

# A Cascade Approach to Continuous Rainfall Data Generation at Point Locations

<sup>1</sup>Sivakumar, B. and <sup>2</sup>A. Sharma

<sup>1</sup>University of California, Davis, <sup>2</sup>University of New South Wales, E-Mail: [sbellie@ucdavis.edu](mailto:sbellie@ucdavis.edu)

**Keywords:** *High-resolution rainfall; Data transformation; Scale-invariance; Random cascade; Moment scaling function; Log-Poisson distribution; Sydney*

## EXTENDED ABSTRACT

High-resolution rainfall data sequences serve as inputs for a range of applications in planning, design and management of small (mostly urban) water resources systems. These applications include continuous flow simulation, rain water tank design, and the evaluation of alternate policies for assessment of environmental impacts. However, high-resolution rainfall data are often not available, since their measurements are costly and time-consuming. The lack of such data is, therefore, becoming an important issue.

As direct measurements of high-resolution rainfall data are hardly possible in practice (especially in developing and under-developed countries), a potential alternative may be to try to derive such data from the available low-resolution (daily) ones, through a data transformation procedure. Research on this (rainfall disaggregation) topic has been an important activity in hydrology and water resources over the last two decades. A host of methods have come into the fore as a consequence. Among these, the scaling-based methods seem to have received noticeable attention, essentially for their underlying concept that: the properties of the rainfall process at different scales not only are related to each other (a fundamental requirement for data transformation between such scales) but also are independent of the scale of observation (a desirable property, for data transformation over a wide range of scales).

In view of this, an attempt is made in the present study to test the utility and suitability of a random cascade scaling-based approach for transformation of rainfall data observed in Sydney, Australia. More specifically, disaggregation of rainfall data from daily scale to finer resolutions is attempted. Four different successively doubled resolutions (daily, 12-hr, 6-hr, and 3-hr) are considered. The analysis involves two systematic steps: (1) identification of the presence of scaling behavior in the rainfall process; and (2) generation of

synthetic data possessing same/similar scaling properties of the observed rainfall data.

The scaling identification is attempted for two different types of data sets, towards cross-verification and confirmation of presence of scaling: (1) rainfall data at the four individual scales; and (2) distribution of rainfall amounts (i.e. weights) between the successively doubled scales. The statistical moment scaling function method is employed for this purpose. While this is essentially a multi-scaling method, it has a particular advantage that it can distinguish between mono-scaling and multi-fractal behaviors. For synthetic data generation, a log-Poisson distribution for the weights (rainfall not used in this step) is assumed, and the scaling behavior (moment scaling function) of the observed weights is considered as the "reference" property.

The results indicate the presence of multi-scaling behaviors in all the time series analyzed (i.e. rainfall data as well as weights), as convex moment scaling function curves are observed. The synthetic data generated using the log-Poisson distribution is found to exhibit a scaling behavior that matches very well with that for the observed weights. However, the results also indicate that fitting the scaling function alone does not necessarily mean fitting the data. The use of log-Poisson distribution for synthetic data generation is found to possess an important limitation in that it does not generate zero values, which are a common occurrence in high-resolution data. The use of a scaling-based method in itself needs to be cautiously pursued since: such methods are useful only for distinguishing between mono- and multi-scaling behaviors, but not between scaling and non-scaling.

In light of the present results, future research will focus on comparison of the scaling-based approach with those based on parametric alternating renewal processes, non-parametric fragments, and nonlinear deterministic chaos.

## 1. INTRODUCTION

The importance of high-resolution rainfall data is increasingly realized, especially in the design and management of small (mostly urban) water resources systems, since such data serve as inputs for a range of applications, including continuous flow simulation, and design and management of drainage networks. However, measurements of high-resolution data are costly and time-consuming and, therefore, hardly a competitive alternative in practice, more so in developing and under-developed countries. An alternative way is to try to obtain data at high resolutions from the commonly available low-resolution (e.g. daily) ones, through a data transformation procedure.

