JP2.8

IMPROVED ACCURACY IN INFERRING THE FINE SCALE PROPERTIES OF RAIN RATE FROM RAIN GAUGE TIME SERIES

Paul D. Baxter* and Kevin S. Paulson
University of Leeds, UK and University of Hull, UK

1 INTRODUCTION

This represents interdisciplinary work (involving statisticians, meteorologists and engineers) conducted as part of the RAINMAP initiative (http://www.rainmap.rl.ac.uk).

RAINMAP is a UK research council funded network that aims to foster communication and collaboration between the diverse communities interested in developing and applying models of rain rate variation. Many applications in telecommunications, urban drainage design, pollution transport modeling and erosion control require stochastic models of rain rate variation that statistically reproduced measured rain properties over a very broad range of spatial and temporal scales. Scales as short as seconds and a few tens of metres are important. Information on large-scale variation is available from hydrological pulse models, radar databases and numerical models (NWP). Techniques have been proposed to stochastically downscale this information to much finer scales. Combining large-scale information with downscaling potentially provides broad scale models with wide application and high commercial value.

Disaggregation algorithms aim to produce ensembles of rain rate series, with short integration times, consistent with a time-series measured with longer integration times and some a priori known statistics. To develop and test disaggregation algorithms, it is necessary to have rain rate data measured at the finest scale considered.

Rapid response rain gauges are devices for measuring rain rates with integration times typically less than one minute and often as short as ten seconds. A widely used type of gauge measures the rain water collected in a funnel by forming equally sized drops and detecting their fall. The data recorded from the gauge is the number of drops detected in a 10 second integration time. At the end of each integration period, a partially formed drop will be present within the gauge. For the gauges operated by Rutherford Appleton Laboratory, RAL, (Oxford, UK), each drop in a 10 s period corresponds to a collected rain height of 0.004 mm.

This work develops a downscaling algorithm capable of producing ensembles of rain rate time-series, with integration times as short as 10 seconds, consistent with a time-series of rain rates with integration times as long as 6 hours. The algorithm is based on a stochastic multiplicative cascade using beta distributions as the random generator. The parameters are estimated from one gauge-year of rain gauge data, with a 10 second integration period, collected in the Southern UK. The statistical moments up to third order, of nine gauge-years of data, are calculated for integration times in the range 10 seconds to 6 hours. These data are compared with time-series derived by accumulating data at 10 second scale to larger integration times, and then downscaling using the proposed algorithm.

2 MOMENT SCALING STRUCTURE FUNCTION

The identification of simple or multi-scaling ranges of rain rate variation provides a useful summarising statistic and suggests a number of modelling algorithms. Analysis methods identify ranges of scales where the statistical moments of rain rate are a power law function of the size of the integration interval. The summarizing statistic used as a basis of modelling is the moment scaling structure function. Let $S_q(\lambda)$ be the qth moment of R_λ, the rain rate measured over an integration volume of scale λ i.e. $S_q(\lambda) = \mathrm{E}(R_\lambda^q)$ where $\mathrm{E}(.)$ is the expected value. If $S_q(\lambda) \propto \lambda^{\zeta(q)}$, where $\zeta(q)$ does not depend upon λ, over some range of scales, then the rain rate is said to exhibit scaling. For simple scaling $\zeta(q)$ is linear in q, otherwise it is known as anomalous scaling. The function $\zeta(q)$ yields the multifractal exponents.

Paulson (2002) shows that log rain rate is simple-scaling for scales as short as 10 s and 300 m, and is well modelled as an isotropic, homogeneous, fractional Brownian process with Hurst coefficient $H = 1/3$. This assumption allows the simulation of rain fields and time-series with a model we will call FBH1/3. A log-rain-rate time series can be modelled as a Gaussian sequence with a power-law spectral density with an exponent of -5/3. The rain rate while raining time series can then be calculated by shifting and scaling the log rain rate sequence to the desired mean and variance before exponentiation.

*Corresponding author address: Paul D. Baxter, Univ. of Leeds, Dept. of Statistics, Leeds, LS2 9JT, UK; email: pdbaxt@maths.leeds.ac.uk.

