters to be preserved (section 2), is difficult to incorporate explicitly. However, it can be treated by an indirect manner, as discussed later (sections 3.3 and 3.6).

The second model is a transformation model that modifies the series generated by the first model, so that the daily totals are equal to the given ones. This uses a (multivariate) transformation, which does not affect the stochastic properties of the series. Both models are discussed further below (sections 3.4 and 3.5).

b. Models associated with inputs to a. above. The second category contains models which may optionally be used to provide the required input, should no observed series be available. These may include

- a multivariate daily rainfall model for providing daily rainfall depths, such as the general linear model (GLM) [*Chandler and Wheater*, 1998a, b];
- a single-site model for providing hourly depths at one location such as the Bartlett-Lewis rectangular pulses model [*Rodriguez-Iturbe et al.*, 1987, 1988; *Onof and Wheater*, 1993, 1994];
- a single-site disaggregation model to disaggregate daily depths of one location into hourly depths [e.g. *Koutsoyiannis and Onof*, 2000, 2001].

Such models may be appropriate to operate the proposed disaggregation approach for future climate scenarios.

c. Models associated with spatial-temporal parameters. The third category includes models which play an auxiliary role in the disaggregation framework by providing some of the required parameters of the spatial-temporal rainfall process given the statistical properties of the available data. For example, in this paper, we have adopted the Gaussian Displacement Spatial-Temporal rainfall Model (GDSTM) [*Northrop*, 1996, 1998] to provide the spatial stochastic structure at the hourly level. This model assumes that rainfall is realized as a sequence of storms that arrive following a Poisson process in space and time and each storm consists of a number of cells. Both storms and cells are characterized by their centers, durations and areal extents, and in addition cells have uniform rainfall intensity. Details of the model structure are given in Appendix 2

As explained above, if some statistics of hourly rainfall required for the disaggregation cannot be estimated directly from the data, they might be inferred using the spatial-temporal rainfall model in the following manner:

1. The temporal and spatial correlations at the daily level are estimated using the daily data sets; in addition, if an hourly series is available, its marginal statistics are estimated.
2. The parameters of the spatial-temporal rainfall model are estimated by fitting the spatial-temporal rainfall model using the historical statistics estimated at point 1.
3. The remaining statistics that are required for disaggregation (e.g. spatial correlations at the hourly level) are inferred from the spatial-temporal rainfall model.

This is further clarified in the case study of section 4. It should be noted that the spatial-temporal rainfall model is used in parameter estimation only. To estimate the required statistics in the manner described above it suffices to calibrate the model and there is no need to run it. However, a model run may be necessary if statistics of transformations of rainfall depths are needed for subsequent modeling steps (see section 3.6). Such statistics can be estimated from synthetic hourly series generated by the spatial-temporal rainfall model.

3.2 The need for a simplified rainfall model

Among the different model categories described in section 3.1, the GDSTM (or another model of this type) is the most complete as it describes the rainfall process in continuous time and space in a multidimensional context. However, such a model cannot have the appropriate structure to utilize the given hourly information at site 1 and therefore it cannot be used in the generation phase. In other words, it is not possible to force (or condition) such a multivariate model to produce a given hourly rainfall depth (known for instance from measurement) at a certain point and a certain hour. For this reason, GDSTM is only used in the parameter estimation phase.

In generation phase, we must replace this model with a model that can utilize the known hourly time series explicitly. As already mentioned above, the strategy adopted in this paper is to develop a simplified multivariate model for this purpose. This model is capable of preserving only the statistics already listed in section 2. In principle, the adoption of a less simplified model, which could preserve additional statistics of the rainfall process, is not excluded. For example, it may be possible to design the multivariate model so as to maintain a large number of autocovariance coefficients (for any lags). However, we prefer to use a more parameter parsimonious approach, maintaining the autocovariances for lag one only to provide an efficient model which can be indirectly corrected. Normally, this simplification would have some implications for the structure of the multivariate hourly rainfall process. However, it is expected that the implications will be counterbalanced due to the additional source of information (apart from the statistics themselves), which is incorporated as model input: namely, the single site hourly information (see also section 3.6).

Specifically, assuming that all sites are close to each other and highly spatially correlated, the given hourly series at site 1 can be used, with the simplified multivariate model, to

- “guide” the generation of the hourly series at the daily data sites and act indirectly to preserve properties not modeled explicitly;

- properly locate the rainfall events in time; and
- produce initial hourly rainfall series at the daily sites, whose departures from the actual hourly depths at those sites are not large (even though these known daily totals are not considered at all by this simplified multivariate model).

At a later stage, i.e. when the transformation model is applied, another source of information is additionally incorporated, that is the multi-site daily information. This results in the preservation of other additional properties, which are not captured by the statistics used. For example, as noted above, nonstationarities of the rainfall field (both in space and time) are reproducible, even though the models used are stationary.

3.3 A simplified univariate rainfall model

Before describing the simplified multivariate rainfall model, it will be informative to study it in a univariate form. It is well known that specific peculiarities of the rainfall process on a fine time scale, and mainly the intermittency and the highly skewed, J-shaped, probability density function, deterred from the use of typical stochastic models such as ARMA. Different types of models such as point process models [e.g. *Waymire and Gupta*, 1981; *Rodriguez-Iturbe et al.*, 1987, 1988; *Onof and Wheater*, 1993, 1994], scaling models [e.g. *Koutsoyiannis and Foufoula-Georgiou*, 1993], possibly combined with alternate renewal processes [e.g. as in *Koutsoyiannis and Pachakis*, 1996], and multifractal simulation techniques [e.g. *Marshak et al.*, 1994, *Olsson*, 1996; *Olsson and Berndtsson*, 1997; *Menabde et al.*, 1997] have been used instead. Most of these types of models consider in an explicit manner the two states of the rainfall process, i.e. the dry and the wet state. Nowadays, most of these model types have become widespread in typical applications concerning single site rainfall simulation. Extensions to many dimensions have been developed in a few cases. For example, GDSTM, described earlier, is an extension of a monodimensional point process into three dimensions (time dimension plus two spatial dimensions). But even in one dimension, such models are difficult, if not impossible, to condition, e.g. to force to match a certain known value at a certain time interval. Therefore, they are not appropriate as generation tools for the problem under study.

In seeking a simplified model that can be easily conditioned, it is helpful to revisit the stochastic structure of rainfall on a fine time scale, such as hourly. As an example case, we use a 5-year time series of hourly rainfall during rainy days of January at gage 1 of the Brue catchment (South-Western England; see also section 4). This time series refers to 95 rainy days in total, so it contains 2280 data values. As shown in Table 1, most of these values (1679 values or 73.6%) are zeros. The second most frequently occurring value is 0.2 mm (280 values or 12.3%). In total, 2191 values or 96.1% are smaller than or equal to 1 mm. This signifies the highly skewed, J-shaped probability density function of hourly rainfall.

The value x = 0.2 mm is typically the lowest possible measured value in devices like tipping bucket pluviographs. Higher measured values (0.4 mm, 0.6 mm, etc.) are multiples of this threshold. In fact each measured value x_i represents a continuous interval of actual rainfall depth $c_i \leq x < c_{i+1}$. For example, it can be assumed that the measurement outcome x = 0.2 mm represents an actual rainfall depth in the interval 0.1 mm $\leq x <$ 0.3 mm. This means that in this most frequently occurring range of rainfall depth, the measurement (relative) error is very high: from –100% to +33%. The error becomes progressively lower for higher values of x.

Extending this discussion, we must recognize that there is a significant percentage of small rainfall depths in the interval $0 < x <$ 0.1 mm which are measured as x = 0, although they are positive values. This fact is known to people involved in rainfall measurement (who at times observe rainfall but see their devices recording zero) and also to researchers that have analyzed the rainfall structure. For example, it has been a common practice, when forming sequences of storm events, to incorporate periods of zero rainfall within events. The lengths of such periods range from 2-3 hours [*Grace and Eagleson*, 1966; *Eagleson*, 1970] through 5-

7 hours [*Huff*, 1967; *Koutsoyiannis and Xanthopoulos*, 1990, *Koutsoyiannis and Foufoula-Georgiou*, 1993] to as high as 3 days or more [*Restrepo-Posada and Eagleson*, 1982].

Given this, what is regarded as a dry period (identified from the measurements of a pluviograph) is not necessarily a dry period from a physical point of view. Distinguishing the (false dry) periods with positive rainfall depth in the interval (0, 0.1 mm) from the periods with exactly zero rainfall (true dry) is not an easy task. Intuitively, given the number of occurrences of small values (0.2-1.0 mm) in the example of Table 1, we expect to have a large number, N_0, of rainfall depths in the interval (0, 0.1 mm), at least comparable to N_1, the number of the values 0.2 mm. To get a quantitative estimate of N_0, we can perform extrapolation assuming that the actual rainfall depths vary in the continuous set of positive real numbers, rather than taking the discrete values 0.2 mm, 0.4 mm, etc. Let $F(x)$:= $\text{Prob}(X \leq x | X > 0)$ be the probability distribution function of hourly depth conditional on being positive (not dry). Then, $N_0 = F(c_1)\, N$, $N_1 = [F(c_2) - F(c_1)]\, N$, or more generally

$$N_i = [F(c_{i+1}) - F(c_i)]\, N \tag{1}$$

where $N = N_0 + N_1 + N_2 + \ldots$, $c_0 = 0$, and $c_i = 0.1 \text{ mm} + (i - 1) \times 0.2 \text{ mm}$ (for $i > 0$). To estimate N_0, we tried three alternative types of $F(x)$, the single-parameter exponential and the two-parameter Weibull and Gamma, all having the ability to yield a J-shaped density, as described in more detail in Appendix 1. The best local fit (for rain depths ≤ 1 mm) was attained by the Gamma distribution, which resulted in N_0 as high as 2657, which is higher than the historical total number of zero values (1679). The Weibull distribution resulted in N_0 = 1110 whereas the exponential distribution, which had the worst fit, resulted in N_0 = 291. This exercise shows that the estimation of the number of the very small values of hourly rainfall depths during rainy days cannot be precise. Apparently, however this is a large number, which in our example must be in the range 291 to 1679, with the higher values being more likely. It also shows that if we regard all historical zero hourly depths during rainy days

(1679 values) as very small depths (< 0.1 mm rather than exactly zero), we will not make a vital error. This assumption is in fact very convenient: there is no need to model rainfall during rainy days as a two-state process but it suffices to handle it as a typical stochastic process whose smallest values (< 0.1 mm) are rounded off to zero. One may argue that this approach may be not precisely consistent with the natural rainfall process as there must be some true dry periods during rainy days. The answer is that this is merely a model and no model can be fully consistent with reality. As shown above, even a two-state approach is inconsistent with reality as far as it regards all periods with measurements rounded off to zero as dry periods.

Let us now study the consequences of the proposed simplified univariate single-state model. In Figure 1 we depict the empirical distribution function of hourly rainfall depth during rainy days of the same example time series, now considering all data values and assuming a single-state model (e.g., $F(0.1\ \text{mm}) = N_0 / (N + 1)$, $F(0.3\ \text{mm}) = (N_0 + N_1) / (N + 1)$, etc., with $N_0 = 1679$). To this empirical distribution we fitted the Gamma distribution (now for the entire domain of hourly depth) and also plotted it on Figure 1 (upper panel) along with the corresponding curves of the Kolmogorov-Smirnof test for significance level 10%. The plot indicates that the Gamma distribution is an acceptable marginal distribution for the single-state hourly rainfall process. In particular, the probability of a dry hour, conventionally defined as $F(0.1\ \text{mm})$, is estimated from the Gamma distribution to 74.8%, very close to the empirical value (73.6%).

Given that hourly depths are autocorrelated, a candidate stochastic process to model them is the Gamma autoregressive (GAR) process [*Lawrance and Lewis*, 1981; *Fernandez and Salas*, 1990] which behaves like the typical autoregressive AR(1) process but, in addition, it preserves exactly the Gamma distribution. As demonstrated in the lower panel of Figure 1, a synthetic series generated from the GAR model with length, mean, standard deviation and lag one autocorrelation equal to those of the historical sample (2280, 0.17 mm, 0.48 mm and

0.43, respectively), when rounded off to intervals of 0.2 mm, gives a distribution function plot very close to the historical distribution.

For the same example, we have plotted on Figure 2 the empirical distribution of the lengths of historical “dry” periods, defined as the periods with consecutive zero measurements of hourly depth. In addition, we have plotted similar distributions for two synthetic series of hourly rainfall. In the first series independence of consecutive hourly depths was assumed, whereas the second is the (already mentioned) one generated with the GAR model with lag one autocorrelation 0.43. In the case of independent hourly depths, the resulting lengths are clearly too small in comparison to the historical ones. When the lag one autocorrelation is considered, the distribution of the synthetic dry intervals approaches that of the historical ones with the exception of the very low values, where a departure is observed.