The last two decades have witnessed a number of studies investigating the possibility of transformation of rainfall data from one scale to another. Hershenhorn and Woolhiser (1987) developed a stochastic model to disaggregate daily rainfall into a number of individual storms in a day. Each storm's starting time, duration and amount were simulated using rainfall on the day and on the preceding and following days. However, the generation of internal storm structure was not addressed. Arnold and Williams (1989) proposed a simple stochastic model to generate half-hourly rainfall intensity from daily rainfall. The model assumed that the daily rainfall fell in only one event. Cowpertwait *et al.* (1996) developed a stochastic disaggregation model for rainfall observed in the UK, which allowed historical or generated hourly data to be disaggregated into totals for shorter time intervals. They also observed the necessity of carrying out some smoothing of the disaggregated totals to obtain realistic storm profiles. Connolly *et al.* (1998) developed a model that allows disaggregation of daily rainfall into multiple events on a day and the simulation of time-varying intensity within each event. Bo *et al.* (1994) used the Bartlett-Lewis Rectangular Pulses (BLRP) model (e.g. Rodriguez-Iturbe *et al.*, 1987) to disaggregate daily rainfall into hourly values. They argued that the successful result was due to a scaling behavior of the power spectrum. Another approach was proposed by Glasbey *et al.* (1995), who modified the BLRP model to simulate hourly data consistent with observed daily totals.

Perica and Foufoula-Georgiou (1996) performed scaling-based disaggregation for spatial rainfall. They developed a disaggregation model based on the empirically-observed scaling of probability distributions of rainfall fluctuations and correlation between the scaling parameters and the convective available potential energy. Olsson (1998)

employed a cascade scheme (e.g. Gupta and Waymire, 1993) to model the temporal small-scale structure of rainfall. The main difference between this model and most other previous cascade models is the exact conservation of rainfall volume between successive cascade levels. This is termed a microcanonical property as opposed to canonical cascades where the volume is only on average conserved (e.g. Mandelbrot, 1974; Schertzer and Lovejoy, 1987). Another consequence of this cascade structure is that the weights are not mutually independent, but have a correlation of  $-1$ . However, the pairs of weights associated with different branchings are assumed to be mutually independent. Olsson and Berndtsson (1998) employed this model to disaggregate the daily seasonal rainfall series of 3 years observed in Sweden to 45-min resolution. The disaggregated data was shown to very well reproduce many fundamental characteristics of the observed 45-min data, e.g. division between rainy and dry periods, the event structure, and the scaling behavior. Menabde and Sivapalan (2000) employed a new type of a bounded random cascade model for temporal downscaling of individual rainfall events in Melbourne, Australia. The study was based on observation of a simple scaling property of the annual maximum (mean storm) rainfall intensity over the storm duration. In a similar scaling context, Sivakumar *et al.* (2001) developed a new rainfall disaggregation model using a chaotic dynamic approach, as the distributions of rainfall between different scales (i.e. weights) were found to exhibit nonlinear deterministic dynamic behaviors.

Examples of still other studies of interest from the viewpoint of continuous rainfall data generation are the ones that employ conditional simulation under assumptions of Markovian dependence, as is the logic used in the DRIP simulation model (e.g. Lambert and Kuczera, 1998; Heneker *et al.*, 2001).

It is encouraging that the studies conducted thus far on the transformation of rainfall data from one scale to another nearly unanimously support such a possibility. Particularly the scaling-based studies that employed multi-fractal random cascade schemes seem to be useful to improve our understanding of the rainfall phenomenon and the possibility of rainfall disaggregation and, therefore, can have a prospect of tremendous practical significance. The purpose of the present study is to test the utility and suitability of a random cascade scaling-based approach for temporal transformation of rainfall data observed in Sydney, Australia. More specifically, transformation of daily rainfall data to finer-resolution ones is attempted.