3 GAUGE DATA

Data spanning three complete years, 2000 to 2002, from 3 RAL Rapid Response Drop-Counting Gauges will be used. Two gauges are located at Chilbolton Observatory in the Southern UK, while the third gauge is situated 9 km away at Sparsholt.

The rain accumulation, or rain height, occurring between time $t_0 = 0$ and t is related to the notional instantaneous rain rate, $r(\tau)$ by:

$$H(t) = \int_{\tau=0}^{t} r(\tau)\mathrm{d}\tau. \qquad (1)$$

A drop is detected by the gauge whenever an integer multiple of the rain height quantisation ΔH is reached. The number of drops detected up to time t is:

$$N(t) = \lfloor H(t)/\Delta H \rfloor, \qquad (2)$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x. The measured rain height time-series, recorded at intervals of ΔT is:

$$H_{rg}(i\Delta T) = N(i\Delta T)\Delta H = H(i\Delta T) + \varepsilon, \qquad (3)$$

where ε is a quantisation error which, for $N(i\Delta T) > 0$, is uniformly distributed over the interval $(-\Delta H, \Delta H)$. The magnitude depends on the size of partially formed drops at the beginning and end of the measurement period. Over any interval $[t, t+\Delta t]$ the measured average rain rate is:

$$R_{\Delta t}(t) = \frac{H(t+\Delta t) - H(t)}{\Delta t}. \qquad (4)$$

When the number of drops measured over an interval is small, the relative error in the derived rain rate, due to gauge quantization, is large. For rain rates less than $\Delta R = \Delta H/\Delta T$ the derived rain rate time-series oscillates between zero and ΔR. These rain rates occur for a high percentage of time and so the quantisation artifact leads to large errors in the calculated autocorrelation and power spectral density. Other second order statistics, such as event duration distributions, will also be distorted at low rain rates.

4 MULTIPLICATIVE CASCADE

In this section we develop the statistics of a multiplicative cascade model (MCM) of rain rate averaged over intervals that are halved at each level of the cascade. Each iteration of the cascade yields a rain rate time-series of measurements with half the integration time and consistent with all longer integration time measurements higher in the cascade.

The cascade begins with a rain rate time-series of measurements averaged over intervals of length T_0. This is equivalent to knowledge of the total rain accumulation $H(t_i)$ at times $t_i = iT_0$, relative to t_0 when measurements began. The cascade proceeds by estimating the rain accumulation in the middle of each interval as:

$$H(t_i + T_0/2) - H(t_i) = w[H(t_i + T_0) - H(t_i)], \qquad (5)$$

where $w \in [0,1]$ is an independent and identically distributed sample from a random generator with PDF $G_{T_0}(w)$. The result is a rain rate time-series of measurements averaged over intervals of length $T_1 = T_0/2$, and consistent with the original, coarser scale time-series. The cascade can be iterated, using a random generator $G_{T_n}(w)$, producing consistent time-series of measurements averaged over intervals of length $T_n = 2^{-n}T_0$. If the average rain rate over the interval T_n is R_n, then the average rain rates over the two half-intervals in the next level of cascade are $2wR_n$ and $2(1-w)R_n$.

We propose a functional form (6) for the generator distribution , based on distributions produced from rain gauge measurements, see Figure 1.

$$\begin{aligned} G_{T_n}(w) &= \gamma_{T_n}[\delta(w) + \delta(w-1)] \\ &+ (1-2\gamma_{T_n})\mathrm{B}(\alpha_{T_n}, \alpha_{T_n}; w), \end{aligned} \qquad (6)$$

where $\delta(w)$ is the delta function and $\mathrm{B}(\alpha, \beta; w)$ is a beta distribution with PDF:

$$\mathrm{B}(\alpha, \beta; w) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w_{\alpha-1}(1 - w^{\beta-1}). \qquad (7)$$

The first term in (6) yields $w = 0$ and $w = 1$, each with probability γ_{T_n}, when the rain occurs exclusively in one half of the interval. In this application, symmetric beta distributions will be used with parameters $\alpha = \beta$ and we will use the notation $\mathrm{B}(\alpha; w) = \mathrm{B}(\alpha, \alpha; w)$.

5 PARAMETER VARIATION

To downscale a rain rate time-series, it is necessary to know the generator distribution at each halving of the integration period. Each generator distribution is defined by two variables; γ_T, the probability of all the rain falling in the first half of the measurement period, and α_T the beta distribution parameter.