Thus, the simple single-state GAR model, applied on rainy days, could be a satisfactory simplified model for a single-site rainfall model. However, its extension to a multivariate space is not possible, and thus even this model is not appropriate for our purpose. The next candidate is the AR(1) model (i.e. $X_s = a\ X_{s-1} + b\ V_s$) with three-parameter Gamma distributed residuals V_s. If these parameters are estimated so that X_s are two-parameter Gamma distributed, then AR(1) approaches GAR. This is demonstrated in the lower panel of Figure 1, where, in addition to the GAR series, the distribution function of a synthetic series generated from the AR(1) model (again with length, mean, standard deviation and lag one autocorrelation equal to those of the historical sample) was also plotted. In this case the plot was drawn for the series as generated, without rounding off its values. What is not distinguished in the graph is that a high percentage, 50.1% of the generated values, is negative with values in the interval (–0.045 mm, 0). Another percentage 20.1% is small positive values smaller than 0.1 mm. Of course, both these groups of values will be rounded off to zero, so that we will get a probability dry 70.1%. In addition, if we round off values smaller than a threshold 0.13 mm, instead of 0.10 mm, we get exactly the historical probability dry (73.6%).

Thus, the AR(1) model, although theoretically is apparently unsuitable for rainfall (especially for the generation of negative values), in practice behaves like the more consistent GAR model and, in addition, can be directly extended to many dimensions.

The above observations have been based on a single real world example for demonstration purposes. A wide range of similar examples was investigated, referring to hourly rainfall series from different climates, and resulted in similar conclusions.

3.4 The simplified multivariate rainfall model

If, after the discussion of the section 3.3, a univariate single-state AR(1) model seems to behave relatively well in reproducing the stochastic structure of rainfall, a multivariate AR(1) model, in which at least one component is the actual (rather than modeled) rainfall process at one location, will behave even better. Thus, for n locations, we may assume that the simplified multivariate rainfall model is an AR(1) process, expressed by

$$\mathbf{X}_s = \mathbf{a}\,\mathbf{X}_{s-1} + \mathbf{b}\,\mathbf{V}_s \tag{2}$$

where $\mathbf{X}_s := [X_s^1, X_s^2, \ldots, X_s^n]^T$ represents the hourly rainfall at time (hour) s at n locations, $\mathbf{a}$ and $\mathbf{b}$ are $(n \times n)$ matrices of parameters and $\mathbf{V}_s$ ($s = \ldots, 0, 1, 2, \ldots$) is an independent identically distributed (iid) sequence of size n vectors of innovation random variables (so that the innovations are both spatially and temporally independent). The time index s can take any integer value. $\mathbf{X}_s$ are not necessarily standardized to have zero mean and unit standard deviation, and obviously they are not normally distributed. On the contrary, their distributions are very skewed. The distributions of $\mathbf{V}_s$ are assumed three-parameter Gamma.

Equations to estimate the model parameters $\mathbf{a}$ and $\mathbf{b}$ and the moments of $\mathbf{V}_s$ directly from the statistics to be preserved are given for instance by *Koutsoyiannis* [1999] for the most general case. In the special case examined here, for convenience, the parameter matrix $\mathbf{a}$ is assumed diagonal, which suffices to preserve the statistics listed in section 2, and is given by

$$\mathbf{a} = \mathrm{diag}(\mathrm{Cov}[X_s^l, X_{s-1}^l] \,/\, \mathrm{Var}[X_{s-1}^l], l = 1, \ldots, n) \tag{3}$$

The parameter matrix **b** is determined from

$$\mathbf{b}\,\mathbf{b}^T = \mathrm{Cov}[\mathbf{X}_s, \mathbf{X}_s] - \mathbf{a}_s\,\mathrm{Cov}[\mathbf{X}_{s-1}, \mathbf{X}_{s-1}]\,\mathbf{a}_s \tag{4}$$

If **b** is assumed lower triangular, which facilitates handling of the known hourly rainfall at site 1, then it can be determined from $\mathbf{b}\,\mathbf{b}^T$ using Cholesky decomposition.

Another group of model parameters are the moments of the auxiliary variables $\mathbf{V}_s$. The first moments (means) are obtained by

$$E[\mathbf{V}_s] = \mathbf{b}^{-1}\,(\mathbf{I} - \mathbf{a})\,E[\mathbf{X}_s] \tag{5}$$

where **I** is the identity matrix. The variances are by definition 1, i.e., $\mathrm{Var}[\mathbf{V}_s] = [1, \ldots, 1]^T$ and the third moments are obtained in terms of $\mu_3[\mathbf{X}_s]$, the third moments of $\mathbf{X}_s$, by

$$\mu_3[\mathbf{V}_s] = \left(\mathbf{b}^{(3)}\right)^{-1}(\mathbf{I} - \mathbf{a}^{(3)})\,\mu_3[\mathbf{X}_s] \tag{6}$$

where $\mathbf{a}^{(3)}$ and $\mathbf{b}^{(3)}$ denote the matrices whose elements are the cubes of **a** and **b**, respectively.

At the generation phase, V_s^1, the first component of $\mathbf{V}_s$, is calculated from the series of X_s^1 rather than generated. Given that **b** is lower triangular, its first row will have only one nonzero item, call it b^1, so that from (2)

$$X_s^1 = a^1\,X_{s-1}^1 + b^1\,V_s^1 \tag{7}$$

which can be utilized to determine V_s^1. This can be directly expanded to the case where several gages of hourly information are available provided that **b** is lower triangular.

Alternatively, the model can be expressed in terms of some nonlinear transformations $\mathbf{X}_s'$ of the hourly depths $\mathbf{X}_s$ (e.g., a power transformation, see section 3.6), in which case (2) is replaced by

$$\mathbf{X}'_s = \mathbf{a}\,\mathbf{X}'_{s-1} + \mathbf{b}\,\mathbf{V}_s \tag{8}$$

All above equations (3)-(7)are still valid if $\mathbf{X}'_s$ is substituted for $\mathbf{X}_s$.

3.5 The transformation model

Transformations that can modify a series generated by any stochastic process to satisfy some additive property (i.e. the sum of the values of a number of consecutive variables be equal to a given amount), without affecting the first and second order properties of the process, have been studied previously by *Koutsoyiannis* [1994] and *Koutsoyiannis and Manetas* [1996]. These transformations, more commonly known as adjusting procedures, are appropriate for univariate problems, although they can be applied to multivariate problems as well, but in a repetition framework. More recently, *Koutsoyiannis* [2001] has studied a true multivariate transformation of this type and also proposed a generalized framework for coupling stochastic models at different time scales.

This framework, adapted to the problem examined here, is depicted in Figure 3 where $\mathbf{X}_s$ and $\mathbf{Z}_p$ represent the "actual" hourly- and daily-level processes, related by

$$\sum_{s=(p-1)k+1}^{p\,k} \mathbf{X}_s = \mathbf{Z}_p \tag{9}$$

where k is the number of fine-scale time steps within each coarse-scale time step (24 for the current application). $\tilde{\mathbf{X}}_s$ and $\tilde{\mathbf{Z}}_p$ in Figure 3 denote some auxiliary processes, represented by the simplified rainfall model in our case, which also satisfy a relationship identical to (9).

The problem is: Given a time series $\mathbf{z}_p$ of the actual process $\mathbf{Z}_p$, generate a series $\mathbf{x}_s$ of the actual process $\mathbf{X}_s$. To this aim, we first generate another (auxiliary) time series $\tilde{\mathbf{x}}_s$ using the simplified rainfall process $\tilde{\mathbf{X}}_s$. The latter time series is generated independently of $\mathbf{z}_p$ and, therefore, $\tilde{\mathbf{x}}_s$ do not add up to the corresponding $\mathbf{z}_p$, as required by the additive property (9), but to some other quantities, denoted as $\tilde{\mathbf{z}}_p$. Thus, in a subsequent step, we modify the series $\tilde{\mathbf{x}}_s$ thus producing the series $\mathbf{x}_s$ consistent with $\mathbf{z}_p$ (in the sense that $\mathbf{x}_s$ and $\mathbf{z}_p$ obey (9)) without affecting the stochastic structure of $\tilde{\mathbf{x}}_s$. For this modification we use a so-called coupling transformation, i.e., a linear transformation, $\mathbf{f}(\tilde{\mathbf{X}}_s, \tilde{\mathbf{Z}}_p, \mathbf{Z}_p)$ whose outcome is a process identical to $\mathbf{X}_s$ and consistent to $\mathbf{Z}_p$.

Let $\mathbf{X}_p^* := [\mathbf{X}_{(p-1)k+1}^T, \ldots, \mathbf{X}_{pk}^T]^T$ the vector containing the hourly values of the 24 hours of any day p for all examined locations (i.e., the 24 vectors $\mathbf{X}_s$ for all hours of the day; for 5 locations, $\mathbf{X}_p^*$ contains 24 × 5 = 120 variables). Let also $\mathbf{Z}_p^* := [\mathbf{Z}_p^T, \mathbf{Z}_{p+1}^T, \mathbf{X}_{(p-1)k}^T]^T$ the vector containing the daily values $\mathbf{Z}_p$ for all examined locations, the daily values $\mathbf{Z}_{p+1}$ of the next day for all locations, and the hourly values $\mathbf{X}_{(p-1)k}$ of the last hour of the previous day $p - 1$ for all locations. For instance, for 5 locations $\mathbf{Z}_p^*$ contains 3 × 5 = 15 variables in total. Items (b) and (c) of the vector $\mathbf{Z}_p^*$ were included to assure that the transformation will preserve not only the covariance properties among the hourly values of each day, but the covariances with the previous and next days as well. Note that at the stage of the generation at day p the hourly values of day $p - 1$ are known (therefore, in $\mathbf{Z}_p^*$ we enter hourly values of the previous day) but the hourly values of day $p + 1$ are not known (therefore, in $\mathbf{Z}_p^*$ we enter daily values of the next day, which are known). In an identical manner, we construct the vectors $\tilde{\mathbf{X}}_p^*$ and $\tilde{\mathbf{Z}}_p^*$ from variables $\tilde{\mathbf{X}}_s$ and $\tilde{\mathbf{Z}}_p$.

Koutsoyiannis [2001] showed that the coupling transformation sought is given by

$$\mathbf{X}_p^* = \tilde{\mathbf{X}}_p^* + \mathbf{h}\,(\mathbf{Z}_p^* - \tilde{\mathbf{Z}}_p^*) \tag{10}$$

where

$$\mathbf{h} = \mathrm{Cov}[\mathbf{X}_p^*, \mathbf{Z}_p^*]\,\{\mathrm{Cov}[\mathbf{Z}_p^*, \mathbf{Z}_p^*]\}^{-1} \tag{11}$$

The quantity $\mathbf{h}\,(\mathbf{Z}_p^* - \tilde{\mathbf{Z}}_p^*)$ in (10) represents the correction applied to $\tilde{\mathbf{X}}$ to obtain $\mathbf{X}$. Whatever the value of this correction is, the coupling transformation will ensure preservation of first and second order properties of variables (means and variance-covariance matrix) and linear relationships among them (in our case the additive property (9)). However, it is desirable to have this correction as small as possible in order for the transformation not to affect seriously other properties of the simulated processes (e.g., the skewness). It is possible to make the correction small enough, if we keep repeating the generation process for the variables of each period (rather than performing a single generation only) until a measure of the correction becomes lower that an accepted limit. This measure can be defined as

$$\Delta = \|\mathbf{h}\ (\mathbf{Z}_p^* - \tilde{\mathbf{Z}}_p^*)\| \,/\, (m\ \sigma_X) \tag{12}$$

where m is the common size of $\mathbf{X}_p^*$ and $\tilde{\mathbf{X}}_p^*$, σ_X is standard deviation of hourly depth (common for all locations due to stationarity assumption) and $\|.\|$ denotes the Euclidian norm..

Given the daily process $\mathbf{Z}_p$ and the matrix $\mathbf{h}$, which determines completely the transformation, the steps followed to generate the hourly process $\mathbf{X}_s$ are the following:

1. Use the simplified rainfall model (2) or (8) to produce a series $\tilde{\mathbf{X}}_s$ for all hours of the current day p and the next day $p + 1$, without reference to $\mathbf{Z}_p$. (The series at location 1 will be identical to the given one).
2. At day p evaluate the vectors $\mathbf{Z}_p^*$ and $\tilde{\mathbf{Z}}_p^*$ using the values of $\mathbf{Z}_p$ and $\tilde{\mathbf{X}}_s$ of the current and next day, and $\mathbf{X}_s$ of the previous day.
3. Determine the quantity $\mathbf{h}\ (\mathbf{Z}_p^* - \tilde{\mathbf{Z}}_p^*)$ and the measure of correction Δ. If Δ is greater than an accepted limit Δ_{m}, repeat steps 1-3 (provided that the number of repetitions up to the current repetition has not exceeded a maximum allowed number r_{m}, which is set to avoid unending loops).
4. Apply the coupling transformation to derive $\mathbf{X}_p^*$ of the current day.
5. Repeat steps 1 to 4 for all days.