The investigation is carried out systematically in two steps:

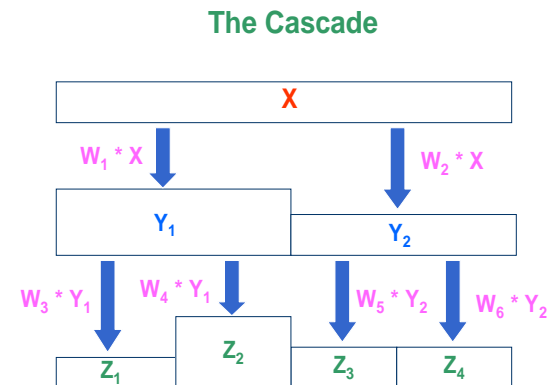
- identification of the presence of scaling in the observed rainfall (both in rainfall amounts at individual scales and in their distributions between successive scales); and
- generation of synthetic data possessing same/similar scale-invariant properties of the observed rainfall (weights)

## 2. SCALING AND CASCADES

Transformation of rainfall data from one scale to another generally involves a fundamental assumption that the properties associated with the rainfall process at these scales are related to each other. In addition, if the properties of the process are assumed to be independent of the scale of observation, then the process is said to exhibit “scaling” or “scale-invariance.” The nature of such a scaling relationship is, in general, dependent upon the behavior of the process at the different scales under consideration, or across such scales to be more precise. For example, the scaling relationship may be either ‘stochastic’ (e.g. Schertzer and Lovejoy, 1987) or ‘deterministic chaotic’ (e.g. Sivakumar *et al.*, 2001), even though most of the studies thus far investigating scaling in rainfall have assumed a stochastic relationship. While this type of distinction can indeed be made, the type of scaling relationship is generally viewed in the context of ‘dimensionality.’ More specifically, if the process can be characterized by a single “fractal” dimension, then it is said to exhibit “mono-scaling;” if more than one fractal dimension (or a dimension function) are needed to characterize the process, then the process is said to exhibit “multi-scaling.” For purpose of simplicity and generality, this type of classification is adopted in the present study, regardless of the nature of the underlying dynamics (stochastic or chaotic). Also, the methods used for scaling identification and synthetic data generation, presented later, are based on the assumption of a “random” cascade.

Figure 1 depicts a typical framework of a random cascade model for rainfall (or any other) process. Assuming transformation (disaggregation) in time, for example, the rainfall amount ( $X$ ) observed at the coarsest-resolution (e.g. daily) is distributed into rainfall amounts ( $Y_i$ ) corresponding to a finer resolution (e.g. 12 hours), where  $i$  is equal to the number of intervals to which  $X$  needs to be distributed (or “broken down”). Obviously, this distribution follows allocation of certain weights to each interval, which are denoted by  $W_i$ . Once  $Y_i$ 's are obtained, they are distributed, in a similar manner, to rainfall amounts ( $Z_i$ ) corresponding to a subsequent finer resolution (e.g. 6 hours), with

(usually) a different set of weights, consistent with the “random” assumption. This procedure is repeated until the rainfall amounts corresponding to the desired (finest) resolution (e.g. 5 minutes) are obtained. The number of intervals,  $i$ , to which rainfall amounts  $X$ 's (or  $Y$ 's or  $Z$ 's and so on) need to be distributed may be fixed or varied, depending upon the desired resolution. In this study, the number of intervals are simply doubled at successive levels, i.e. rainfall amount ( $X$ ) at a given coarser resolution is broken down into two parts to obtain amounts ( $Y_1$  and  $Y_2$ ) at the successive finer resolution.



**Figure 1.** Generic Framework of a Random Cascade Model:  $X$  represents the aggregate value,  $Y_i$  (and  $Z_i$ ) are the disaggregated values, and  $W_i$  are the weights of distribution of  $X$  (and  $Y_i$ ).

## 3. METHODOLOGY

The methodology for high-resolution rainfall data generation herein involves two steps: (1) identification of scaling behavior in the rainfall process (both in rainfall amounts and their distributions between successive scales); and (2) generation of synthetic data possessing similar scaling properties to the observed ‘process.’ Several (mono-scaling and multi-scaling) methods are available for scaling identification. Among these, the statistical moment scaling function method (e.g. Frisch and Parisi, 1985; Over and Gupta, 1994) is used herein. While this method is essentially considered as a multi-scaling method, it has also a particular advantage in that it can distinguish between mono-scaling and multi-scaling behaviors. In the second step of synthetic data generation, the distribution of weights are assumed as log-Poisson, and the moment scaling function is considered as the “reference” scaling property [see also Onof and Townend (2004)]. The procedures involved in these scaling identification and synthetic data generation methods are presented next.