In practice, the moment scaling function, $S_q(t)$, of a measured time-series is calculated as samples are accumulated to coarser scales. As a time-series generated by a multiplicative cascade is accumulated to calculate longer integration time measurements, the PDF and moments at different accumulation scales are not those present in the original cascade. Therefore, the parameters γ_T and α_T cannot be directly estimated from accumulated rain gauge data. However, the variation of parameters estimated directly from data suggests functional forms for parameter variations, and this is demonstrated in this section.

The distribution of rain height fractions w can be estimated from gauge data using:

$$w_i = \frac{H(t_0 + T_{n-1}) - H(t_0)}{H(t_0 + T_n) - H(t_0)},$$

$$\approx \frac{N[(i+2^{n-1})\Delta T] - N(i\Delta T)}{N[(i+2^{n})\Delta T] - N(i\Delta T)}. \quad (8)$$

The estimates w_i are unbiased but values that are ratios of small integers e.g. 1/2, 1/3, 2/3 etc. occur disproportionately due to the quantization introduced by the gauge. Figure 1 illustrates the cumulative probability functions (CDF) of w calculated from one gauge-year of data, with $w_i = 0$ and $w_i = 1$ excluded. Also plotted is the beta distribution CDF with parameters derived by maximum likelihood estimation from the estimates w_i.

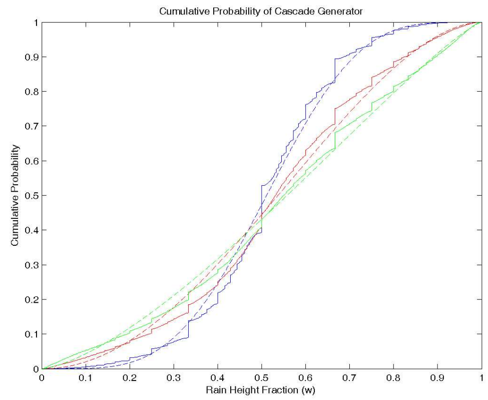

Figure 1: Cumulative probability functions of rain height fraction w ($w \neq 0$ and $w \neq 1$) for three integration periods: 200 s (blue) 1000 s (red) and 2000 s (green). The solid curves are derived from rain gauge measurements while the dashed curves are the maximum likelihood beta distributions

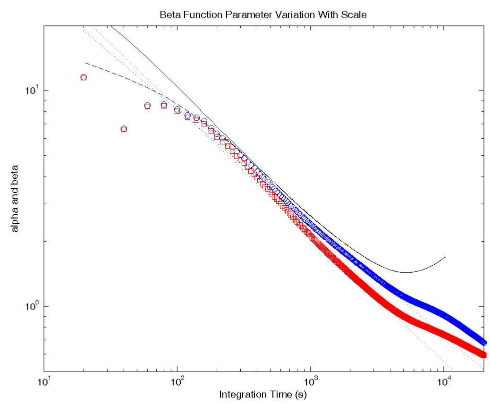

Figure 2: Variation of the measured rain gauge beta function parameters, α_T (circles) and β_T (squares), with scale. The dotted lines indicate the best fit power laws to the measured data. The solid line is the alpha and beta parameters for the FBH1/3 model while the dashed line is simulated gauge data based on the FBH1/3 time-series

Figure 2 illustrates the variation of the beta function parameters α_T and β_T with accumulation scale. Nine gauge-years of gauge data were accumulated to integration times that were even multiples of 10 s up to 10^4 s. Rain height fractions were calculated using (8) and beta functions we fitted using maximum likelihood estimation. Also plotted are the parameters for simulated FBH1/3 data. Simulated rain gauge time-series are calculated from the FBH1/3 time-series using (2) and (3) to generate $H_{rg}(t)$ from $H(t)$. Due to the symmetry in the beta distribution parameters $\alpha = \beta$, only one line is visible for each of the simulated rain rate and rain gauge sequences. The simulated rain gauge time-series has a similar flattening of the alpha and beta curves below 200 s, as the curves derived from measured rain gauge data. This suggests that the flattening is an artifact due to gauge quantization. The FBH1/3 model produces continuous rain, with no zero rain rates, and has a scale break at 3000 s linked to the size of continuous rain events. These lead to the deviation from power-law variation as this integration time is approached. Slight asymmetry exists between α_T and β_T derived from measured data. This is consistent with slightly more rain falling in the latter half of periods suggesting that the onset of rain may be more gradual than the end of rain events. However, for this initial investigation, we will assume that $\alpha_T = \beta_T$ and that this parameter follows a power-law over the range 10 s to 10800 s (approximately 3 hours).