3.6 Specific difficulties

The peculiarities of the rainfall process at a fine time scale were already studied in section 3.3 for the univariate case demonstrated. Here we describe how these are handled in the multivariate modeling scheme.

Negative values. As demonstrated in section 3.3, the negative values, unavoidably generated by any linear stochastic model when the coefficient of variation is high (possibly in a high proportion but with low values), is not a major problem in our case. They are simply truncated to zero, thus having a beneficial effect in preserving the proportion of dry intervals (as also shown in next paragraph). A negative effect is the fact that truncation may be a potential source of bias to statistical properties that are to be preserved. Specifically, it is anticipated to result in overprediction of cross-correlations as it is very probable that negative values are contemporary.

Dry intervals. As already mentioned in section 3.1, the proportion of dry intervals cannot be preserved by linear stochastic models in an explicit manner. However, as demonstrated in section 3.3, after rounding off the generated values, a significant number of zero values emerge, which are added to the significant number of zero values resulting from the truncation of negative values. The total percentage of zero values resulting in this way may be comparable to (usually somewhat smaller than) the historical probability dry. It was demonstrated in section 3.3 that we can match exactly the historical probability dry by slightly modifying the rounding-off rule. For the multivariate case, the following technique was found effective: A proportion π_0 of the very small positive values, chosen at random among the generated values that are smaller than a threshold l_0 (e.g., 0.1-0.3 mm), are set to zero.

An alternative technique, based on a two-state (wet-dry) representation of hourly rainfall within a rainy day, was also studied. According to this technique, at periods when the known hourly time series (location 1) indicates dry condition (zero depth) the unknown hourly time series are forced, with a specified probability φ_0, to take zero depth as well.

Preservation of skewness. Although the coupling transformation preserves the first and second order statistics of the processes, it does not ensure the preservation of third order statistics. Thus, it is anticipated that it will result in underprediction of skewness. However, the repetition technique described in section 3.5 can result in a good approximation of skewness.

Homoscedasticity of innovations. By definition, the innovations $\mathbf{V}_s$ in the simplified multivariate rainfall model are homoscedastic, in the sense that their variances are constant, independent of the values of rainfall depths $\mathbf{X}_s$. Therefore, if, for instance, we estimate (or generate) the value at location 2, given that at location 1, we assume that the conditional variance is constant and independent of the value at location 1. This, however, does not comply with reality: by examining simultaneous hyetographs at two locations we can observe that the variance is larger during the periods of high rainfall (peaks) and smaller in periods of low rainfall (heteroscedasticity). As a result of this inconsistency, synthesized hyetographs will tend to have unrealistically similar peaks. One way to mitigate this problem is to apply a nonlinear transformation to rainfall depths.

The first candidate for nonlinear transformation is the logarithmic one,

$$\mathbf{X}_s' := \ln(\mathbf{X}_s + \boldsymbol{\zeta}) \tag{13}$$

with constant vector $\boldsymbol{\zeta} > \mathbf{0}$, where the logarithmic transformation should be read as an item to item one. The stationarity assumption allows considering all items of vector $\boldsymbol{\zeta}$ equal to a constant ζ. This transformation would be an appropriate selection if ζ were estimated so that the transformed series of known hourly depths have zero skewness, in which case the transformed variables could be assumed to be normally distributed. Then, preservation of first and second order properties of the untransformed variables is equivalent to preservation of first- and second-order properties of the transformed variables [*Koutsoyiannis*, 2001]. However, evidence from the examined data sets shows that the skewness of the transformed variables increases with increasing ζ and it still remains positive even if very small ζ are chosen. This means that the lognormal assumption is not appropriate for hourly rainfall.

A second candidate is the power transformation

$$\mathbf{X}_s' := \mathbf{X}_s^{(m)} \tag{14}$$

where the symbol (m) means that all items of the vector $\mathbf{X}_s$ are raised to the power m (item to item) where $0 < m < 1$. The stationarity assumption complies with the assumption that m is the same for all items. The preservation of the statistics of the untransformed variables does not necessarily lead to the preservation of the corresponding statistics of the transformed variables. However, the discrepancies are expected to be insignificant if m is not too low (e.g., for $m \geq 0.5$).

An alternative approach to remedy heteroscedasticity would be to classify different wet days in a month into different categories, defining different means, variances, etc., for each category. Such a classification could be based either on additional meteorological information (e.g. convective activity) for each day, or on merely the rainfall depths of each day (e.g. low, intermediate and high daily depths). This, however, was not studied further.

3.7 Methodology implementation

The methodology was implemented in a computer program with the name MuDRain (abbreviating Multivariate Disaggregation of Rainfall). The program automates most tasks of parameter estimation, performs the disaggregation and provides tabulated and graphical comparisons of historical and simulated statistics of hourly rainfall. In the parameter estimation phase, the program estimates all statistics to be preserved that are listed in section 2, apart from hourly cross-correlation coefficients whose estimation, as explained earlier, requires the GDSTM. The program offers three categories of options that must be specified by the user: (a) the use or not of repetition in the generation phase, (b) the use or not of one of the transformations mentioned in section 3.6, and (c) the use or not of the two-state representation of hourly rainfall. In case of the adoption of each of these options, the user must specify some additional parameters for the generation, which are: for (a) the maximum allowed distance Δ_{m} and the maximum allowed number of repetitions r_{m} (as defined in point 3 of section 3.5); for

(b) the transformation constant ζ or m (as defined in equation (13) or (14), respectively); and for (c) the probability φ_0, to force a dry state in each of the locations. Two additional parameters are used, which are related to the rounding off rule of generated hourly depths, i.e. the proportion π_0 and the threshold l_0.

In the current program configuration, the options and the additional parameters are specified by the user in a trial-and-error manner, i.e., starting with different trial values until the resulting statistics in the synthetic series match the actual ones. This can be seen as a fine-tuning of the model, which is manual. An automatic fine-tuning procedure, based on stochastic optimization, seems to be possible but has not been studied so far.

4 Case study

The methodology described above has been applied to the Brue catchment located in South-Western England. The catchment is equipped with 49 raingages of which only 5 were used in this case study. Five years of data were available, covering the period September 1993 to August 1998 (in fact the data of September 1993 was suspect for most of the raingages and was excluded from the analyses). Specifically, the simulation with the proposed disaggregation framework was performed using hourly data of one raingage only (raingage 1) and daily data from another four raingages (raingages 2-5) shown in Figure 4. The hourly data of raingages 2-5 was later used for tests and comparisons with simulated data. As usual, simulations were performed for each month separately. Of the different months, January and July are those with the wettest and driest regime, respectively, and, simultaneously, the strongest and the weakest cross-correlation between the different raingages. Therefore, the case study presented here is concerned with these two months.

The statistics estimated from the single-site hourly rainfall data (gage 1) are shown in Table 2 for January and Table 3 for July (interestingly, the maximum observed hourly rainfall is higher in July than in wetter January, which can be attributed to more convective weather during summer). For comparison and verification, in both tables we have included the corresponding statistics of the four raingages 2-5 estimated from the hourly data of the same period, although the latter were not used in any phase of simulation. Using the statistics of gage 1 at the hourly level, as well as the corresponding statistics at the daily level together with the cross-correlations at daily level we fitted the GDSTM in a manner described in Appendix 2. From this model we inferred the cross-correlation coefficients at hourly level, which are shown in Table 4 (for January and July). For comparison, the historical cross-correlation coefficients among the five gages are also shown in this table, although again these were not used at all in any phase of the modeling. The differences between the two groups of values are less than ±0.15 for January but much higher for July (up to ±0.30). Notably, the differences in the cross-correlation coefficients, estimated by the spatial-temporal rainfall model in this case study, from the values estimated again by the same model but using a more complete set of statistics to fit, including the hourly cross-correlations between all gages, are less than ±0.05 for January and ±0.10 for July. These figures indicate that the fit of the GDSTM is very good for January and somewhat poorer for July. However, we proceeded with the hourly cross-correlations estimated by this model for both months.

In both simulations (January and July) the single-state approach was adopted, so the option of forcing a dry condition and the related parameter φ_0 were deactivated. For January, the simplified multivariate model was used in its form (8) along with the power transformation (14). The exponent m was chosen equal to 0.5, a value that was found (after trials) to prevent discrepancies in the statistics to be preserved (see section 3.6). The statistics of the transformed variables are shown in Table 5 in a format similar to that of Table 2. These statistics were used in the simplified rainfall model but the cross-correlations were unchanged as given in Table 4. This assumption was checked to be realistic for m = 0.5 or higher. For July, no transformation was assumed and the simplified multivariate model was used in its form (2).

Repetition was not necessary for January where the cross-correlation coefficients were very high, but was adopted for July. In this case, Δ_m was set 1% and r_{m} was set 1000. For the control of the proportion of dry intervals the technique described in section 3.6 was used with l_0 = 0.2 mm and π_0 = 0.40 for January and l_0 = 0.3 mm and π_0 = 0.50 for July.

Applying the disaggregation framework, synthetic hourly rainfall series were produced for the five gages, that of gage 1 being identical to the historical series. The statistics of the synthetic series are compared to the historical and model statistics in Table 2 through Table 4. It can be observed that the statistics of the synthetic series are in good agreement with historical and model statistics. A graphical comparison of the entire distribution function of historical and simulated hourly rainfall depth during wet days is given in Figure 5 for gage 3 and for the month of July; wet days have been defined here as those with areal average daily rainfall greater than 1 mm. The historical and simulated distribution functions in general, and the corresponding probabilities dry in particular, emerging as the values of the distribution functions for hourly rain depth 0.1 mm, are in good agreement to each other as shown in Figure 5.

The lag-zero cross-correlation coefficients for the five gages at hourly level, calculated both for the entire period of the data set and for merely wet days, are shown in Table 4. For the month of July, where, as discussed above, the cross-correlations modeled by GDSTM depart significantly from historical ones, we observe that the deviations of simulated cross-correlations from either the modeled or the historical ones are generally less that the deviation of the modeled cross-correlations from the historical ones. This indicates that the disaggregation reduces the discrepancies due to inappropriateness or misspecification of modeled parameters. For the month of January we have also calculated the lag-one cross-correlation coefficients, shown in Table 6. These statistics (apart for the diagonal values which in fact are autocorrelations) have not been entered in the calculations and therefore their preservation could not be assured. Nevertheless, the figures in Table 6 indicate that acceptable approximations of these statistics have been attained, with the modeled values being somewhat below the historical ones. The differences must be attributed to the fact that the autocorrelation coefficients of each site that have been entered to the model as input are in all cases somewhat smaller that the historical ones.

A further comparison is given in Figure 6 for January (upper panel) and July (lower panel) in terms of the autocorrelation function for higher lags, up to lag 10 (again, not explicitly modeled apart from lag 1) for two of the examined gages. Clearly, the autocorrelation function that corresponds to the AR(1) model departs significantly from the historical one (e.g. for January and for lag 3, the model autocorrelation is virtually zero whereas the historical one is 0.15-0.20). The autocorrelation function of the GDSTM agrees better with the historical one. One would expect that the synthetic autocorrelations would agree with those of the AR(1) model that was used to generate them. However, they agree much better with the historical autocorrelations. As discussed earlier, what forced the synthetic autocorrelations to agree with the historical ones was the given hourly rainfall series at gage 1 along with the significant cross-correlations among the different gages. This indicates that in cross-correlated sites there is no need for a higher order AR or ARMA model, since in the multivariate framework studied (with one data series known), even the AR(1) model can reproduce historical autocorrelations (for lags > 1) adequately.

The length of dry intervals is another variable that has not been explicitly modeled in this approach. Nevertheless, a comparison of historical and simulated probability distribution functions of this variable during wet days, depicted in Figure 7 for gage 3 and for the month of July, indicates an encouraging performance of the model, with a slight overprediction of the very low values of the length of dry intervals (similar to that discussed in section 3.3 with reference to Figure 2).