### 3.1. Scaling Identification: Statistical Moment Scaling Function Method

In the statistical moment scaling function method, the time series,  $X_t$ ,  $t = 1, 2, \dots, N$ , is divided into non-overlapping intervals of a certain time resolution. The ratio of the maximum scale of the field to this interval is termed the "scale ratio,"  $\lambda$ . Thus,  $\lambda$  is inversely proportional to the size of the scale examined. For different scale ratios,  $\lambda$ , the average threshold,  $\varepsilon(\lambda, i)$ , in each interval,  $i$ , is computed and raised to power  $q$ , and subsequently summed to obtain the statistical moment,  $M(\lambda, q)$ :

$$M(\lambda, q) = \sum_i \varepsilon(\lambda, i)^q \quad (1)$$

For a scaling field, the moment,  $M(\lambda, q)$ , relates to the scale ratio,  $\lambda$ , as follows:

$$M(\lambda, q) = \lambda^{\tau(q)} \quad (2)$$

where  $\tau(q)$  may be regarded as a characteristic function of the fractal behavior, and is given by the slope of the  $\log M(\lambda, q)$  vs.  $\log \lambda$  plot. If  $\tau(q)$  versus  $q$  is a straight line, the time series is mono-scaling. If  $\tau(q)$  versus  $q$  is a convex function, then the time series is multi-scaling (e.g. Frisch and Parisi, 1985; Svensson *et al.*, 1996).

### 3.2. Synthetic Data Generation: Log-Poisson Distribution

For synthetic data generation, at each subdivision (or level) of the cascade model, the total intensity (i.e. rainfall value) at the coarser resolution is multiplied by two random numbers to obtain the intensities at the finer resolution. The distribution of the random numbers  $W$  assumed is log-Poisson [following the encouraging results reported by Onof and Townsend (2004) for UK rainfall data]. More specifically, the distribution is given as:

$$W = A\beta^N \quad (3)$$

where  $P\{N = m\} = c^m e^{-c} / m!$

This apparently leaves three parameters to be identified:  $A$ ,  $\beta$  and  $c$ . However, the conservation of the mean intensity entails that the mean of  $W$  must be equal to 1. This leads to:

$$A = \exp[c(1-\beta)] \quad (4)$$

which leaves two parameters to be identified:  $\beta$  and  $c$ . The random cascade is designed to reproduce the observed properties of the (multi-scaling of the (rainfall) time series, expressed in terms of moments:

$$\ln\{E[I(\Delta t)^q]\} = -\tau(q)\ln(\Delta t) + \rho(q) \quad (5)$$

with  $\tau(q)$ , the moment scaling function, as the *structure function* (i.e. a convex function, if multi-scaling behavior holds).

The calibration then involves fitting the structure function derived from the data to a theoretical form

$$\tau(q) = c[q(1-\beta) + (\beta^q - 1)] / \ln 2 \quad (6)$$

This fitting is carried out by using the least-squares method. This means that, if  $\tau_{obs}(q_i)$  is the observed structure function and  $\tau_{\theta}(q_i)$  is the theoretical one given by the formula above, then the parameters are found with the following minimization program:

$$O = \sum_{i=1,n} [\tau_{obs}(q_i) - \tau_{\theta}(q_i)]^2 \quad (7)$$

subject to:

$$\beta, c \geq 0$$

where  $n$  is the number of values of  $q$  used in the fitting, and  $q_i$  are these values of  $q$ .

