Rainfall disaggregation using adjusting procedures on a Poisson cluster model

Demetris Koutsoyiannis [a] and Christian Onof [b]

[a] Department of Water Resources, Faculty of Civil Engineering, National Technical University, Athens, Greece

[b] Department of Civil and Environmental Engineering, Imperial College, London, UK.

Abstract.

A disaggregation methodology for the generation of hourly data that aggregate up to given daily totals is developed. This combines a rainfall simulation model based upon the Bartlett-Lewis process with proven techniques developed for the purpose of adjusting the finer scale (hourly) values so as to obtain the required coarser scale (daily) values. The methodology directly answers the question of the possible extension of the short hourly time-series with the use of longer-term daily data at the same point and provides the theoretical basis for an operational use of this methodology when no hourly data are available. The algorithm has been validated in full test mode in the case where hourly data are available. Specifically, two case studies (from the UK and US) are examined whose results indicate a good performance of the methodology in preserving the most important statistical properties of the rainfall process.

Keywords Rainfall, Disaggregation, Stochastic processes, Point processes.

1. Introduction

In Europe and many other countries in the world, there is a large number of daily raingauges which have often been operational for a few decades. However, the number of raingauges providing hourly or sub-hourly resolution data is smaller by about an order of magnitude. This situation reflects a general relative paucity of rainfall data for time-scales of one hour or less, both in numbers of gauges and length of the recorded series. The need for hourly data for hydrological applications, especially in flood studies, suggests the use of a disaggregation model to make use of the available daily information and provide the user with possible realisations of hourly precipitation which aggregate up to the given daily data. In this way, the model would provide a continuous simulation tool for use for simulation studies and design. This definition of disaggregation distinguishes it from downscaling, which aims at producing hourly data with the required statistics but that do not necessarily add up to the observed hourly data. The latter is in particular used for hydrological applications of GCM output where the exact values of the large-scale GCM totals are not considered particularly reliable. In both problem types, synthetic fine-scale (here hourly) data should reproduce the important statistical features of the observed rainfall, when the latter are available.

This problem has been examined by Koutsoyiannis and Xanthopoulos (1990) who developed a mathematical model for disaggregating hourly rainfall from monthly totals. Initially, they proposed a mathematical tool that could disaggregate a total amount into k partial amounts in k – 1 steps, using a stepwise procedure. Then they applied this procedure four times successively to (a) locate the starting points of storm events within a month; (b) disaggregate the monthly rainfall duration into event durations; (c) disaggregate the monthly rainfall into depths of individual events; and (d) to disaggregate the total depth of each individual event into shorter period (i.e., hourly) depths. This model could however not straightforwardly be applied to the case of the daily to hourly disaggregation. The issue of disaggregating a daily total into individual storm amounts in a day was earlier studied in a different manner by Hershenhorn and Woolhiser (1987). The issue of the disaggregation of a storm event to finer time-scales was earlier studied by Woolhiser and Osborn (1985) and Marien and Vandewiele (1986) and later was addressed in a mathematically simpler yet accurate manner by Koutsoyiannis (1994). The aim of the latter technique was however not to produce a continuous simulation tool (Zarris et al., 1998).

Such a tool was proposed by Glasbey et al. (1995). Using a random parameter Bartlett-Lewis rectangular pulse model proposed by Rodriguez-Iturbe et al. (1987), the authors examine a method based upon simulating data until a good match of daily totals is obtained for the duration of a given event in the daily data. Re-scaling is then required to reproduce daily totals exactly. This method leads to inflated hourly intensity variances. Another more ad hoc model consists in trying to reproduce the sequence of 3 daily rainfall totals with the simulation and then adjusting the data. This was more satisfactory, with a good reproduction of the main hourly statistics. However, the extreme values were not investigated and the decrease of the autocorrelations with the lag was somewhat faster than in the historical sequence, probably as a result of the adjustments which have to be carried out every three days. As the authors point out moreover, this method is not theoretically justified since three-day periods are not independent. This raises the issue of the applicability of the method to other data sets.

The development of multifractal simulation techniques has provided a potentially powerful tool for the exploration of problems such as disaggregation. An application of this approach to the disaggregation problem was proposed by Olsson (1996) and Olsson and Berndtsson (1997). The use of a self-similar microcanonical cascade enables the reproduction of the exact total daily rainfall, but it does not allow for the reproduction of the observed hourly autocorrelations (Tang, 1999). Bounded microcanonical cascades (Marshak et al., 1994) do however provide a tool which could be used for disaggregation. Such approaches are promising, as illustrated by the successful reproduction of rainfall statistics with canonical bounded cascades (Menabde et al., 1997), but require more analysis, particularly in their ability to reproduce the dry period structure at different scales.

The approach adopted here is to combine a successful rainfall generation model (Onof and Wheater, 1993) based upon a Poisson cluster process with proven techniques (Koutsoyiannis and Manetas, 1996) developed for the purpose of adjusting the hourly totals so as to obtain the required daily totals. The method will moreover use the particular structure of the rainfall model, which is that the wet/dry structure can be generated independently of the intensity profile as well as the independence of successive storms to reduce the number of computations required. This paper will seek to validate this model in the case where hourly data are available. It therefore directly answers the question of the possible extension of the short hourly time-series with the use of longer-term daily data at the same point and provides the theoretical basis for an operational use of this methodology when no hourly data are available. The paper first presents the rainfall model and disaggregation algorithm separately and then the algorithm used in their combination. After considering the issue of implementation, two case studies (from the UK and US) are examined to validate the methodology. The paper concludes by indicating the further research required.

2. General model characteristics

Many disaggregation models of the literature are ad hoc models designed as such from the beginning. On the contrary, our approach combines an existing typical rainfall simulation model along with disaggregation by adjusting, i.e., an appropriate technique for modifying the rainfall model output, so as to be consistent with the original rainfall depths at the higher-level time scale, thus performing disaggregation.

2.1 The rainfall model

As an appropriate rainfall model, the Bartlett-Lewis model was chosen due to its wide applicability and experience in calibrating and applying it to several climates. Accumulated evidence on its ability to reproduce important features of the rainfall field from the hourly to the daily scale and above can be found in the literature (Rodriguez-Iturbe et al., 1987, 1988; Onof and Wheater, 1993, 1994). This type of model has the important feature of representing rainfall in continuous time. It is therefore particularly useful in a disaggregation framework where it may be used at a time-step different from that at which it is fitted (e.g. in operational mode described in section 3; see also Gyasi-Agyei, 1999).

The general assumptions of the Bartlett-Lewis Rectangular Pulse model are (see Figure 1): (1) Storm origins t_i occur following a Poisson process with rate λ; (2) Origins t_{ij} of cells of each storm i arrive following a Poisson process with rate β; (3) Cell arrivals of each storm i terminate after a time v_i exponentially distributed with parameter γ; (4) Each cell has a duration w_{ij} exponentially distributed with parameter η; and (5) Each cell has a uniform intensity X_{ij} with a specified distribution.

In the original version of the model, all model parameters are assumed constant. In the modified version, the parameter η is randomly varied from storm to storm with a gamma distribution with shape parameter α and scale parameter ν. Subsequently, parameters β and γ also vary so that the ratios $\kappa := \beta / \eta$ and $\phi := \gamma / \eta$ are constant.

The distribution of the uniform intensity X_{ij} is typically assumed exponential with parameter $1 / \mu_X$. Alternatively, it can be chosen as two-parameter gamma with mean μ_X and standard deviation σ_X. Thus, in its most simplified version the model uses five parameters, namely λ, β, γ, η, and μ_X (or equivalently, λ, κ, ϕ, η, and μ_X) and its most enriched version seven parameters, namely λ, κ, ϕ, α, ν, μ_X and σ_X.