As an additional means of comparison, two hyetographs are given in Figure 8. It can be seen that, generally, the disaggregation model, by virtue of its multivariate character, the high cross-correlation coefficients and the conditioning on an observed hourly series at one gage, reproduced well the actual hyetographs at the other gages (two of which are shown in Figure 8). Two kinds of discrepancies worth mentioning that appear in some hyetographs (not shown in Figure 8) are (a) the generation of low intensity tails at hours where the actual intensity was zero during days with high rainfall; and (b) the generation of hyetographs that, even though they have a realistic shape, may depart from historical ones during days with low rainfall. The first kind appeared mostly in January and is most probably the effect of allocating a significant correction at the stage of applying the coupling transformation. Therefore, it could be avoided by using repetition so as to decrease the quantity of correction. The second kind appears mostly in July and may be unavoidable because of the relatively lower cross-correlations of hourly rainfall in summer months.

5 Conclusions and discussion

A new methodology for spatial-temporal disaggregation of rainfall, with wide potential hydrological applicability, has been proposed. The methodology involves the combination of several univariate and multivariate rainfall models operating at different time scales, in a disaggregation framework that can appropriately modify outputs of finer time scale series so as to become consistent with given coarser time scale series.

Potential hydrologic applications include enhancement of historical data series and generation of simulated data series. Specifically, the methodology can be applied to derive spatially consistent hourly rainfall series in raingages where only daily data are available. In addition, in a simulation framework, the methodology provides a way to take simulations of multivariate daily rainfall (incorporating spatial and temporal non-stationarity) and generate multivariate fields at fine temporal resolution. A minimum requirement for the methodology is a single-site fine-scale time series. As indicated in section 3.1, if this is not available from measurements, it may be obtained from a single-site point process model or from an existing single-site disaggregation scheme.

The method is general enough and its application is not restricted by the geographical location of the sites that are to be disaggregated, which can be arbitrary. That is, the method can preserve the modeled statistics of the rainfall process no matter how close or far to each other the sites are. However, if the distances among sites are high (of the order of 10^2 km or more) it may be not necessary to use a multivariate method like this, as the different sites will be not correlated. Conversely, if the distances are small (of the order of 10 km or less) the multivariate approach becomes necessary and, in addition, the performance of the method is improved as the cross-correlations among different sites become higher. In the latter case, apart from preserving the modeled statistics, the method is able to reproduce or approach a wider category of statistics (e.g. autocorrelations for high lags, lagged cross-correlations) and also mimic the actual hyetographs. The beneficial effect of high cross-correlations in the proposed multivariate method can be also utilized in applications to large catchments, where one may think of daisy-chaining clusters of gages together so as to disaggregate rainfall over a domain that is larger than the spatial correlation length of the rain fields.

The results presented here are extremely encouraging. Specifically, the case study presented, regarding the disaggregation of daily historical data of five raingages into hourly series, showed that the methodology results in good preservation of important properties of the rainfall process such as marginal moments, temporal and spatial correlations, as well as proportions and lengths of dry intervals. In addition, it provides a good reproduction of the actual hyetographs.

Among the weaknesses of the current configuration of method, as experienced in the case study presented, are the slight overprediction of the very low values of the length of dry intervals and some long tails with low intensity emerging in certain rainfall events during periods where the intensity was actually zero. There is considerable flexibility in the proposed scheme, and hence potential for further refinement and remediation of these weaknesses. A list of possible further investigations and improvements could include: the validation of the method with rainfall time series from different climates; the use of different simplified rainfall models, in the place of the AR(1) model, such as AR models with parameters dependent on the rainfall magnitude; the use of different types of parameter matrices **a** and **b** (for example to avoid problems associated with the lower triangular form of matrix **b**, which makes results depend on the order on which different locations are entered into the vector of hourly depths; see *Koutsoyiannis* [1999]); the automatic estimation of fine-tuning parameters by means of stochastic optimization; and the development of a methodology to infer the hourly cross-correlations from daily ones and other indicators, so as to avoid using the complex (and not necessarily accurate in case of limited data availability) GDSTM and make the method more convenient for every day applications. In particular, the study of simple sub-daily correlation structures (e.g. based on a power relationship of daily cross-correlations), which could be used instead of fitting GDSTM, is under way with encouraging results so far [*Fytilas*, 2002].

Appendix 1: Estimation of the number of 'false dry' periods in section 3.3

With reference to the discussion of section 3.3, we observe that the number of 'false dry' hours N_0 is unknown and, therefore, the total number of wet hours N is unknown, too. First, let us estimate the unknown N_0 extrapolating from the values N_1 and N_2 and assuming that the hourly depth follows an exponential distribution with parameter λ, i.e. $F(x) = 1 - e^{-\lambda x}$. From (1) we obtain

$$N_1 = (e^{-\lambda c_1} - e^{-\lambda c_2})N, \quad N_2 = (e^{-\lambda c_2} - e^{-\lambda c_3})N \tag{A.1}$$

and solving for the unknowns λ and N we find

$$\lambda = \ln(N_1 / N_2) / (c_2 - c_1), \quad N = \frac{e^{\lambda c_1}}{1 - e^{-\lambda(c_2 - c_1)}} N_1 \tag{A.2}$$

Besides

$$N_0 = (1 - e^{-\lambda c_1})N \tag{A.3}$$

so that

$$N_0 = \frac{e^{\lambda c_1} - 1}{1 - e^{-\lambda(c_2 - c_1)}} N_1 \tag{A.4}$$

and given that $e^{\lambda(c_2 - c_1)} = N_1 / N_2$ (from (A.2)) we obtain

$$N_0 = \frac{(N_1/N_2)^{c_1/(c_2 - c_1)} - 1}{1 - N_2/N_1} N_1 \tag{A.5}$$

In our example, $c_1/(c_2 - c_1) = 1/2$ and finally, after algebraic manipulations

$$N_0 = \frac{N_1}{N_2 + \sqrt{N_1 N_2}} N_1 \tag{A.6}$$

The values of our example data set shown in Table 1 result in $N_0 = 291$.

In a similar manner, it can shown that for equidistant c_i with $\Delta c = c_{i+1} - c_i$, the ratio N_i / N_{i+1} is constant, independent of i and equal to $e^{\lambda \Delta c}$. In our example data set we can observe that this property is not verified by the data ($N_1 / N_2 = 2.77$, $N_2 / N_3 = 1.66$, $N_3 / N_4 = 1.36$, etc.). This means that the exponential distribution will not fit well the data beyond c_2. This is shown in Figure A1 (upper panel), where we have plotted N_i versus x_i and we observe that the theoretical values of N_i assuming a (single-parameter) exponential distribution, depart significantly from the empirical ones for $x > 0.3$.

It is expected that the two-parameter Weibull and Gamma distributions, both having the ability to yield a J-shaped density, will give a better fit to a wider range of data values, and consequently, they will give a better estimate of N_0. With these distributions it is no longer possible to derive a closed solution for N_0. Therefore, we developed a different method that initially fits the distribution parameters based on some N_i ($i = 1, \ldots, k$ with $k \geq 3$) and then extrapolates to find N_0.

Let $M := N_1 + \ldots + N_k$. Analogously to (1) we write

$$M = [F(c_{k+1}) - F(c_1)]\, N \tag{A.7}$$

which combined with (1) yields

$$N_i = \frac{F(c_{i+1}) - F(c_i)}{F(c_{k+1}) - F(c_1)} M \tag{A8}$$

where the fraction of the right-hand side is a function of the distribution function parameters (call them α and β). This, however, cannot be satisfied precisely for all N_i, so we try to estimate α and β on the basis of the minimum fitting square error. That is, we try to

$$\text{minimize } g(\alpha, \beta) := \sum_{i=1}^{k} \left(N_i - \frac{F(c_{i+1}) - F(c_i)}{F(c_{k+1}) - F(c_1)} M \right)^2 \tag{A9}$$

Although a closed solution is not possible for this nonlinear optimization problem, it is trivial to solve it numerically using widely available software tools (e.g. in spreadsheets). Having determined the distribution parameters, we can then estimate N_0 from

$$N_0 = \frac{F(c_1)}{F(c_{k+1}) - F(c_1)} M \tag{A10}$$

In our example we performed a local fitting for the Weibull and the Gamma distribution to x values from 0.2 to 1.0 mm (i.e. using N_1 to N_5). Using the optimized parameters in each case we estimated all N_i from (A8) which we plotted on Figure A1 (upper panel) in comparison with the historical values of Table 1. Both distributions gave good fits, almost indistinguishable from each other in Figure A1 (upper panel), with the Gamma distribution resulting in slightly smaller square error. The resulting N_0 for the Weibull distribution is 1110 and for the Gamma distribution 2657, which is higher than the historical total number of zero values (1679). For additional comparison, the density functions of the three distributions used have been plotted in Figure A1 (lower panel).

Appendix 2 – Brief description of the Gaussian displacement spatial-temporal rainfall model

The Gaussian displacement spatial-temporal rainfall model used in this study is a spatial analogue of a point process model used for the rainfall process at a single site. More specifically, it has a temporal structure similar to that of the Bartlett-Lewis rectangular pulses model [*Rodriguez-Iturbe et al.*, 1987, 1988] and, in addition, it has a spatial structure known as the Gaussian displacement structure, as introduced by *Cox and Isham* [1988] and further developed by *Northrop* [1996, 1998]. In the following paragraphs there is a synopsis of the model for the purposes of this paper; the interested reader is referred to *Northrop* [1996, 1998]. Some adaptations and simplifications of the original model, developed for radar data, were necessary in order to make it appropriate for raingage input data.

The model, known as Gaussian displacement spatial-temporal rainfall model (GDSTM) assumes that rainfall is realized as a sequence of storms, each consisting of a number of cells. Both storms and cells are characterized by their centers, durations and areal extends, and in addition cells have certain uniform rainfall intensity. Specifically, the following assumptions characterize storms and cells.

Storms. Storm centers arrive in a homogeneous Poisson process of rate λ in two-dimensional space (denoted by x, y) and time (denoted by t) and are moving with a uniform velocity (V_x, V_y). Thus, if t_0 is the time that a storm is generated and (x_0, y_0) is the location of the center at that time, then the location of the center at any time $t > t_0$ is

$$x = x_0 + V_x\,(t - t_0), \quad y = y_0 + V_y\,(t - t_0) \tag{A11}$$

Each storm has a finite duration L, which is assumed exponentially distributed with parameter $\beta = 1 / \mu_L$.

Each storm has an infinite areal extent, represented by an elliptical geometry with eccentricity ε and orientation θ, and incorporates a certain number of rainfall cells. However, a storm can be assigned a finite "storm area", the area that contains a certain percentage of rainfall cells. The storm area varies randomly and in each storm is determined in terms of the realization of a random variable w, which determines uniquely (for the specific storm) a set of parameters σ_x^2, σ_y^2 and ρ that determine the (dispersion of the) displacement of cell centers from the storm center. Specifically, w is a Gamma-distributed random variable (whose value remains constant for one storm) with shape and scale parameters determined in terms of the eccentricity ε and the mean storm area μ_s. At the same time, parameter ρ is determined in terms of the eccentricity ε and the storm orientation, θ. Following the generation of w, parameters σ_x^2 and σ_y^2 are determined in terms of the eccentricity ε, the storm orientation θ, and the value of w.

Cells. Each rainfall cell is assigned a center (x_c, y_c, t_c). The time origin t_c follows a Poisson process starting at the time ordinate of the storm origin t_0 (with the first cell being located at this point) and ending at $t_0 + L$. The expected number of cells within that time interval is $\mu_c = 1 + \beta / \gamma$ where γ is the cell generation Poisson process parameter, i.e., $\gamma = (\mu_c - 1) / \mu_L$. The spatial displacements from storm center, i.e.,

$$\varDelta x = x_c - x_0 - V_x (t_c - t_0), \;\; \Delta y = y_c - y_0 - V_y (t_c - t_0) \qquad \text{(A12)}$$

are random variables jointly normally distributed with zero means, variances σ_x^2 and σ_y^2, and correlation ρ. Given these parameters, the displacement $\varDelta x$ of each cell is generated as a normal variate (0, σ_x) and the displacement $\varDelta y$ as a normal variate ($\mu_{y|x}$, $\sigma_{y}|_{x}$).

Each cell has a finite duration D, which is assumed exponentially distributed with parameter $1 / \mu_D$. Also, it has an elliptical area with major axis a, forming an angle θ with the x axis (west-east), and minor axis $b = \sqrt{1 - \varepsilon^2}\, a$. It is assumed that a is a random variable

Gamma distributed with shape and scale parameters depending on the mean storm area μ_A and the eccentricity ε, respectively.

Each cell has an intensity X independent of any other variable, exponentially distributed with parameter $1 / \mu_X$.