## 4. DATA, ANALYSIS AND RESULTS

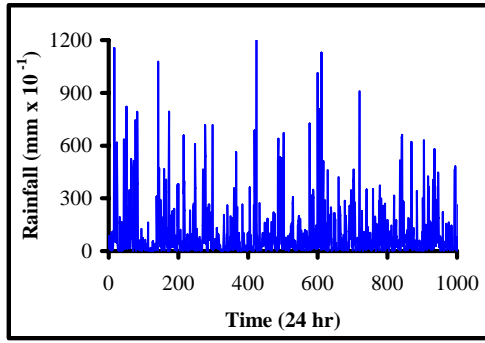
### 4.1. Data

In this study, rainfall data observed at the Sydney Observatory Hill, Sydney, Australia, are studied. Rainfall data are available at 6-min intervals. As a preliminary attempt herein to study the utility and suitability of the scaling-based approach, rainfall data of only four successively doubled resolutions between daily and 3-hr (i.e., daily, 12-hr, 6-hr, and 3-hr) are considered. The presence of scaling in rainfall is investigated both for the rainfall data at the individual scales and for the distributions (weights) of rainfall between them, in order to cross-verify and confirm the results. For the synthetic data generation, however, only the results from the distributions of weights are used, consistent with the log-Poisson assumption above.

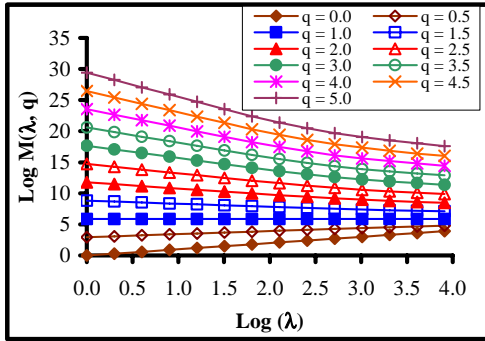
### 4.2. Rainfall Scaling Behavior

In the statistical moment scaling function method, for each rainfall series or weight distributions at each level of the cascade model, as appropriate, the non-overlapping intervals (or box-sizes) used are: 1, 2, 4, ...,  $N/4$ ,  $N/2$ , and  $N$ . Consequently, the scale ratios,  $\lambda$ , are:  $N$ ,  $N/2$ ,  $N/4$ , ..., 4, 2, and 1. For each  $\lambda$ , the  $q^{\text{th}}$  statistical moment is calculated according to Eq. (1). The  $q$  values considered are between 0.0 and 5.0, in steps of 0.5.

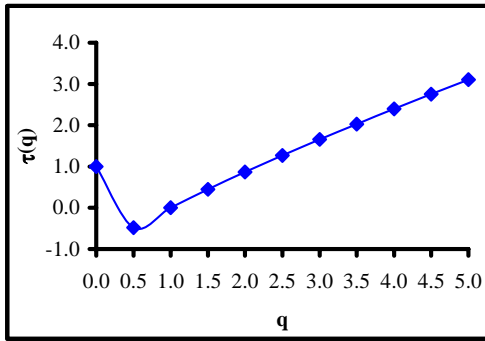
Figure 2, for instance, shows the time series [(a)] and the moment scaling results [(b) and (c)] for the daily rainfall data. As can be seen [Figure 2(b)], the log-log relationship between  $M(\lambda, q)$  as a function of  $\lambda$  exhibits large linear scaling regions for all values of  $q$  [except for higher  $q$ 's that cause divergence of moments], allowing fairly accurate estimation of the slopes,  $\tau(q)$ . The  $[-\tau(q)]$  versus  $q$  function [Figure 2(c)] is a convex curvature, rather than a straight line. This is an indication of the presence of multi-scaling behavior in the daily rainfall series. Similar multi-scaling behavior is observed also for the 12-hr, 6-hr, and 3-hr rainfall series [figures not shown].



(a)



(b)

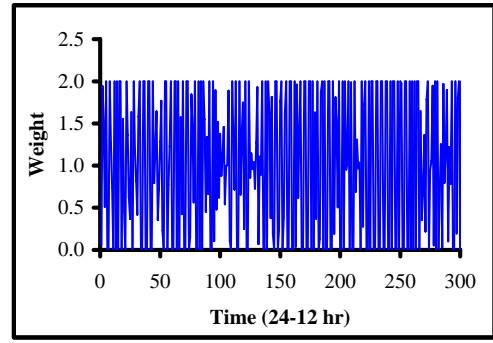


(c)