2.2 The adjusting procedures

'Adjusting' of a time-series refers here to a modification of a fine scale (lower-level, such as hourly) time series, generated by a specific stochastic model, so as to be consistent with a given coarser scale (higher-level, such as daily) time series, and simultaneously not affect the stochastic structure implied by the model. Techniques for disaggregation by adjusting, have been studied by Koutsoyiannis (1994) and Koutsoyiannis and Manetas (1996).

Provided that a data series Z_p (p = 1, 2, …) is known at a higher-level time scale (e.g., daily) and a lower-level (e.g. hourly) synthetic series $\tilde{X}_s$ (s = 1, 2, …) has been generated by some stochastic model (in our case, the Bartlett-Lewis model), disaggregation by adjusting procedures is a methodology to modify the lower-level series (thus getting a modified series X_s, s = 1, 2, …) so as to make it consistent with the higher-level one. To achieve this, it uses accurate adjusting procedures to allocate the error in the additive property, i.e., the departure of the sum of lower-level variables within a period from the corresponding higher-level variable. These procedures are accurate in the sense that they preserve explicitly (at least under some specified conditions) certain statistics or even the complete distribution of lower-level variables. In addition, the methodology uses repetitive sampling in order to improve the approximations of statistics that are not explicitly preserved by the adjusting procedures.

Three such adjusting procedures have been developed and studied (Koutsoyiannis, 1994; Koutsoyiannis and Manetas, 1996). Here are some of their more important properties

2.2.1 Proportional adjusting procedure

This procedure modifies the initially generated values $\tilde{X}_s$ to get the adjusted values X_s according to

$$X_s = \tilde{X}_s \left(Z \Big/ \sum_{j=1}^{k} \tilde{X}_j \right) \qquad (s = 1, \ldots, k) \tag{1}$$

where Z is the higher-level variable and k is the number of lower-level variables within one higher-level period.

The proportional adjusting procedure is the simplest in application, among the three procedures. As shown by Koutsoyiannis (1994), it is exact for complete preservation of distributions if variables X_s are independent with two-parameter gamma distribution and common scale parameter. It also provides good approximation for dependent variables with gamma distribution. It has the advantage of not resulting in negative values X_s.

2.2.2 Linear adjusting procedure

The linear adjusting procedure modifies the initially generated values $\tilde{X}_s$ to get the adjusted values X_s according to

$$X_s = \tilde{X}_s + \lambda_s \left(Z - \sum_{j=1}^{k} \tilde{X}_j \right) \qquad (s = 1, \ldots, k) \tag{2}$$

where λ_s are unique coefficients depending on the covariances of X_s with Z. As shown by Koutsoyiannis and Manetas (1996), it is exact for complete preservation of distributions for Gaussian (dependent or independent) variables. In addition, it is exact in preserving second order moments of (dependent or independent) variables with any distribution. Its main disadvantage is that it may result in negative values, which can then be corrected using repetitions (that is, by setting them to zero and then reapplying the same procedure to adjust the resulting error).

2.2.3 Power adjusting procedure

The power adjusting procedure modifies the initially generated values $\tilde{X}_s$ to get the adjusted values X_s according to

$$X_s = \widetilde{X}_s \left(Z / \sum_{j=1}^{k} \widetilde{X}_j \right)^{\lambda_s / \eta_s} \qquad (s = 1, \ldots, k) \tag{3}$$

where λ_s are appropriate coefficients depending on the covariances of X_s with Z and η_s are coefficients depending on the mean values of X_s and Z. It is approximate apart from special cases where it is exact (e.g., when it coincides with proportional procedure) and its application requires repetitions. As in the case of the proportional adjusting procedure, it does not result in negative values. For stationary processes the power adjusting procedure is identical to the proportional procedure.

2.3 Choice of the appropriate adjusting procedure and sources of bias

The examined rainfall disaggregation problem is characterised by a large proportion of zeros (sometimes reaching or exceeding 90% in rainy days). This creates difficulties if the adjusting procedure does not prohibit negative values, as all zero values can become negative after the adjusting. For example, if the linear adjusting procedure is used and the term in parenthesis in (2) is negative, all zero values become negative after adjustment. Therefore, the linear adjusting procedure is not ideal for the problem examined.

Besides, the rainfall process can be assumed stationary (within a specific period, e.g., month) and thus the power adjusting procedure becomes identical to the proportional one. Given that the rainfall depths in rainy intervals can be assumed approximately gamma distributed, the proportional adjusting procedure seems to be the most appropriate one for our disaggregation problem.

However, as it was mentioned above, the proportional adjusting procedure is not exact in the strict sense, apart from the case of independent gamma distributed lower level variables with common scale parameter. Among these conditions, the independence one is not valid in the rainfall process at the fine time scale, and this may be considered as a potential source of bias. Koutsoyiannis and Manetas (1996) proposed repetition as a means for reducing bias in such situations. Specifically, instead of running the generation routine of the rainfall model (in our case, the Bartlett-Lewis model) once for a certain rainy period, it is run several times and the sequence of generated values that is in closest agreement with the known sequence of the higher-level (daily) variables is finally chosen.

In the rainfall disaggregation problem examined, another, more significant, source of bias, which may not be remedied by repetition, is the varying number of zero values within any specified period. We used simulations to demonstrate this, whose results are shown in Figure 2 and Figure 3. In this investigation we considered the disaggregation of a daily depth into 24 hourly depths each having a mean value $\mu_X = 1$, a standard deviation $\sigma_X = 2$ (arbitrary units) and a lag-one autocorrelation coefficient $\rho_X \geq 0$. For simplification and full compliance to the gamma distribution assumption, we assumed that the hourly rainfall process is given by a gamma autoregressive (GAR) process (Lawrance and Lewis, 1981; Fernandez and Salas, 1990), rather than the Bartlett-Lewis process. In the GAR process we also incorporated a randomly varying number (given by a binomial distribution) of zero values so that the probability of nonzero values is p. In each simulation step we generated a sequence of 24 initial variates and calculated the value of the higher-level variable Z as the sum of these initial variates. Then, using the same model, we generated another sequence of 24 variates $\tilde{X}_i$, their sum $\tilde{Z}$, and the logarithmic distance d of the latter from the initial value Z, i.e., $d = |\ln(Z/\tilde{Z})|$. We kept generating sequences $\tilde{X}_i$ until the distance became smaller than an accepted value d_a. Eventually, we adjusted the final sequence of $\tilde{X}_i$ using (1). We clarify that the sequence of d obtained with this kind of repetition is not a convergent sequence towards zero or d_a. Rather, it is a random sequence with a certain probability for d being equal to or smaller than d_a. Simply, when we reach at a realisation having this property ($d \leq d_a$) we stop performing other repetitions.

The graphs of the first and second row of Figure 2 correspond to the case that $\rho_X = 0.4$ and $p = 0.1$. For the simulations whose results are depicted in the first row of Figure 2 it was assumed that the number of nonzero values is known (i.e., equal to the number obtained in the phase of the generation of Z) whereas in those of the second row the number of nonzero values was assumed not known (generated independently). We observe that the adjusting procedure does not introduce any bias if the number of nonzero values is known (first row) but it results in a notable increase of variation and skewness if the number of nonzero values is not known. In the latter case we observe some increase in the correlation coefficient if the allowed distance d_a is large (and the number of repetitions small) but no bias is introduced if the allowed distance is smaller than about 1. In the case of variation and skewness, the bias is not eliminated even if the allowed distance becomes as low as about 0.01. More careful investigation shows that if the allowed distance becomes too small, then bias is introduced by repetition, rather than by the adjusting procedure. This is confirmed by the dotted lines with squares in Figure 2, which correspond to the case where the adjusting procedure is not applied at all.