Model parameters and rainfall statistics. The model is defined in terms of 11 independent parameters, which have some physical or geometrical meaning, namely:

Rate of storm arrivals, λ (number of storms per km^2 per hour)

Mean cell duration, μ_D (h)

Mean storm duration, μ_L (h)

Mean cell area, μ_A (km^2)

Mean storm area, μ_s (km^2)

Mean number of cells per storm, μ_c

Mean cell intensity, μ_X (mm/h)

Component of cell/storm velocity in the x direction (east), V_x (km/h)

Component of cell/storm velocity in the y direction (north), V_y (km/h)

Cell/storm eccentricity, ε

Cell/storm orientation (in radians from east), θ

All other parameters are derived in terms of these 11 parameters. In its initial formulation, the model assumes that parameter values are constant during a rainfall event but vary among different events. However, in the application of this paper we assumed that the parameters have constant values within each month, as the parsimony of available data (raingage measurements) did not allow reliable parameter estimation for smaller time periods.

The parameters can be estimated in terms of the first and second order rainfall statistics. In the case of radar data, the rainfall statistics are calculated for spatially averaged rainfall instant intensity. On the contrary, in the case of raingage data the statistics can be calculated only in terms of the temporally aggregated rainfall intensity process at point basis, that is, in terms of the discrete time process

$$Y_i(x, y) = \int_{(i-1)h}^{ih} Y(x, y, t)\, dt \tag{A13}$$

where $Y_i(x, y)$ denotes the mean rainfall intensity at the discrete time interval i with a fixed length h, and $Y(x, y, t)$ is the point instant rainfall intensity at the point (x, y) at time t.

The mean rainfall intensity is independent of the time scale of aggregation h and is given by

$$E[Y_i(x, y)] = \lambda\, \mu_D\, \mu_A\, \mu_c\, \mu_X \tag{A14}$$

where $E[\ ]$ denotes expectation. The second order properties depend on the scale of aggregation and are determined in terms of the point-instant covariance function

$$c(u_x, u_y, \tau) := \mathrm{Cov}[Y(x, y, t),\ Y(x + u_x, y + u_y, t + \tau) \tag{A15}$$

where u_x, u_y and τ are spatial and temporal displacements (lags). This is a complicated function of all model parameters that can be evaluated only numerically. Its expression in terms of Taylor series expansion is given by *Northrop* [1996]. Given that function, the second order properties of the temporally aggregated process Y_i are given by

$$\mathrm{Var}[Y_i(x, y)] = 2\int_0^h (h - t)\, c(0, 0, t)\, dt \tag{A16}$$

$$\mathrm{Cov}[Y_i(x, y),\ Y_{i+k}(x + u_x, y + u_y)] = \int_{-h}^{h} (h - |t|)\, c(u_x, u_y, k\,h + t)\, dt \tag{A17}$$

Apparently, these statistics can be evaluated only numerically (by numerical integration, apart from special simple cases).

Model fitting. For the case study of this paper the model was simplified by ignoring (setting to zero) four of its parameters, i.e., the storm eccentricity ε, the storm orientation θ and the two velocity components V_x and V_y. This simplification was justified by analyses using hourly data of raingages of the Brue catchment [*Wheater et al.*, 2000]. For the estimation of the six parameters λ, μ_D, μ_L, μ_A, μ_s, μ_c we used an optimization technique based on second order properties of the process whereas the remaining parameter μ_X was estimated from the mean intensity (equation (A14)).

Specifically, the estimation of the six parameters was based on the variance, lag-one and lag-two autocorrelation of both hourly and daily discrete time processes, and lag zero cross-correlations of the daily process. The theoretical values are given in terms of the unknown parameters by equations (A15)-(A17). The objective function to be minimized has the form

$$f(\lambda, \mu_D, \mu_L, \mu_A, \mu_s, \mu_c) = \sum_i w_i \left(\frac{m_i - h_i}{h_i}\right)^2 \tag{A18}$$

where h_i and m_i denote historical and modeled (theoretical) statistics, respectively, and w_i are weights, which were considered equal to one for all statistics except for lag-two autocorrelations, where a weight equal to 0.1 was assumed (these weights were found appropriate for the problem studied but, apparently, could be different in other application studies). The optimization was performed using a software tool based on the generalized reduced gradient method.

Acknowledgments. The research leading to this paper was performed within the framework of the project *Generation of Spatially Consistent Rainfall Data* commissioned by the Ministry of Agriculture, Fisheries and Food of the UK. The authors wish to thank the personnel of the Ministry for the support of this research. They also thank the other members of the research group V. S. Isham, R. E. Chandler, P. J. Northrop, P. Guiblin, S. M. Bate, D. Cox, and R. Bird for their collaboration and their comments that helped improve the methodology. D.K. wishes to acknowledge the hospitality of the research group during his visit in the winter term of 1999 at the Imperial College. The constructive reviews on the earlier version of the paper (submitted in June 2000) by K. Srinivasan, N. Kumar and a critical anonymous reviewer, the comments of two anonymous reviewers of the current version and those of two WRR Associated Editors, the Acting Editor C. Welty and the editor K. E. Bencala resulted in substantial improvements and a much better presentation of the methodology.

References

Chandler, R. and H. Wheater, Climate change detection using Generalized Linear Models for rainfall, A case study from the West of Ireland, I, Preliminary analysis and modelling of rainfall occurrence, Technical report, no. 194, Department of Statistical Science, University College London, 1998a (http://www.ucl.ac.uk/Stats/research/abstracts.html)

Chandler, R. and H. Wheater, Climate change detection using Generalized Linear Models for rainfall, A case study from the West of Ireland, II, Modelling of rainfall amounts on wet days, Technical report, no. 195, Department of Statistical Science, University College London, 1998b (http://www.ucl.ac.uk/Stats/research/abstracts.html)

Connolly, R. D., J. Schirmer, and P. K. Dunn, A daily rainfall disaggregation model, *Agricultural And Forest Meteorology*, 92(2), 105-117, 1998.

Eagleson, P. S., *Dynamic Hydrology*, McGraw-Hill, 1970.

Fernandez, B., and J. D. Salas, Gamma-autoregressive models for stream-flow simulation, *J. Hydraul. Eng.*, 116(11) 1403-1414, 1990.

Fytilas, P., *Multivariate rainfall disaggregation at a fine time scale*, Diploma thesis, 171 pages, Universita degli Studi di Roma "La Sapienza", Roma, 2002.

Glasbey, C. A., G. Cooper, and M. B. McGechan, Disaggregation of daily rainfall by conditional simulation from a point-process model, *J. of Hydrol.*, 165, 1-9, 1995.

Grace, R. A., and P. S. Eagleson, *The synthesis of Short-Time-Increment Rainfall Sequences*, Report no. 91, Hydrodynamics Laboratory, Massachusetts Institute of technology, Cambridge, Mass., 1966.

Grygier, J. C. and Stedinger, J. R., Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resour. Res.*, 24(10), 1574-1584, 1988.

Grygier, J. C. and J. R. Stedinger, SPIGOT, A synthetic streamflow generation software package, Technical description, School of Civil and Environmental Engineering, Cornell University, Ithaca, NY., Version 2.5, 1990.

Gyasi-Agyei Y., Identification of regional parameters of a stochastic model for rainfall disaggregation, *J. of Hydrol.*, 223(3-4), 148-163, 1999.

Hershenhorn, J., and D. A. Woolhiser, Disaggregation of daily rainfall, *J. of Hydrol.*, 95, 299-322, 1987.

Hoshi, K. and S. J. Burges, Disaggregation of streamflow volumes, *Journal of the Hydraulics Division, Proceedings ASCE*, 105 (HY1), 27-41, 1979.

Huff, F. A., Time distribution of rainfall in heavy storms, *Water Resources Research*, 3(4), 1007-1019, 1967.

Koutsoyiannis, D., A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series, *Water Resources Research*, 28(12), 3175-3191, 1992.

Koutsoyiannis, D., A stochastic disaggregation method for design storm and flood synthesis, *J. of Hydrol.*, 156, 193-225, 1994.

Koutsoyiannis, D., Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Water Resour. Res.*, 35(4), 1219-1229, 1999.

Koutsoyiannis, D., Coupling stochastic models of different time scales, *Water Resour. Res.*, 37(2), 379-392, 2001.

Koutsoyiannis, D., and E. Foufoula-Georgiou, A scaling model of storm hyetograph, *Water Resour. Res.*,, 29(7), 2345-2361, 1993.

Koutsoyiannis, D., and A. Manetas, Simple disaggregation by accurate adjusting procedures, *Water Resour. Res.*, 32(7) 2105-2117, 1996.

Koutsoyiannis, D., and C. Onof, A computer program for temporal rainfall disaggregation using adjusting procedures, XXV General Assembly of European Geophysical Society, Nice, *Geophys. Res. Abstracts*, 2, 2000. (http://www.itia.ntua.gr/e/docinfo/59/)

Koutsoyiannis, D., and C. Onof, Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *J. of Hydrol.*, 246, 109-122, 2001.

Koutsoyiannis, D., and D. Pachakis, Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series, *Journal of Geophysical Research-Atmospheres,* 101(D21), 26444-26451, 1996.

Koutsoyiannis, D., and Th. Xanthopoulos, A dynamic model for short-scale rainfall disaggregation, *Hydrol. Sci. J.*, 35(3), 303-322, 1990.

Lane, W. L., Applied Stochastic Techniques, User's Manual, Bureau of Reclamation, Engineering and Research Center, Denver, Co., 1979.

Lane, W. L., Corrected parameter estimates for disaggregation schemes, in *Statistical Analysis of Rainfall and Runoff,* edited by V. P. Singh, Water Resources Publications, Littleton, Colo., 1982.

Lane, W. L. and D. K. Frevert, Applied Stochastic Techniques, User's Manual, Bureau of Reclamation, Engineering and Research Center, Denver, Co., Personal Computer Version 1990.

Lawrance, A. J., and P. A. W. Lewis, A new autoregressive time series model in exponential variables [NEAR(1)], *Adv. Appl. Prob.*, 13(4), 826-845, 1981.

Marien, J. L. and Vandewiele, G. L., A point rainfall generator with internal storm structure, *Water Resour. Res.*, 22(4), 475-482, 1986.

Marshak, A., A. Davis, R. Cahalan and W. Wiscombe, Bounded cascade models as nonstationary multifractals, *Phys. Rev. E*, 49(1), 55-69, 1994.

Menabde, M., D. Harris, A. Seed, G. Austin, and D. Stow, Multiscaling properties of rainfall and bounded random cascades, *Water Resour. Res.*, 33(12), 2823-2830, 1997.

Mejia, J. M. and J. Rousselle, Disaggregation models in Hydrology revisited, *Water Resour. Res.*, 12(2), 185-186, 1976.

Northrop, P., Modelling and statistical analysis of spatial-temporal rainfall fields, PhD thesis, Department of Statistical Science, University College London, 1996.

Northrop, P. J., A clustered spatial-temporal model of rainfall, Proc. R. Soc. Lond. A454, 1875-1888, 1998.

Oliveira, G. C., J. Kelman, M. V. F. Pereira, and J. R. Stedinger, Representation of spatial cross-correlation in a seasonal streamflow model, *Water Resour. Res.*, 24(5), 781-785, 1988.

Olsson, J., Scaling and fractal properties of rainfall, PhD Thesis, University of Lund, Sweden, 1996.

Olsson, J., and R. Berdtsson, Temporal rainfall disaggregation based on scaling properties, Proc. Third Intern. Workshop on Rainfall in Urban Areas: Use of Historical Rainfall Series in Hydrological Modelling, Pontresina, Switrzerland, IHP-V, Technical Documents in Hydrology, pp. 125-131, UNESCO, Paris, 1997.

Onof, C. and H. S. Wheater, Modelling of British rainfall using a Random Parameter Bartlett-Lewis Rectangular Pulse Model, *J. Hydrol.*, 149, 67-95, 1993.

Onof, C. and H. S. Wheater, Improvements to the modeling of British rainfall using a Modified Random Parameter Bartlett-Lewis Rectangular Pulses Model, *J. Hydrol.*, 157, 177-195, 1994.

Pereira, M. V. F., G. C. Oliveira, C. C. G. Costa, and J. Kelman, Stochastic streamflow models for hydroelectric systems, *Water Resour. Res*., 20(3), 379-390, 1984.

Restrepo-Posada, P. J. and P. S. Eagleson, Identification of independent rainstorms, *Journal of Hydrology*, 55, 303-319, 1982.

Rodriguez-Iturbe, D. R. Cox, and V. Isham, Some models for rainfall based on stochastic point processes, *Proc. R. Soc. Lond.*, A 410, 269-298, 1987.

Rodriguez-Iturbe, D. R. Cox, and V. Isham, A point process model for rainfall: Further developments, *Proc. R. Soc. Lond.*, A 417, 283-298, 1988.

Salas, J. D., Analysis and modeling of hydrologic time series, Chapter 19, *Handbook of hydrology*, edited by D. Maidment, McGraw-Hill, New York, 1993.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colo. 1980.