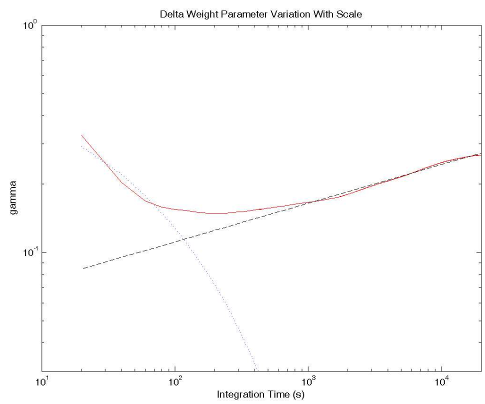

Figure 3: Variation of the γ_T parameter from the measured rain gauge data (solid) and the simulated gauge data based on the FBH1/3 time-series (dotted). The dashed line is a power-law fit to large scale variation of measured γ_T

Figure 3 illustrates the variation of γ_T estimated from the measured rain height fraction distributions and from simulated rain gauge measurements based on FBH1/3 data. Gauge quantisation leads to bias in the estimation of this parameter. For light rain events, and at the edges of all rain events, gauge quantization leads to oscillation between samples with zero and one drops detected, and hence large over-estimation of γ_T. The FBH1/3 model yields continuous rain and so has $\gamma_T = 0$. However, after the gauge simulation introduces quantization to the FBH1/3 sequence, the simulated data yields large values of γ_T for integration periods up to 2000 s. Above this scale the variation is well approximated by a power law.

6 PARAMETER ESTIMATION

The generator distributions of the multiplicative cascade (5) are different from the distributions of the w variable (8) as a time-series is accumulated. In this section the cascade parameters are calculated by minimising the difference between the scaling moments of measured rain rate time-series and time-series formed by downscaling.

Based on observations in Section 5, the parameters γ_T and α_T are assumed to follow power laws i.e.

$$\gamma_T = A_\gamma T^{B_\gamma}, \quad (9)$$

$$\alpha_T = A_\alpha T^{B_\alpha}. \quad (10)$$

One gauge-year of data has been used to estimate the four parameters A_α, A_γ, B_α and B_γ by numerical minimization of the difference between statistics of the original and downscaled time-series. For each set of parameters, the process was repeated 50 times. An error function was defined as the sum of the absolute differences between the second and third moments the original rain gauge time-series $M_q(T_k)$, over the twelve scales $T_k = 2^k \times 10$ s for $k = 0, 1, \ldots, 11$, and $M_q^{DS}(T_k)$ the moments of the downscaled series i.e.

$$\varepsilon(A_\alpha, A_\gamma, B_\alpha, B_\gamma) = \sum_{k=0}^{11} \sum_{q=2}^{3} |M_q^{DS}(T_k) - M_q(T_k)|. \quad (11)$$

The first moment $q = 1$ is not included in the error function as the cascade method conserves rain accumulation during downscaling, and so no error is introduced. Higher order moments could have been included, possibly with declining weights. However, these moments become increasingly sensitive to extreme rain rates, with return times longer than the span of the dataset, and it was felt that this would lead to over-fitting to the data. The accumulation times tested were chosen to span the downscaling scales of interest.

The calculation of the error associated with a set of parameters can be summarized as follows.

1. Calculate the scaling moments of one gauge-year of rain rate data i.e. $M_q(T_k) = \mathsf{E}(R_{T_k})$ for $q = 2, 3$ and for 12 accumulation scales T_k.

2. Accumulate the gauge data to an integration time of 20480 s and then downscale to 10 s using the multiplicative cascade with parameters A_α, A_γ, B_α and B_γ using the cascade defined by (5) based on generator distribution (6) and parameters from (9) and (10).