The third row of graphs of Figure 2 corresponds to simulations with unknown number of nonzero values with $\rho_X = 0.6$ and $p = 0.1$. We observe there that the bias becomes higher due to the increase in the correlation coefficient ρ_X. A more systematic investigation of the effect of ρ_X on bias is depicted in the graphs of the first row of Figure 3, where the bias clearly increases with the increase of ρ_X if we keep the allowed distance d_a constant, equal to 0.1 and the probability of nonzero values p equal to 0.1. Likewise, the bias increases, too, with the decrease in the probability of nonzero values p, as depicted in the graphs of the second row of Figure 3, which correspond to allowed distance d_a constant, equal to 0.1, and autocorrelation ρ_X equal to 0.4.

In conclusion, this investigation shows that the use of the proportional adjusting procedure along with repetition with a fairly low allowed distance results in good preservation of the process autocorrelation. On the contrary, the use of the proportional adjusting procedure, combined with repetition or not, may introduce bias in the variation and skewness of the process if the autocorrelation or the probability of zero values is high. This problem may be remedied by introducing negative bias to the theoretical variation and skewness before simulation (see Koutsoyiannis, 2001). This may require a trial and error procedure to determine the value of negative bias, that is, some theoretical values of variation and skewness that after simulation and adjusting will result in the desired values.

This investigation is rather abstractive and appropriate to explore the model behaviour at rather extreme cases. We must note that in all real world case studies, including those presented in section 4 below, bias was always practically negligible and therefore no need emerged to apply an additional technique, like negative bias.

2.4 Coupling of the Bartlett-Lewis model with the adjusting procedure

The Bartlett-Lewis rainfall model is a continuous time model whereas the disaggregation operates on discrete time with two characteristic time scales, the higher-level (e.g., daily) and lower-level (e.g., hourly) ones. The storms and cells generated by the Bartlett-Lewis model may lie on more than one higher- or lower-level time steps. Therefore, the application of the adjusting procedure on these storms and cells must extend to more than one day. However, if applied over a long simulation period, the methodology could be extremely computer time consuming as, in addition to adjustment, it uses repetition to match the simulated to observed higher-level values. To avoid this, the simulation period must be separated to as many subperiods as possible. For this purpose, we observe that different sequences (clusters) of wet days, separated by at least one dry day, can be assumed independent. This empirical observation is consistent with the Bartlett-Lewis model, which assumes Poisson arrivals of storms. This allows independent treatment of each cluster of wet days, which reduces computer time rapidly. Thus, the Bartlett-Lewis model runs separately for each cluster of wet days. Several runs are performed for each cluster, until the departures of the sequence of daily sums from the given sequence of daily rainfall becomes lower than an acceptable limit.

Details of the repetition and disaggregation scheme are shown in Figure 4, with reference to the disaggregation of daily rainfall depths of a cluster of L wet days (preceded and followed by at least one dry day). The scheme was assembled so as to optimise computer time and incorporates four levels of repetition. Initially (Level 0), the Bartlett-Lewis model runs several times until a sequence of exactly L wet days is generated. Then (Level 1), the intensities of all cells and storms are generated and the resulting daily depths are calculated. These are compared to the original ones by means of the logarithmic distance

$$d = \left[\sum_{i=1}^{L} \ln \left(\frac{Z_i + c}{\tilde{Z}_i + c} \right)^2 \right]^{1/2} \tag{4}$$

where Z_i and $\tilde{Z}_i$ are the original and generated, respectively, daily depths of day i of the wet day sequence and c a small constant (= 0.1 mm). The logarithmic transformation is selected to avoid domination by the very high values and the constant c was inserted to avoid domination by the very low values. If the distance d is greater than an accepted limit d_a, then we re-generate the intensities of cells (Level 1 repetitions) without modifying the time locations of storms and their cells. If, however, after a large number of Level 1 repetitions, the distance remains higher than the accepted limit, this may mean that the arrangement of storms and cells is not consistent with the original (and unknown) one. In this case we discard this arrangement and generate a new one, thus entering Level 2 repetitions. Furthermore, in the case of a very long sequence of wet days it is practically impossible to get a sequence of wet days with a departure of the daily sum from the given daily rainfall smaller than the specified limit. In these cases the sequence is subdivided into sub-sequences (in a random manner), each treated independently from the others (Level 3 repetitions). The algorithm allows nested subdivisions. Eventually, the sequence with distance smaller than the accepted limit is chosen and further processed by determining the lower-level (e.g., hourly) rainfall depths through the application of the proportional adjusting procedure.

We must note that this repetition and adjustment scheme, although it has some similarities to the earlier work by Glasbey et al. (1995) is structurally different. In that work, a long array (e.g. 1000 years) was generated and stored, and then retrieved each time, seeking for a matching rainfall pattern. On the contrary, the proposed scheme does not use any auxiliary database of synthetic records. Because it avoids storing and retrieval of a database it is faster, and because it uses a unique allowed distance for each sequence of wet days (instead of seeking for the best match, whose distance from the original sequence differ from sequence to sequence) it is expected to be more accurate.

3. Model implementation

The model is implemented in a computer program (available on request from the authors) under the name Hyetos (Koutsoyiannis and Onof, 2000), which operates on a windows environment with several graphical capabilities. Hyetos supports both the original and the modified Bartlett-Lewis rectangular pulses model version with exponential or gamma intensities. For practical reasons, the model implementation is specified for the daily higher-level and the hourly lower-level scales, although the methodology described above can be used for other, coarser or finer, time scales as well. Hyetos can perform in each of the following modes:

1. Disaggregation test mode (without input). An initial sequence of storms is generated using the Bartlett-Lewis model with the given parameters and then aggregated to the hourly and daily scales. The daily sequence then serves as an "original" series, which is disaggregated, thus producing another synthetic hourly series. This mode is appropriate for testing the disaggregation model itself (e.g. by comparing original and disaggregated statistics).
2. Full test mode (with hourly input). In this mode an input file containing hourly historical data must be available. The difference from Mode 1 is that the original sequence is read from the file rather than generated. This mode is appropriate for testing (e.g. by comparing original and disaggregated statistics) the entire model performance including the appropriateness of the Bartlett-Lewis model with its parameters and the disaggregation model.
3. Operational mode (with daily input). This is similar to Mode 2 the difference being that the input file contains no hourly data but only daily. This is the usual case for the model application. It cannot provide any means for testing.
4. Rainfall model test mode (with hourly input). This is similar to Mode 2 but with synthetic data not disaggregated but generated from the Bartlett-Lewis model with the given parameters. This mode is appropriate for testing whether the Bartlett-Lewis model fits the historical data (in terms of several statistics).
5. Simple rainfall generation mode (without input and without disaggregation). This is similar to Mode 4 but with no input provided (simply the Bartlett-Lewis model parameters are entered). This mode is appropriate for the generation of rainfall series using the Bartlett-Lewis model with the given parameters without performing any disaggregation.

In all modes the Bartlett-Lewis model can be implemented either in its original or modified version with a number of parameters from 5 to 7.

4. Case studies

As test cases for the model, datasets of two raingauges with extremely different climatic conditions were used: the Heathrow airport raingauge (London, UK) and the Walnut Gulch (Arizona, USA) Gauge 13. Heathrow airport is in a wet region with almost half of the days of a year being rainy and the mean annual rainfall depth exceeding 600 mm. A characteristic climatic condition of the Heathrow region is its stability throughout the year with respect to rainfall depth. Thus, January and July, which are characteristic winter and summer months, respectively, the first being the wettest and the second the driest in terms of the proportion of dry days (47% and 63%, respectively) have almost the same mean monthly rainfall depth, around 50 mm (see Table 1).

On the contrary, Walnut Gulch is a semiarid region, the mean annual rainfall at Gauge 13 being less than 300 mm and in some years falling below 200 mm. It is characterised by a strong variability throughout the year. Thus, in May, the driest month, the mean monthly rainfall depth is as low as 4 mm and the proportion of dry days 97%, whereas in July, the wettest month, the corresponding figures are 84 mm and 61% (see Table 1).