Santos, E. G., and J. D. Salas, Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering*, 118(5), 765-784, 1992.

Stedinger, J. R., and R. M. Vogel, Disaggregation procedures for generating serially correlated flow vectors, *Water Resour. Res.,* 20(1) 47-56, 1984.

Stedinger, J.R., D. Pei, and T. A. Cohn, A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resour. Res.*, 21(5) 665-675, 1985.

Tao, P.C. and J. W. Delleur, Multistation, multiyear synthesis of hydrologic time series by disaggregation, *Water Resour. Res.*, 12(6), 1303-1312, 1976.

Todini, E., The preservation of skewness in linear disaggregation schemes, *J. Hydrol.*, 47, 199-214, 1980.

Valencia, D. and J. C. Schaake, A disaggregation model for time series analysis and synthesis, Report no. 149, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Massachusetts Institute of Technology, Cambridge, Mass., 1972.

Valencia, D. and J. C. Schaake, Disaggregation processes in Stochastic Hydrology, *Water Resour. Res.*, 9(3) 211-219, 1973.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations: 3, Some applications of the point process theory to rainfall processes, *Water Resources Research*, 17(5), 1287-1294, 1981.

Wheater, H. S., V. S. Isham, C. Onof, R. E. Chandler, P. J. Northrop, P. Guiblin, S. M. Bate, D. Cox, and D. Koutsoyiannis, Generation of Spatially Consistent Rainfall Data, Volume I, Report to MAFF, Imperial College and University College London, 170 pp., London, 2000.

Woolhiser, D. A. and Osborn, H. B., A stochastic model of dimensionless thunderstorm rainfall, *Water Resour. Res.*, 21(4) 511-522, 1985.

List of Figures

Figure 1 Plots on Weibull probability paper of probability distribution function of hourly rainfall depth during rainy days at gage 1 for the month of January: (upper panel) comparison of empirical (historical) and theoretical (Gamma) distribution functions; (lower panel) comparison of historical and simulated distribution functions using the GAR model with rounding-off of resulting rainfall depths, and the AR(1) model without rounding-off of resulting rainfall depths.

Figure 2 Plots on exponential probability paper of probability distribution function of the length of dry intervals (historical and simulated) during rainy days at gage 1 for the month of January.

Figure 3 Schematic representation of actual and auxiliary processes, their links, and the steps followed to construct the actual hourly-level rainfall series from the actual daily-level rainfall series.

Figure 4 Schematic of the case study area and the raingages used for the case study (coordinates in meters according to the UK National Ordnance Survey system that is based on a true origin at 49^{o}N, 2^{o}W). The details given for each raingage are the gage number used in this case study, the official gage number, and the altitude. Hourly data was used for gage 1 only (circle) whereas only daily values were used for other gages (squares).

Figure 5 Comparison of historical and simulated probability distribution functions of hourly rainfall depth during wet days at gage 3 for the month of July (plots on Weibull probability paper).

Figure 6 Comparison of autocorrelation functions of hourly rainfall as determined from historical (H2, H5 for gages 2 and 5, respectively), or simulated (S2, S5 for gages 2 and 5, respectively) series, or predicted from the AR(1) (Markov) and GDSTM models: (upper panel) January; (lower panel) July.

Figure 7 Comparison of historical and simulated probability distribution functions of the length of dry intervals during wet days at gage 3 for the month of July (plots on exponential probability paper).

Figure 8 Comparison of historical (H2, H5 for gages 2 and 5, respectively) and simulated (S2, S5 for gages 2 and 5, respectively) hyetographs at two days with high rainfall (average daily rainfall depths 14.3 mm at 17/01/95 and 16.2 mm at 12/07/98).

Figure A1 (Upper panel) Comparison of historical number of occurrences of small hourly rainfall depths during rainy days at gage 1 for the month of January and modeled number of occurrences estimated from three distribution functions; and (lower panel) plots of the density functions of the three distributions in comparison with the empirical histogram whose first bar corresponds to observed zero values.

Tables

Table 1 Number of occurrences of small hourly rainfall depths during rainy days at gage 1 for the month of January.

Value, x_i (mm)	Corresponding interval (mm)	Number of occurrences, N_i
0	$0 \leq x < 0.1$	1679
0.2	$0.1 \leq x < 0.3$	280
0.4	$0.3 \leq x < 0.5$	101
0.6	$0.5 \leq x < 0.7$	61
0.8	$0.7 \leq x < 0.9$	45
1	$0.9 \leq x < 1.1$	25
>1	$1.1 \leq x$	89
Total		2280

Table 2 Statistics of hourly rainfall depths at each gage for the month of January.

Gage	1	2	3	4	5
Proportion dry	0.84	0.85	0.84	0.85	0.84
	0.84	0.84	0.84	0.84	0.84
	0.84	0.82	0.80	0.82	0.80
Mean (mm)	0.10	0.10	0.12	0.11	0.12
	0.10	0.10	0.10	0.10	0.10
	0.10	0.10	0.12	0.11	0.12
Maximum value	8.0	7.6	5.8	4.8	4.8
(mm)	8.0	8.0	8.0	8.0	8.0
	8.0	6.7	7.1	6.9	7.2
Standard deviation	0.39	0.40	0.44	0.40	0.41
(mm)	0.39	0.39	0.39	0.39	0.39
	0.39	0.37	0.41	0.39	0.42
Skewness	7.6	7.2	5.8	6.0	5.3
	7.6	7.6	7.6	7.6	7.6
	7.6	6.9	6.9	7.9	7.2
Lag 1 autocorrelation	0.46	0.48	0.50	0.53	0.55
	0.46	0.46	0.46	0.46	0.46
	0.46	0.44	0.44	0.47	0.43

Key: For each gage three figures are given, which are: (top) the historical value, not used in the disaggregation model (apart from values of gage 1); (middle) the value used in the disaggregation model, which is the historical value of gage 1; (bottom) the synthetic value.

Table 3 Statistics of hourly rainfall depths at each gage for the month of July.

Gage	1	2	3	4	5
Proportion dry	0.95	0.95	0.95	0.94	0.95
	0.95	0.95	0.95	0.95	0.95
	0.95	0.95	0.95	0.94	0.95
Mean (mm)	0.034	0.039	0.050	0.054	0.047
	0.034	0.034	0.034	0.034	0.034
	0.034	0.039	0.050	0.054	0.047
Maximum value	14.0	8.2	12.4	11.2	8.6
(mm)	14.0	14.0	14.0	14.0	14.0
	14.0	8.2	8.5	7.6	9.8
Standard deviation	0.32	0.28	0.39	0.39	0.38
(mm)	0.32	0.32	0.32	0.32	0.32
	0.32	0.29	0.35	0.34	0.34
Skewness	26.8	13.6	16.6	14.8	14.5
	26.8	26.8	26.8	26.8	26.8
	26.8	15.6	13.2	11.6	15.9
Lag 1 autocorrelation	0.37	0.36	0.42	0.34	0.37
	0.37	0.37	0.37	0.37	0.37
	0.37	0.34	0.39	0.33	0.34

Key: For each gage three figures are given, which are: (top) the historical value, not used in the disaggregation model (apart from values of gage 1); (middle) the value used in the disaggregation model, which is the historical value of gage 1; (bottom) the synthetic value.

Table 4 Lag-zero cross-correlation coefficients for the five gages at hourly level for the months of January (upper triangle of the table) and July (lower triangle of the table).

Gage	1	2	3	4	5
1	1.00	0.84 (0.83)	0.80 (0.78)	0.72 (0.69)	0.83 (0.81)
	1.00	0.89	0.89	0.87	0.88
	1.00	0.93 (0.93)	0.92 (0.92)	0.84 (0.83)	0.86 (0.84)
2	0.59 (0.57)	1.00	0.82 (0.80)	0.80 (0.79)	0.83 (0.81)
	0.69	1.00	0.87	0.82	0.77
	0.67 (0.67)	1.00	0.90 (0.90)	0.84 (0.84)	0.83 (0.82)
3	0.51 (0.49)	0.58 (0.56)	1.00	0.79 (0.77)	0.86 (0.84)
	0.70	0.64	1.00	0.77	0.82
	0.56 (0.54)	0.54 (0.51)	1.00	0.81 (0.79)	0.83 (0.81)
4	0.34 (0.31)	0.49 (0.48)	0.44 (0.41)	1.00	0.87 (0.86)
	0.64	0.49	0.35	1.00	0.84
	0.58 (0.58)	0.56 (0.56)	0.41 (0.38)	1.00	0.79 (0.77)
5	0.58 (0.57)	0.57 (0.56)	0.48 (0.45)	0.53 (0.51)	1.00
	0.67	0.36	0.48	0.55	1.00
	0.70 (0.69)	0.57 (0.55)	0.48 (0.44)	0.61 (0.60)	1.00

Key: For each pair of gages three rows of values are given, which are: (top) the historical value, not used in the disaggregation model; (middle) the value predicted by the spatial temporal model, which was used in the disaggregation model; (bottom) the synthetic value. The values without parentheses were calculated from the entire samples and the values in parentheses were calculated from the samples of wet days only.

Table 5 Statistics of the power transformation of hourly rainfall depths at each gage for the month of January.

Gage	1	2	3	4	5
Proportion dry	0.84	0.85	0.84	0.85	0.84
	0.84	0.84	0.84	0.84	0.84
	0.84	0.82	0.80	0.82	0.80
Mean ($mm^{0.5}$)	0.11	0.11	0.13	0.11	0.13
	0.11	0.11	0.11	0.11	0.11
	0.11	0.12	0.14	0.13	0.14
Maximum value	2.83	2.76	2.41	2.19	2.19
($mm^{0.5}$)	2.83	2.83	2.83	2.83	2.83
	2.83	2.59	2.66	2.63	2.68
Standard deviation	0.30	0.30	0.33	0.31	0.32
($mm^{0.5}$)	0.30	0.30	0.30	0.30	0.30
	0.30	0.30	0.32	0.30	0.32
Skewness	3.13	3.18	3.00	3.13	2.90
	3.13	3.13	3.13	3.13	3.13
	3.13	2.90	2.73	2.88	2.85
Lag 1 autocorrelation	0.60	0.60	0.60	0.60	0.63
	0.60	0.60	0.60	0.60	0.60
	0.60	0.62	0.64	0.66	0.64

Key: For each gage three figures are given, which are: (top) the historical value, not used in the disaggregation model (apart from values of gage 1); (middle) the value used in the disaggregation model, which is the historical value of gage 1; (bottom) the synthetic value.

Table 6 Lag one cross-correlation coefficients (Corr[X_{s-1}^{i}, X_{s}^{j}]) for the five gages at hourly level for the month of January.

Gage: $i\downarrow, j\rightarrow$	1	2	3	4	5
1	0.46 0.46 0.46	0.45 0.43	0.46 0.44	0.44 0.39	0.48 0.42
2	0.49 0.44	0.48 0.46 0.44	0.47 0.43	0.48 0.40	0.48 0.41
3	0.47 0.43	0.45 0.42	0.50 0.46 0.44	0.48 0.38	0.52 0.40
4	0.48 0.42	0.47 0.41	0.48 0.41	0.53 0.46 0.47	0.52 0.41
5	0.51 0.43	0.48 0.41	0.50 0.41	0.51 0.40	0.55 0.46 0.43

Key: For each pair of gages two or three values are given, which are: (top) the historical value; (middle) the value used in the disaggregation model (i.e., the assumed lag one autocorrelation, applicable only to diagonal elements); (bottom) the synthetic value.