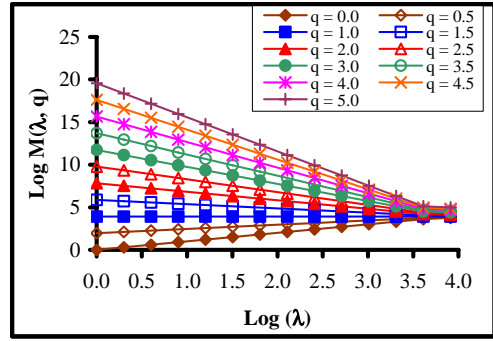
**Figure 2.** Sample Time Series [(a)] and Moment Scaling Results [(b) and (c)] for Daily Rainfall

Figure 3, for instance, shows the time series [(a)] and moment scaling results [(b) and (c)] for the weights between daily and 12-hr scales. The relationship between  $[-]\tau(q)$  and  $q$  [(c)] exhibits a convex curvature, indicating the presence of multi-scaling behavior. Similar behavior is observed also for the weights between 12-hr and 6-hr scales and between 6-hr and 3-hr scales [figures not shown].

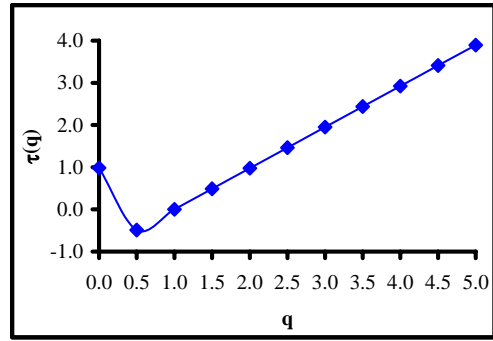
The presence of multi-scaling behavior in all of the time series analyzed, and especially in the distributions of weights, indicates the possibility of obtaining rainfall data at a finer resolution (e.g. 12-hr) from those at a coarser resolution (e.g. daily). With these results, synthetic data that possess the scaling property (moment scaling function) of the observed weights are generated next.



(a)



(b)