The results of the model application in the two test cases are given in Figure 5 through Figure 8 in graphical form. All graphs are referred to the wettest and driest months, which are January and July, respectively, for Heathrow airport, and July and May, respectively, for Walnut Gulch gauge 13. In each graph results of four cases are plotted, namely (1) historical data; (2) simulated data using the Bartlett-Lewis rectangular pulse model with length equal to that of the historical record and parameters estimated from the historical data using the generalised method of moments; (3) data produced by disaggregating the historical data series 1; and (4) data produced by disaggregating the simulated data series 2. The inclusion of results from all four data series allows us to distinguish the performance of the Bartlett-Lewis model, the disaggregation model, and the combination of the two, in preserving several characteristics of the historical data series. In cases (3) and (4) the Hyetos model was applied with maximum allowed distance d = 0.1 and maximum number of repetitions (total for all levels) 5000. As indicated in the investigation of section 2.3 and Figure 2, there is no reason to use a maximum allowed distance d smaller than 0.1 because there is no gain in bias reduction. This was verified in the test cases examined here; in addition, by experimentation with these data sets it was found that significantly larger values of d, although they lead to a smaller computer time, may increase significantly the process variation and skewness. Thus, the value d = 0.1 proved to be the most appropriate.

In the full test and simple rainfall generation modes (cases (2)-(4)), the availability of hourly data to fit the model allows for the use of a set of statistics recommended in Onof and Wheater (1993) in the method of moments fitting procedure. Thus, the mean, variance and covariance lag-1 of hourly rainfall, the covariance lag-1 of the 6-hourly rainfall and the proportions of dry periods in the hourly and daily data (all with weight 1) are used.

Figure 5 depicts the proportions of dry hours and dry days in the entire period as well as the proportions of dry hours in wet days. In addition to the four sets of values estimated from data records, a fifth one is also plotted which corresponds to the theoretical values as given by the Bartlett-Lewis model equations for the estimated parameters (Rodriguez-Iturbe et al., 1987, 1988; Onof and Wheater, 1993, 1994). We observe that in all cases the empirical values of all three simulated series agree very well with the theoretically expected values.

Figure 6 depicts the coefficients of variation (standard deviation divided by mean value) and skewness of the hourly rainfall intensities. In terms of the variation, we observe a generally good preservation of the disaggregated series with respect to the original series but with a slight positive bias, which must be attributed to the reasons explained in section 2.3. However, this positive bias is generally smaller than the departure of the historical and simulated series generated by the Bartlett-Lewis model without disaggregation. Therefore, no additional technique for fine-tuning of the simulated variation seems to be necessary here. In terms of the coefficient of skewness, we must mention that the Bartlett-Lewis model does not preserve the skewness of the hourly rainfall intensities and therefore the disaggregation model cannot explicitly preserve the skewness. Nevertheless, the model produced good approximations of the historical coefficients of skewness in all cases.

Figure 7 depicts the autocorrelation coefficients of the hourly rainfall intensities for lags up to 10. We observe that autocorrelations of the disaggregated series are in good agreement with those of the Bartlett-Lewis model itself (some differences in the month May in Walnut Gulch must be attributed to estimation errors because there are only 43 wet days in 36 years), which in turn are very close to the historical ones. This means that the disaggregation model does not enter any bias in the autocorrelation function.

Finally, Figure 8 depicts the results of a detailed analysis of hourly maxima on Gumbel probability plots. Clearly, the results of all simulated series agree well with those of the historical ones and moreover, the series of hourly maxima obtained by disaggregation from the historical daily rainfall depths is closer to the historical series than the one synthesised by the Bartlett-Lewis model without disaggregation. This means that the application of the disaggregation model improves the Bartlett-Lewis model as far as the properties of maximum intensities are considered.

In addition, Figure 9 concentrates the most important comparisons of Figure 5 through Figure 7 for one of the studied cases (Walnut Gulch Gauge 13, month of July) also providing information for timescales greater than one hour. Specifically, it compares the coefficient of variation and skewness, probability of dry intervals and lag-1 autocorrelation coefficient of historical and synthetic data at timescales (aggregation levels) 1 to 24 hours. The agreement of historical and synthetic statistics is impressively good at all timescales.

5. Conclusions and discussion

The paper has developed a disaggregation methodology for the generation of hourly data which aggregate up to given daily totals. This combines a rainfall simulation model based upon the Bartlett-Lewis process with repetition techniques and adjustment procedures. The algorithm has been validated in full test mode, which means that it can as such be used when limited hourly information is available to fit it. Specifically, two case studies (from the UK and US) are examined whose results indicate a good performance of the methodology in preserving the most important statistical properties of the rainfall process. Three important extensions of this work appear natural.

- First is the use of the methodology in the fully operational mode which supposes that no hourly information at the given point is used as input. This can take on two forms: in a first case, one can assume the availability of neighbouring hourly information; in a second, one can try and dispense with any such data. An analysis of the sensitivity of the disaggregation procedure to the parameters of the Bartlett-Lewis model forms part of this study. These issues are being examined and will be reported upon in a future publication.
- A second possible extension is the more general use of an algorithm of this nature to disaggregate between other time-scales. Here, for the purpose of the downscaling of global circulation models (GCM) output – a topic of prime importance today – a simpler version of this methodology could be considered.
- As a third extension, the use of other rainfall simulation models (in particular of the random cascade type), even if they do not have the feature of a separation between the generation of the wet/dry scenarios and the intensity profiles, may be attractive particularly in the case of sub-hourly time-scales. This is under examination.

Acknowledgments. We thank Y. Gyasi-Agyei, and an anonymous reviewer for their positive and constructive reviews. D.K. wishes to acknowledge the hospitality of the rainfall research group at the Imperial College during his visit in the winter term of 1999. The Walnut Gulch Gauge 13 record is made available on the Internet by the Agricultural Research Service (ARS) research organisation.

References

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

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

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

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

Koutsoyiannis, D. (1994). A stochastic disaggregation method for design storm and flood synthesis, *Journal of Hydrology*, 156, 193-225.

Koutsoyiannis, D. (2001). Coupling stochastic models of different time scales, *Water Resources Research*, 37(2), 379-391.

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

Koutsoyiannis, D., and C. Onof (2000). A computer program for temporal rainfall disaggregation using adjusting procedures, XXV General Assembly of European Geophysical Society, Nice, *Geophys. Res. Abstracts*, vol. 2. (Presentation also available on line at http://www.hydro.ntua.gr/faculty/dk/idocs/4/2000EGSHyetos.pdf)

Koutsoyiannis, D., and T. Xanthopoulos (1990). A dynamic model for short-scale rainfall disaggregation, *Hydrol Sci. J.*, 35(3), 303-321.

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

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

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

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

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

Olsson, J. and Berndtsson, R. (1997). Temporal rainfall disaggregation based on scaling properties, *Third International Workshop on Rainfall in Urban Areas, ed. Fankhauser, R., Einfalt and Th., Arnbjerg-Nielsen, K., IHP-V, Technical Documents in Hydrology,* UNESCO.

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

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

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

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

Tang, A.C.K. (1999). Multifractal rainfall disaggregation, Final Year Project, Dept. of Civil/Environment. Engin., Imperial College.

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

Zarris, D., D. Koutsoyiannis and G. Karavokiros (1998). A simple stochastic rainfall disaggregation scheme for urban drainage modelling, *Proc. Fourth Int. Conf. on Developments in Urban Drainage Modelling*, International Association of Water Quality & Imperial College of Science, Technology and Medicine, London, pp. 85-92, 21-24 September 1998.Tables

Table 1 Characteristics and parameters of the datasets used in case studies.