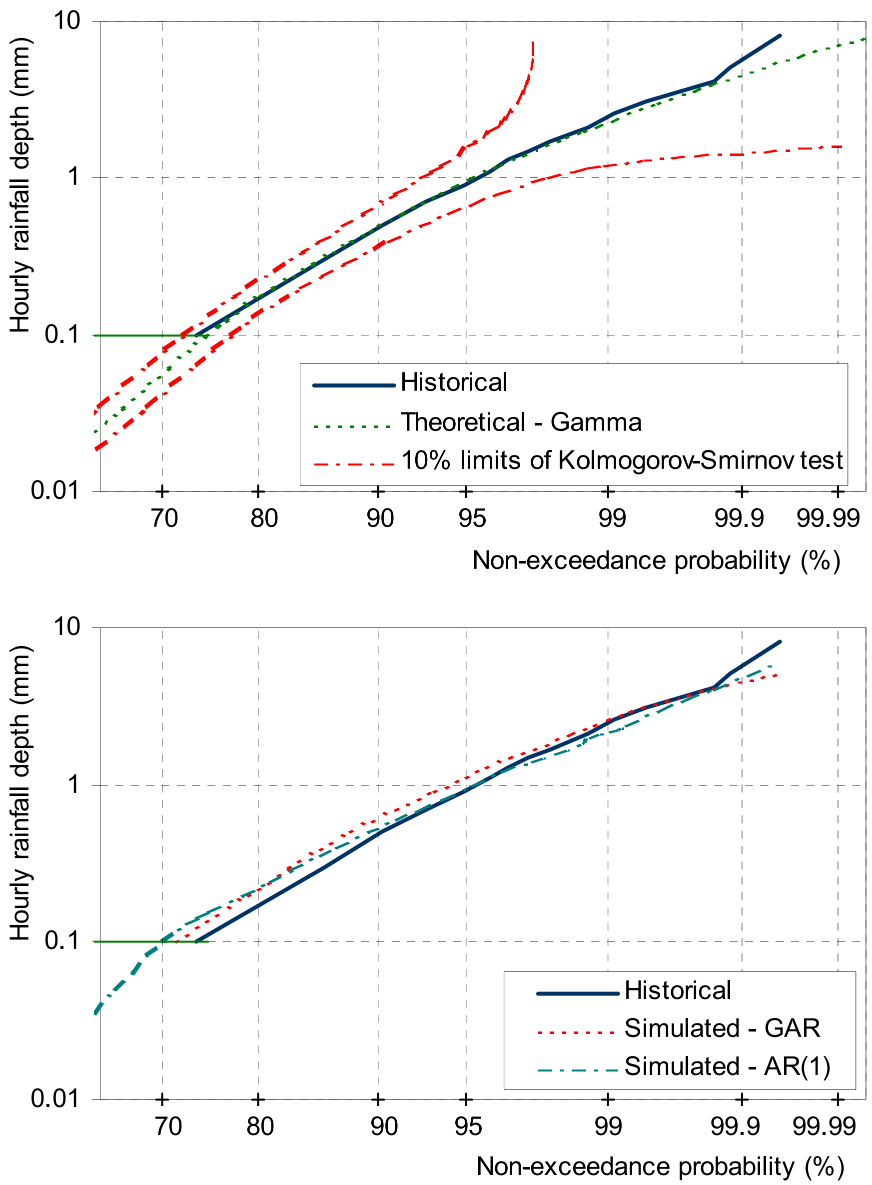

Figure 1 Plots on Weibull probability paper of probability distribution function of hourly rainfall depth during rainy days at gage 1 for the month of January: (upper panel) comparison of empirical (historical) and theoretical (Gamma) distribution functions; (lower panel) comparison of historical and simulated distribution functions using the GAR model with rounding-off of resulting rainfall depths, and the AR(1) model without rounding-off of resulting rainfall depths.

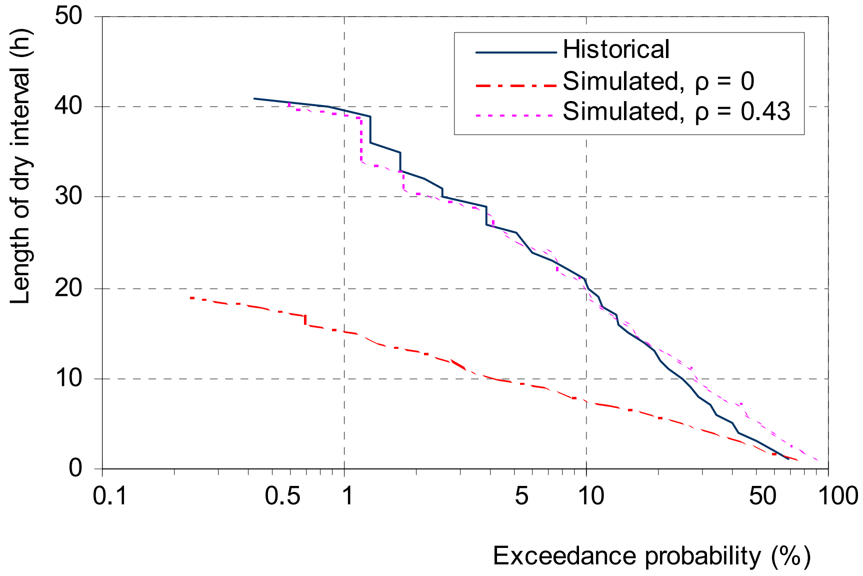

Figure 2 Plots on exponential probability paper of probability distribution function of the length of dry intervals (historical and simulated) during rainy days at gage 1 for the month of January.

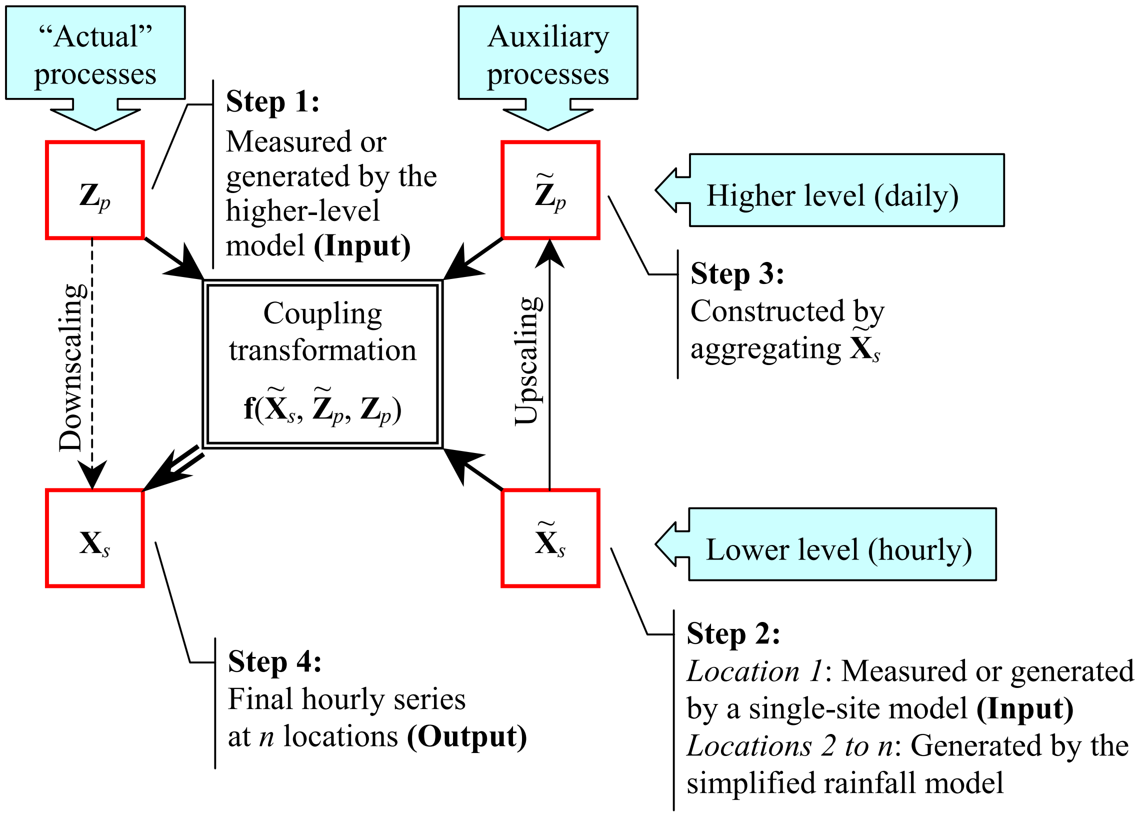

Figure 3 Schematic representation of actual and auxiliary processes, their links, and the steps followed to construct the actual hourly-level rainfall series from the actual daily-level rainfall series.

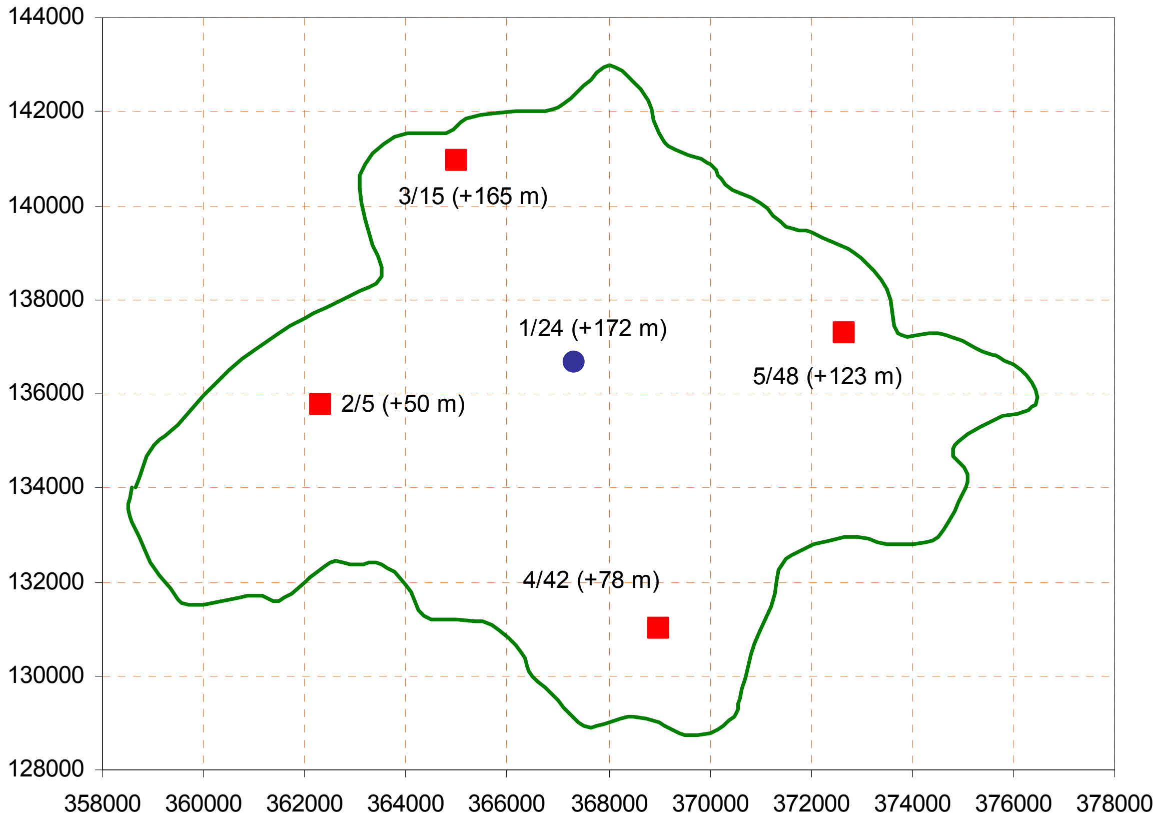

Figure 4 Schematic of the case study area and the raingages used for the case study (coordinates in meters according to the UK National Ordnance Survey system that is based on a true origin at 49°N, 2°W). The details given for each raingage are the gage number used in this case study, the official gage number, and the altitude. Hourly data was used for gage 1 only (circle) whereas only daily values were used for other gages (squares).

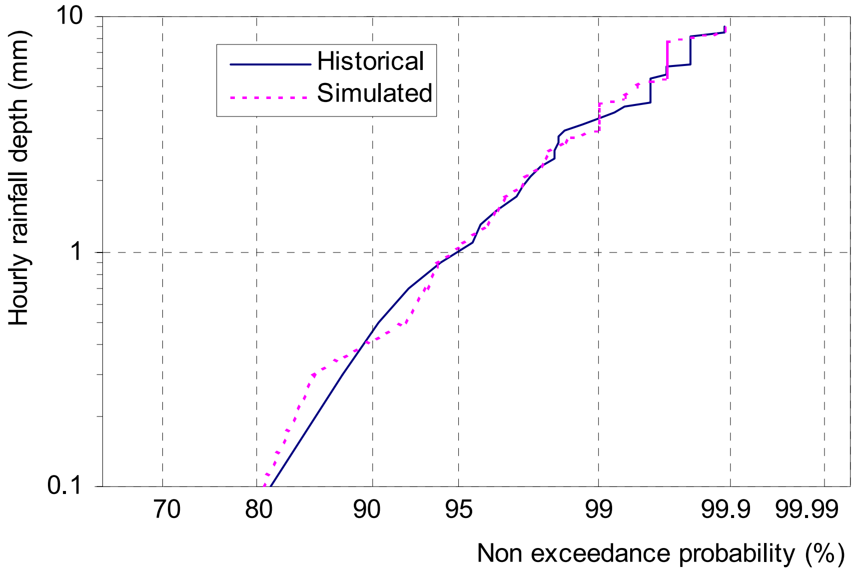

Figure 5 Comparison of historical and simulated probability distribution functions of hourly rainfall depth during wet days at gage 3 for the month of July (plots on Weibull probability paper).