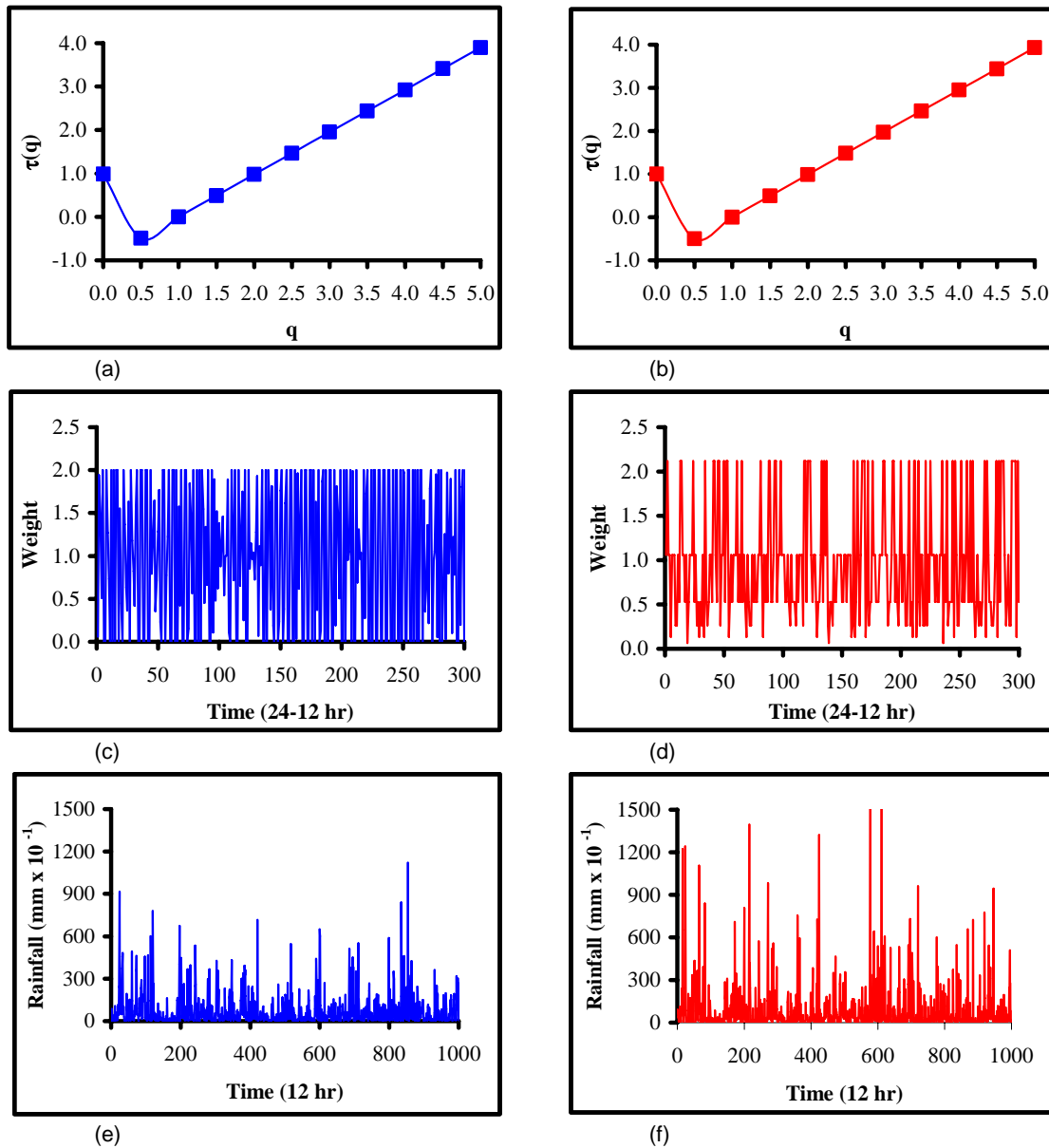


(c)

**Figure 3.** Sample Time Series [(a)] and Moment Scaling Results [(b) and (c)] for Weights Between Daily and 12-hr Rainfall Series

### 4.3. Synthetic Rainfall Data Generation

Fitting the theoretical scaling function [Eq. (6)] to that of the observed weights for the distribution between 24-hr and 12-hr [Figures 3(c)] yields:  $\beta = 0.5$ ,  $c = 1.5$ . Substituting these values in Eq. (4) gives  $A = 2.117$ . With this, synthetic weights are generated according to [Eq. (3)]:  $W = 2.117 * (0.5)^n$ , where  $n$  are the random numbers. In a similar manner, weights are generated also for the two other cases: between 12-hr and 6-hr scales, and between 6-hr and 3-hr scales. For brevity, only the results obtained for the case between daily and 12-hr scales are presented. However, it is also found that the observations made below for this case are also applicable to the other two cases.



**Figure 4.** Observed versus Modeled Data Between Daily and 12-hr Scales: (from top to bottom) Moment Scaling Function, Weight Distribution, and Time Series (at 12-hr)

Figure 4(a) and (b) presents the moment scaling functions for the observed and the generated weights, respectively, for the daily to 12-hr scales. The functions match very well, qualitatively and quantitatively. This seems to indicate the utility of the log-Poisson distribution approach for synthetic data generation, with the moment scaling function as a characteristic. While these results are encouraging, one must also recognize that matching the scaling function alone may not be sufficient to “replicate” the “observed” data.

To verify this, comparisons are also made using direct time series plots of weight distributions [Figure 4(c) and (d)] and of rainfall data at the 12-

hr scale [Figure 4(e) and (f)]. While the synthetically generated data seems to possess the “random” distributions present in the observed data, including the “peaks” and “dips”, a closer look at the plots tells a different story. Some conspicuous observations to this end are as follows: The observed weights have a maximum value equal to 2, a minimum value equal to 0, and over 30% of values as zeros. On the other hand, the modeled weights have a maximum value of 2.117, a minimum value of 0.002 (and obviously no zeros at all). This situation is essentially due to the log-Poisson distribution used. It is obvious that these discrepancies also eventually influence the rainfall values [Figure 4(e) and (f)].