Raingauge		**Heathrow airport**		**Walnut Gulch Gauge 13**	
Month		**January**	**July**	**May**	**July**
Record length (yr)		39 (1949-87)	39 (1949-87)	36 (1955-90)	36 (1955-90)
Total number wet days		641	447	43	1116
Number of clusters of wet days		232	201	34	219
Monthly rainfall	Mean (mm)	50.04	50.96	3.62	84.22
	Standard deviation (mm)	23.09	28.39	5.31	39.85
Daily rainfall	Mean (mm)	1.61	1.64	0.12	2.72
	Standard deviation (mm)	3.05	4.79	0.92	6.21
Hourly rainfall	Mean (mm)	0.067	0.068	0.005	0.113
	Standard deviation (mm)	0.305	0.580	0.124	0.956
Proportion dry	Daily	0.466	0.630	0.966	0.613
	Hourly	0.891	0.939	0.996	0.961
Parameters of BL model	α	5.675	3.038	17.624	96.612
	κ	0.5551	0.5509	0.0726	0.1983
	ϕ	0.1011	0.1037	0.0120	0.1261
	λ (d^{-1})	0.6386	0.4405	0.0352	0.4977
	μ_X (mm d^{-1})	20.33	118.56	357.21	270.34
	ν (d)	0.0896	0.0102	0.0220	0.7506

List of Figures

Figure 6 Comparison of coefficients of variation and skewness of the historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

Figure 7 Comparison of autocorrelation functions of the historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

Figure 8 Comparison of empirical distributions of maximum hourly rainfall of historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

Figure 9 Comparison of coefficient of variation and skewness, probability of dry intervals and lag one autocorrelation coefficient of historical and synthetic data at timescales (aggregation levels) 1 to 24 hours for the case study of Walnut Gulch Gauge 13, month of July.

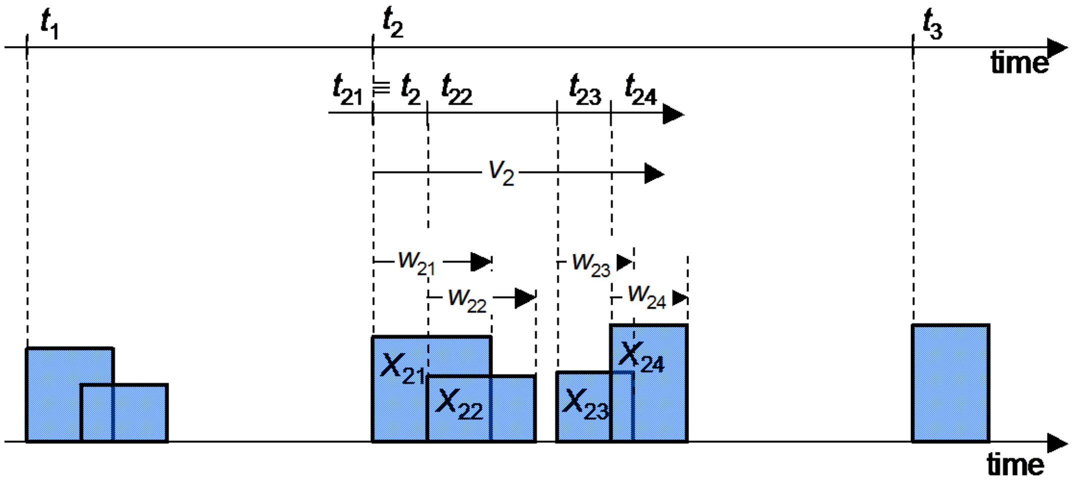

Figure 1 Explanatory sketch for the Bartlett-Lewis rectangular pulses model.

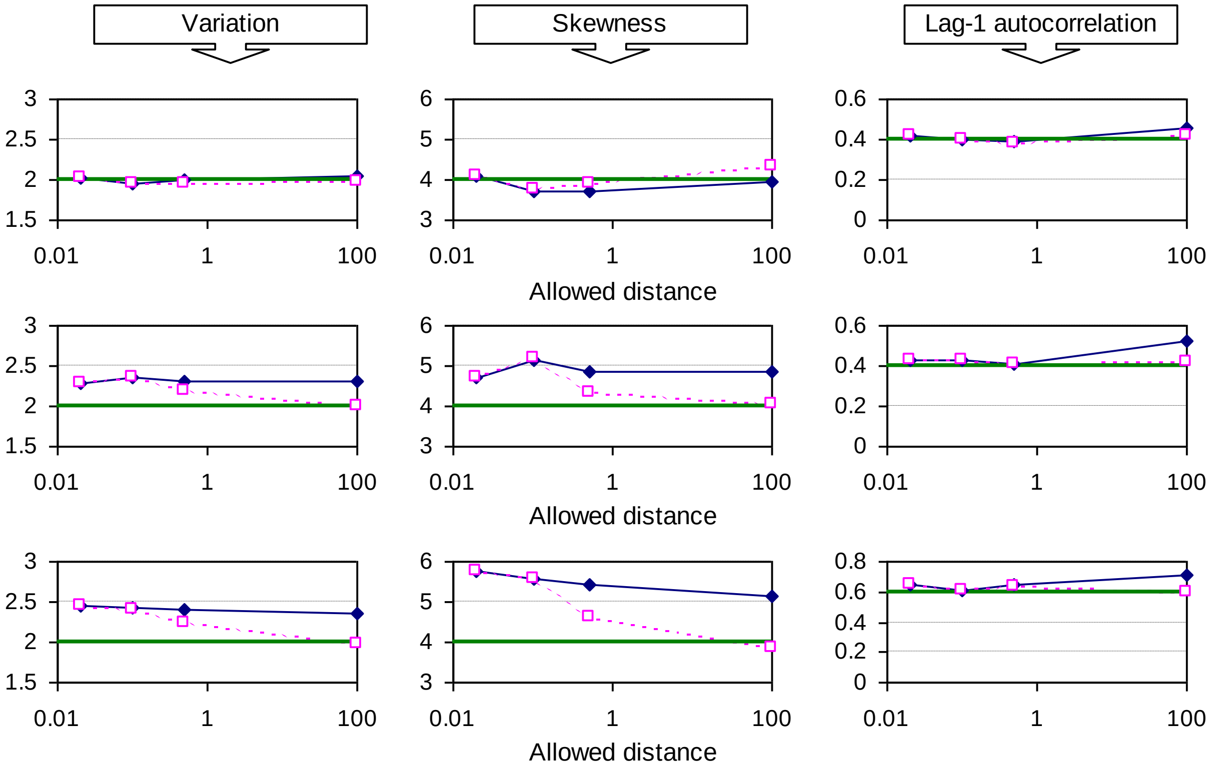

Figure 2 Investigation of bias introduced by repetition and adjustment to the variation, skewness and lag-1 autocorrelation for an intermittent GAR process (incorporating a randomly varying number of zero values), as a function of the allowed distance of repetition. Solid continuous lines represent theoretical values whereas lines with diamonds and squares represent simulated values with and without adjustment, respectively. In all cases the theoretical mean of the process is 1, the theoretical standard deviation is 2 and the probability of nonzero values is 0.1. The theoretical lag-1 autocorrelation is 0.4 for rows 1 and 2 and 0.6 for row 3. Row 1 corresponds to the case where the number of nonzero values is known whereas in rows 2 and 3 the number of nonzero values is not known.