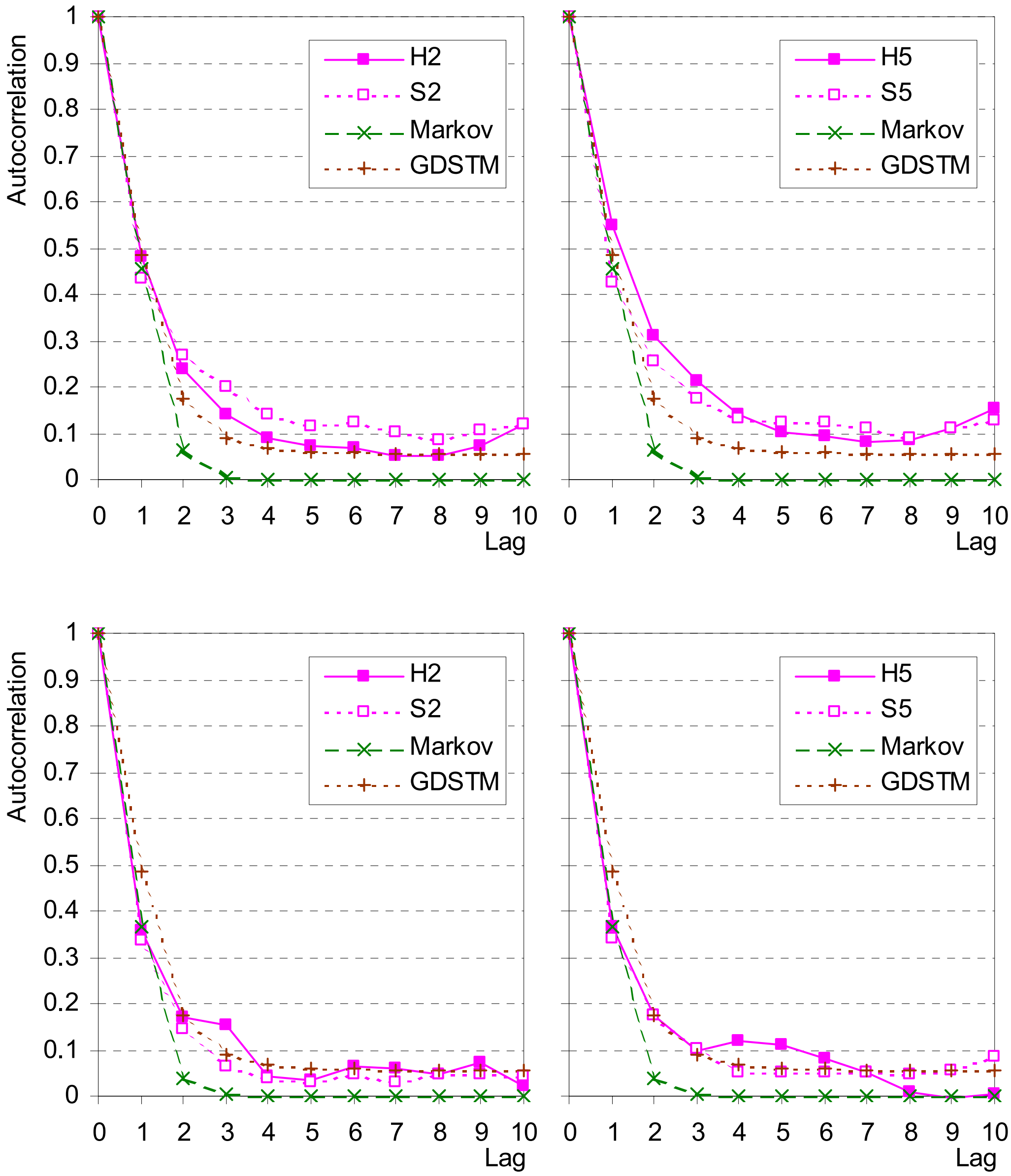

Figure 6 Comparison of autocorrelation functions of hourly rainfall as determined from historical (H2, H5 for gages 2 and 5, respectively), or simulated (S2, S5 for gages 2 and 5, respectively) series, or predicted from the AR(1) (Markov) and GDSTM models: (upper panel) January; (lower panel) July.

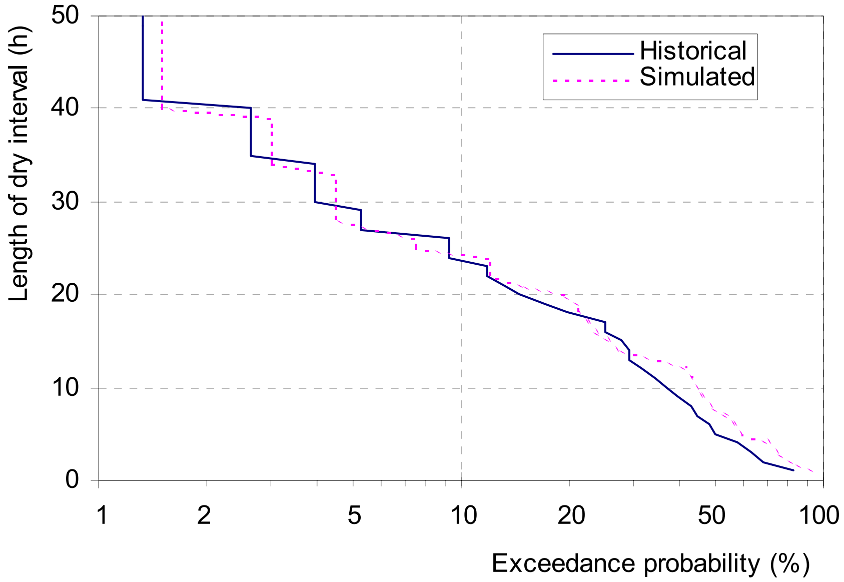

Figure 7 Comparison of historical and simulated probability distribution functions of the length of dry intervals during wet days at gage 3 for the month of July (plots on exponential probability paper).

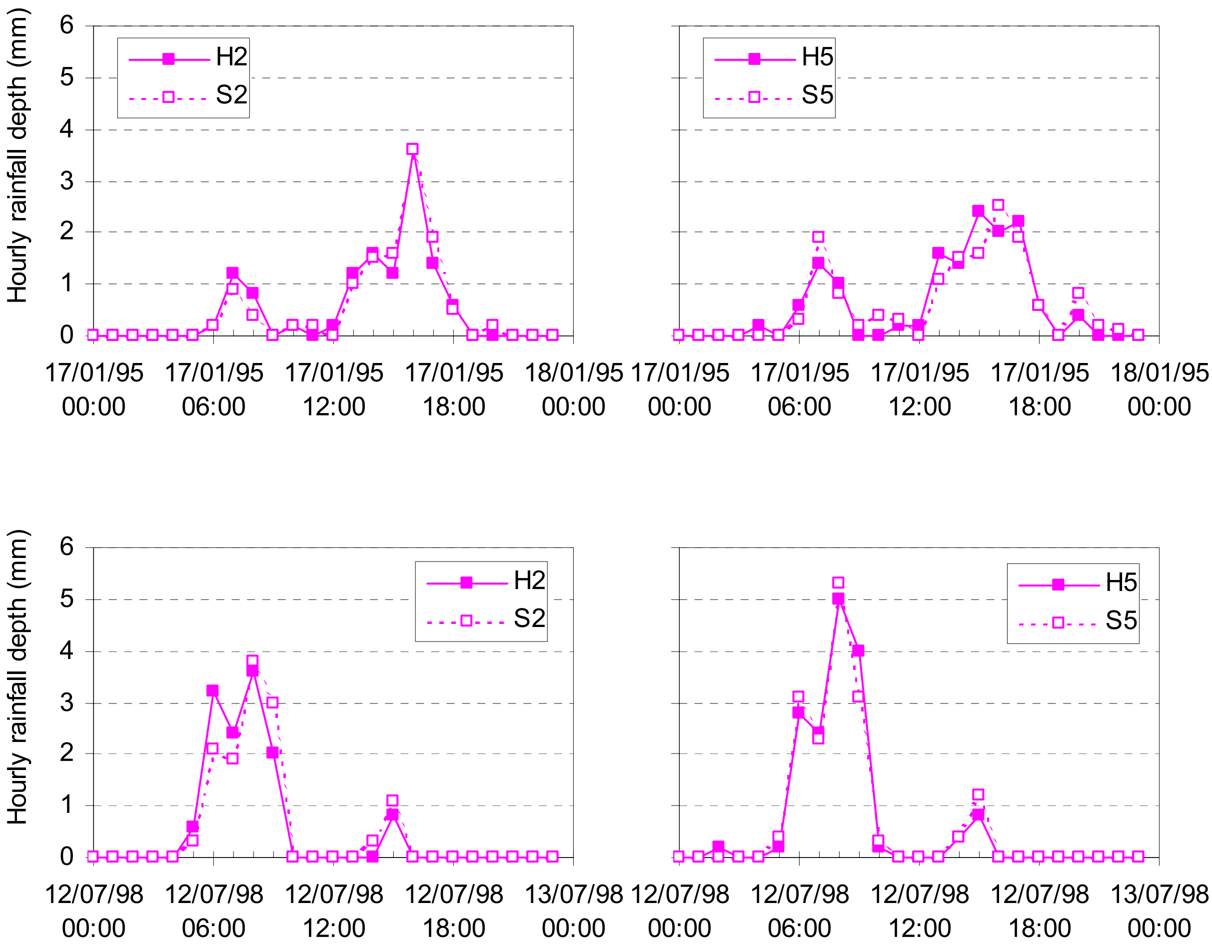

Figure 8 Comparison of historical (H2, H5 for gages 2 and 5, respectively) and simulated (S2, S5 for gages 2 and 5, respectively) hyetographs at two days with high rainfall (average daily rainfall depths 14.3 mm at 17/01/95 and 16.2 mm at 12/07/98).

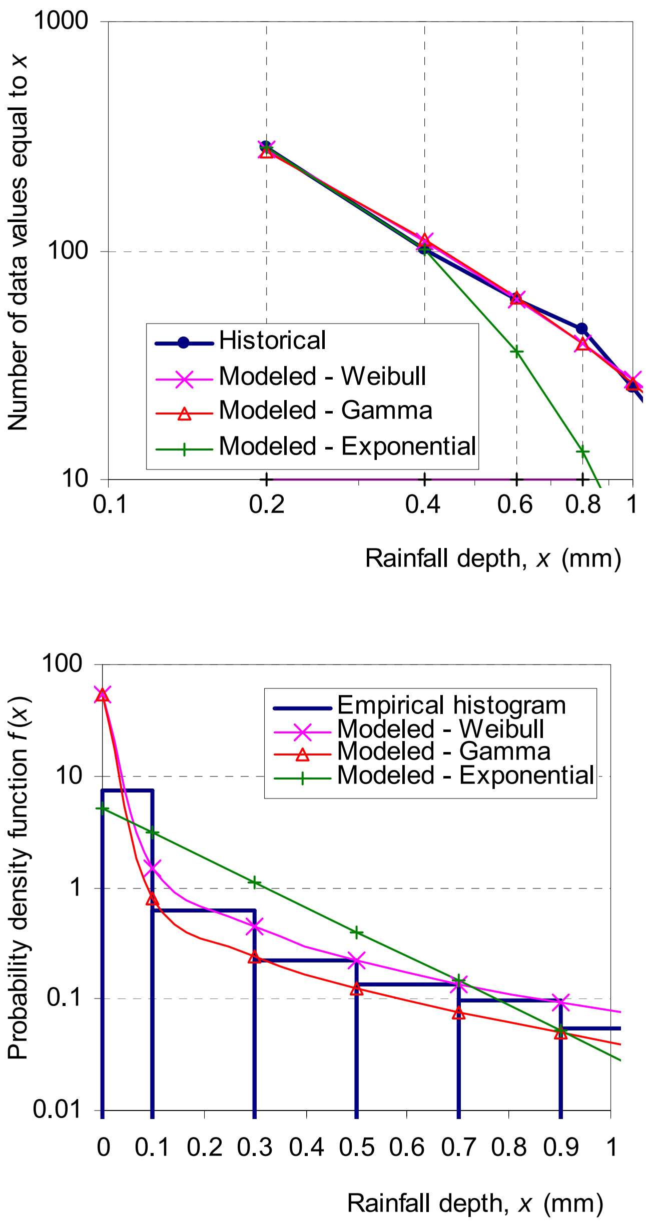

Figure A1 (Upper panel) Comparison of historical number of occurrences of small hourly rainfall depths during rainy days at gage 1 for the month of January and modeled number of occurrences estimated from three distribution functions; and (lower panel) plots of the density functions of the three distributions in comparison with the empirical histogram whose first bar corresponds to observed zero values.