## 5. CONCLUSIONS

This study tested the utility of a random cascade scaling-based approach for rainfall disaggregation. The methodology involved identification of scaling behavior in rainfall, followed by synthetic generation of data consistent with this scaling behavior. Application of the approach to rainfall observed in Sydney indicates its potential, but the results also suggest caution, because: (1) fitting the scaling function alone does not necessarily mean fitting the data; and (2) certain existing methods for synthetic data generation (e.g. log-Poisson distribution) may not suit the observed data, especially in generating zeros. Two other limitations that exist in the use of scaling identification methods also need attention: (1) most methods (e.g. statistical moment scaling) are useful only for distinguishing between mono- and multi-scaling, but not between scaling and non-scaling; and (2) since the methods generally require  $2^n$  number of data, some of the available data may go unused. Future research will address these issues.

## 6. REFERENCES

- Arnold, J.G., and J.R. Williams (1989), Stochastic generation of internal storm structure, *Transactions, ASAE*, 32(1), 161-166, 1989.
- Bo, Z., S. Islam, and E.A.B. Eltahir (1994), Aggregation-disaggregation properties of a stochastic rainfall model, *Water Resources Research*, 30(12), 3423-3435.
- Connolly, R.D., J. Schirmer, and P.K. Dunn (1998), A daily rainfall disaggregation model, *Agricultural and Forest Meteorology*, 92, 105-117.
- Cowpertwait, P.S.P., P.E. O'Connell, A.V. Metcalfe, and J.A. Mawdsley (1996), Stochastic point process modeling of rainfall. II. Regionalization and disaggregation, *Journal of Hydrology*, 175, 47-65.
- Frisch, U., and G. Parisi (1985), On the singularity structure of fully developed turbulence, In: Ghil, M., R. Benzi, and G. Parisi (Eds.), *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, North-Holland, New York, 84-88.
- Glasbey, C.A., G. Cooper, and M.B. McGechan (1995), Disaggregation of daily rainfall by conditional simulation from a point-process model, *Journal of Hydrology*, 165, 1-9.
- Gupta, V.K., and E. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *Journal of Applied Meteorology*, 32, 251-267.
- Heneker, T.M., M.F. Lambert, and G. Kuczera (2001), A point rainfall model for risk-based design, *Journal of Hydrology*, 247(1-2), 54-71.
- Hershendorff, J., and D.A. Woolhiser (1987), Disaggregation of daily rainfall, *Journal of Hydrology*, 95, 299-322.
- Lambert, M.L., and G. Kuczera (1998), Seasonal generalized exponential probability models with application to interstorm and storm durations, *Water Resources Research*, 34(1), 143-148.
- Mandelbrot, B.B. (1974), Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, *Journal of Fluid Mechanics*, 62, 331-358.
- Menabde, M., and M. Sivapalan (2000), Modeling of rainfall time series and extremes using bounded random cascades and Levy-stable distributions, *Water Resources Research*, 36(11), 3293-3300.
- Olsson, J. (1998), Evaluation of a scaling cascade model for temporal rainfall disaggregation, *Hydrology and Earth System Sciences*, 2(1), 19-30.
- Olsson, J., and R. Berndtsson (1998), Temporal rainfall disaggregation based on scaling properties, *Water Science and Technology*, 37(11), 73-79.
- Onof, C., and J. Townend (2004), Modeling 5-minute rainfall extremes, In: *Hydrology: Science & Practice for the 21st Century*, British Hydrological Society, pp. 377-388.
- Over, T.M., and V.K. Gupta (1994), Statistical analysis of mesoscale rainfall: dependence of a random cascade generator on large-scaling forcing, *Journal of Applied Meteorology*, 33, 1526-1542.
- Perica, S., and E. Foufoula-Georgiou (1996), Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions, *Journal of Geophysical Research*, 101, 26347-26361.
- Rodriguez-Iturbe, I., D.R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proceedings of the Royal Society of London A*, 410, 269-288.
- Schertzer, D., and S. Lovejoy (1987), Physical modeling and analysis of rain and clouds by scaling multiplicative processes, *Journal of Geophysical Research*, 92, 9693-9714.
- Sivakumar, B., S. Sorooshian, H.V. Gupta, and X. Gao (2001), A chaotic approach to rainfall disaggregation, *Water Resources Research*, 37(1), 61-72.
- Svensson, C., J. Olsson, and R. Berndtsson (1996), Multifractal properties of daily rainfall in two different climates, *Water Resources Research*, 32(8), 2463-2472.