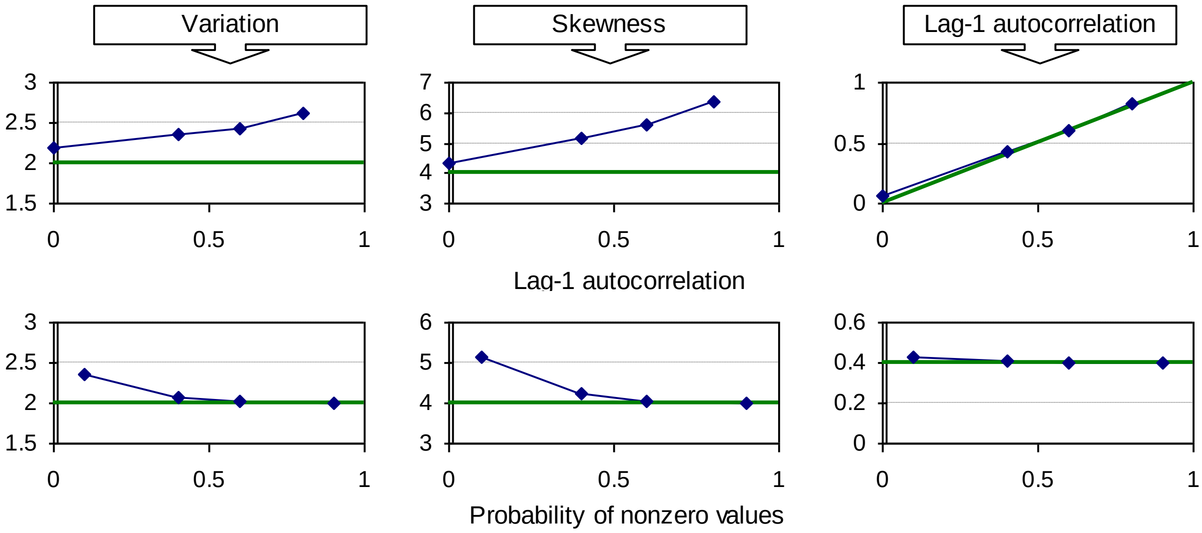

Figure 3 Investigation of bias introduced by repetition and adjustment to the variation, skewness and lag-1 autocorrelation for an intermittent GAR process (incorporating a randomly varying number of zero values), as a function of the theoretical lag-1 autocorrelation (row 1) and probability of nonzero values (row 2). Solid continuous lines represent theoretical values whereas lines with diamonds represent simulated values with adjustment. In all cases the theoretical mean of the process is 1 and the theoretical standard deviation is 2. In row 1 the probability of nonzero values is 0.1. In row 2 the theoretical lag-1 autocorrelation is 0.4. In both rows the allowed distance in repetition is 0.1.

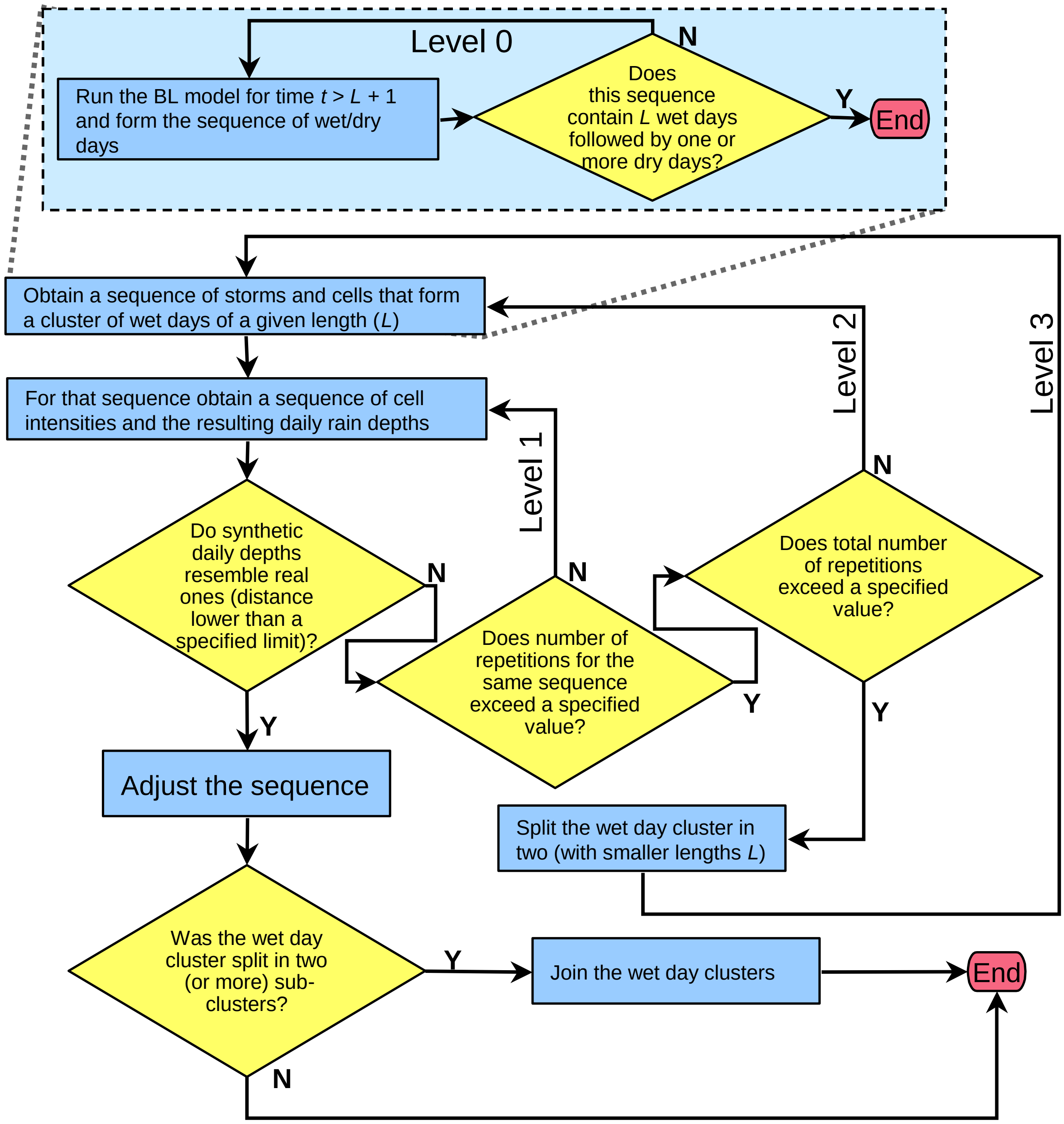

Figure 4 Flow diagram of the repetition scheme.

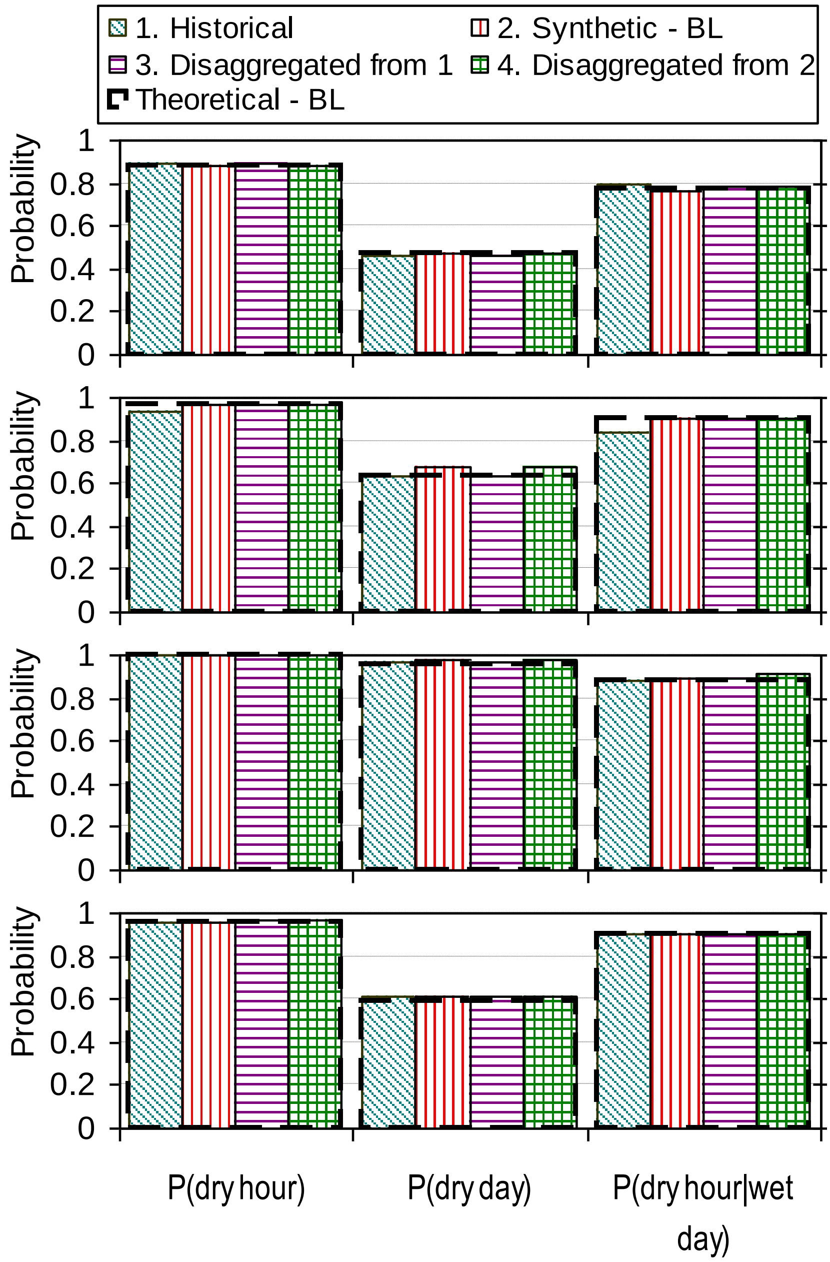

Figure 5 Comparison of dry/wet probabilities of the historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

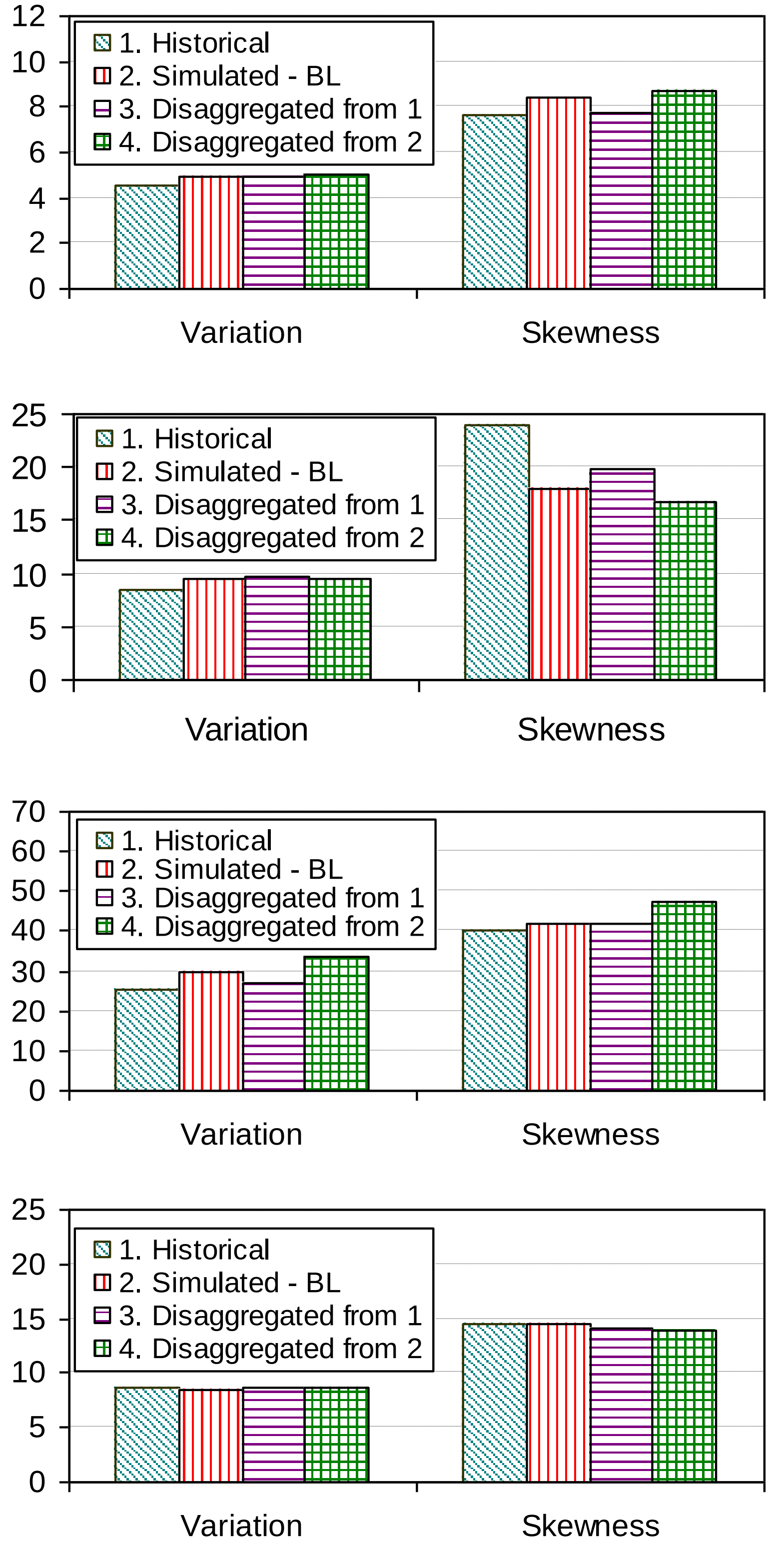

Figure 6 Comparison of coefficients of variation and skewness of the historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

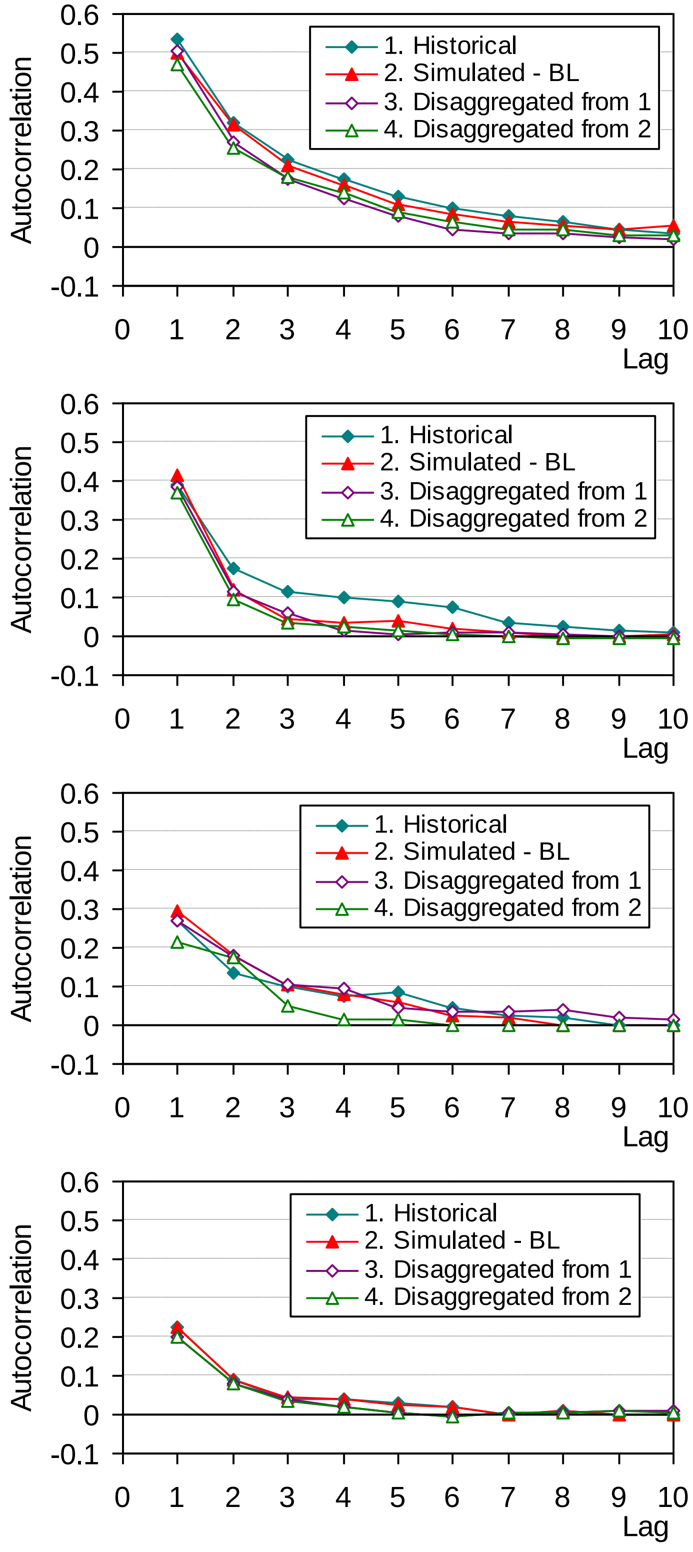

Figure 7 Comparison of autocorrelation functions of the historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

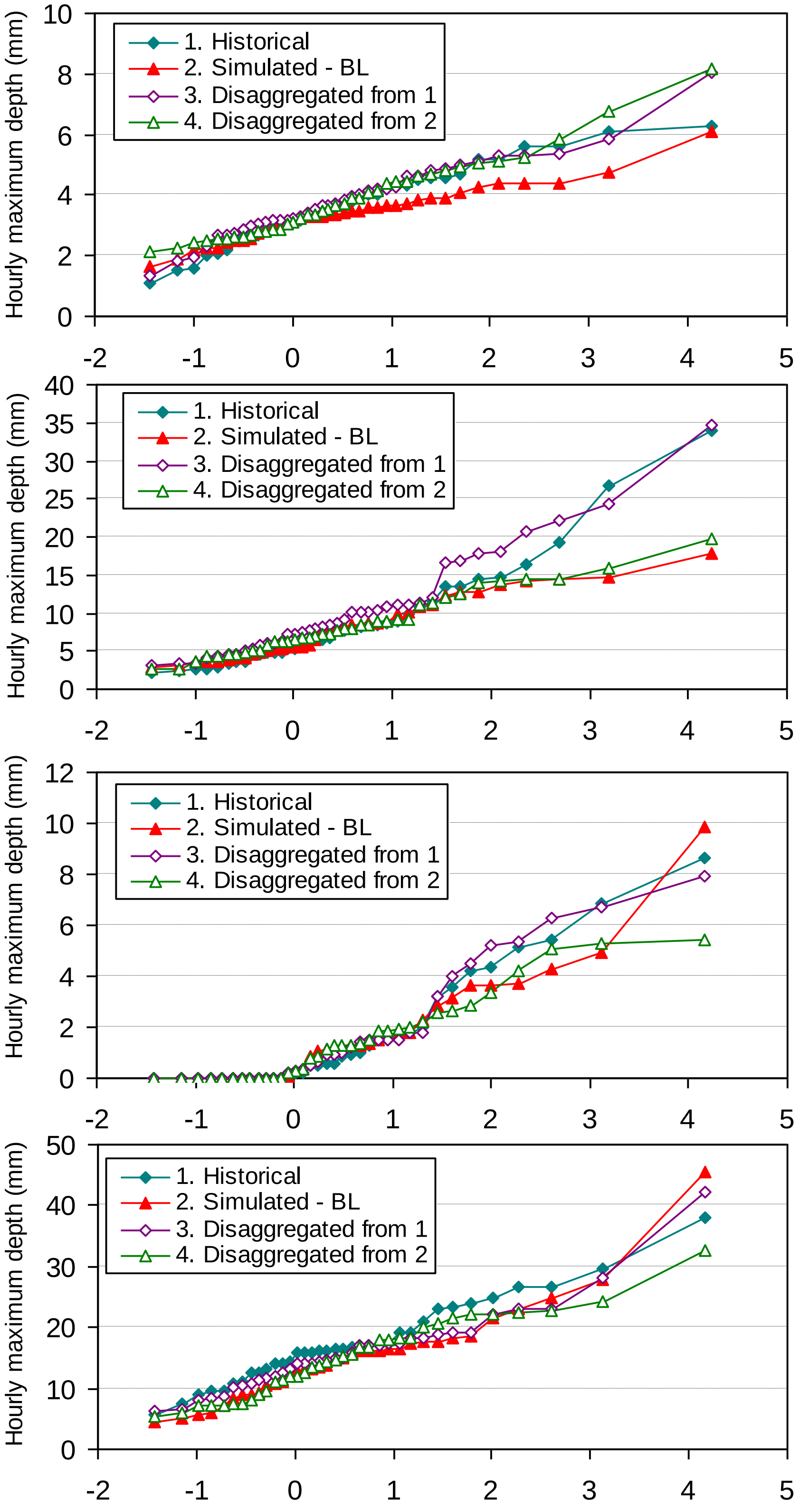

Figure 8 Comparison of empirical distributions of maximum hourly rainfall of historical and synthetic data records for the case studies of Heathrow airport, months of January (top) and July (second graph), and Walnut Gulch Gauge 13, months of May (third graph) and July (bottom).

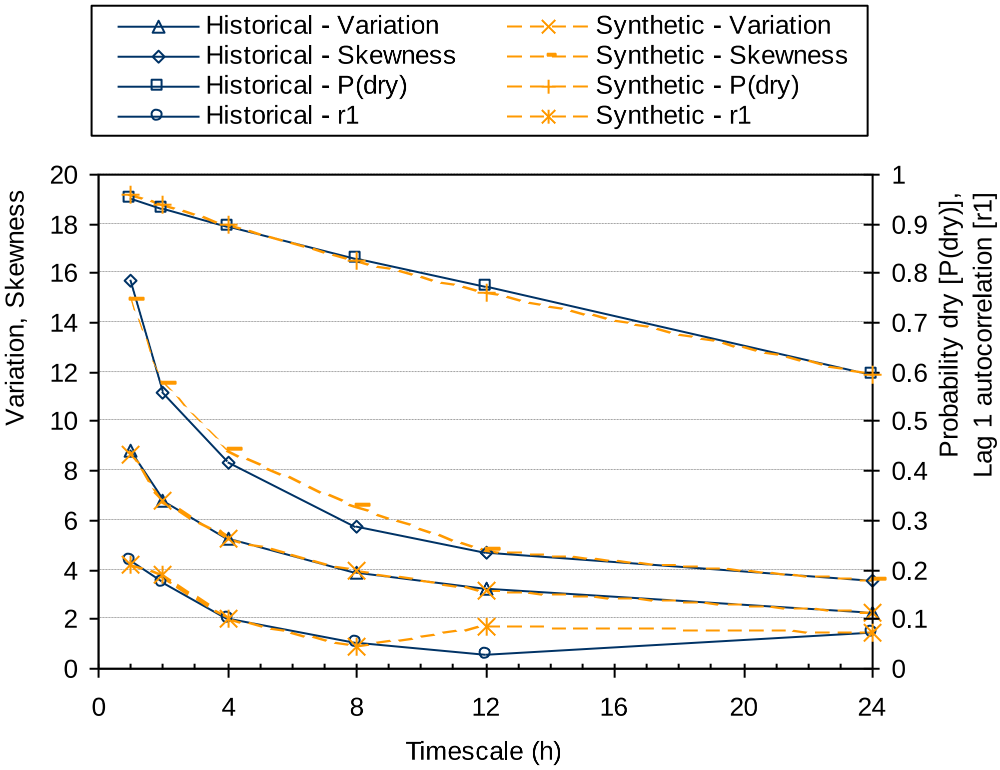

Figure 9 Comparison of coefficient of variation and skewness, probability of dry intervals and lag one autocorrelation coefficient of historical and synthetic data at timescales (aggregation levels) 1 to 24 hours for the case study of Walnut Gulch Gauge 13, month of July.