3. Pass the accumulated, downscaled time-series through a rain gauge simulator to introduce rain rate quantization.

4. Use 50 repetitions of steps 2 and 3 to estimate scaling moments of accumulated, downscaled, quantized time-series.

5. Calculate the error using (11).

Each downscaling cascade yields a different fine-scale rain rate time series consistent with the coarse-scale series, due to the stochastic nature of the cascade. Fifty repetitions was chosen as it yielded an estimate of (11) with a relative error of approximately 5%.

The Nelder-Meads simplex method, Nelder and Mead (1965), was used to minimize (11), starting with initial values based on measured estimates, see Figures 2 and 3. The best fit parameters, for T measured in seconds, were:

$$A_\alpha = 145, A_\gamma = 0.00612, B_\alpha = -0.531, B_\gamma = 0.174. \quad (12)$$

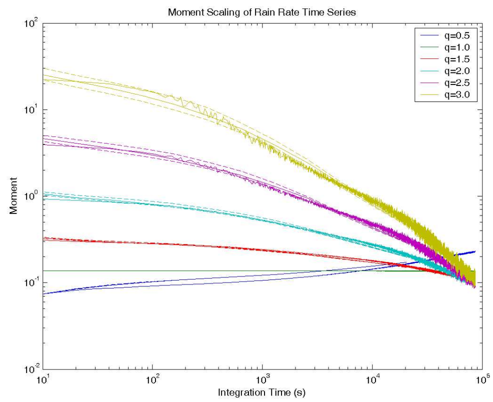

Figure 4: The moment scaling structure function $S_q(T)$ for one rain gauge-year of data (solid lines) and for the simulated rain gauge data based on the time-series downscaled from approximately 6 hours. The upper and lower quartiles from 50 downscaling simulations are indicated by the dashed lines. The curves correspond to moments, from bottom to top, of q=0.5, 1, 1.5, 2, 2.5 and 3

Figure 4 illustrates the variation of the moment scaling structure function $S_q(T)$ for one rain gauge-year of data and for the simulated rain gauge data based on the accumulated, downscaled time-series. For the simulated data the median and upper and lower quartiles of 50 simulations are plotted. Neither the measured nor simulated data is multi-scaling over the range of scales considered, although ranges which are approximately scaling can be identified e.g. 300 to 10000 s. For $q > 1$ the measured statistics lie between the lower and upper quartile of the simulated results. For higher order q the measured moments become increasingly sensitive to small numbers of extreme rain rate values in the time-series. The relatively poor agreement for $q = 0.5$ suggests that the statistics of light rain rates are less well reproduced by the downscaling algorithm.

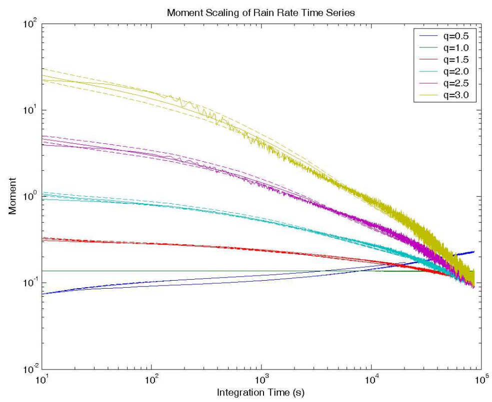

Figure 5: The same statistics as Figure 4 but using all 9 gauge-years of data

The parameters derived from a single gauge-year of data were then used to downscale all the 9 gauge-years of available data. These data include the one gauge-year of data used to derive the cascade parameters. Also included are data from a different site and years to that used to train the cascade. Figure 5 illustrates the variation of $S_q(T)$ derived from all 9 gauge-years of data and compares these with the results from 50 simulated gauge series produced by accumulation and downscaling. For $q > 1$ the measured statistics closely track the upper quartile curves implying higher extreme values of rain rate occur in the data outside the one gauge-year used for parameter estimation.

In a further investigation, the four parameters A_α, A_γ, B_α and B_γ were estimated using each gauge-year of data individually. The purpose of this study was to investigate year-to-year and inter-site variability. The estimated parameters varied by approximately 3% with no consistent inter-year or inter-site pattern.

7 CONCLUSIONS

Nine gauge-years of 10 s integration time rain rate measurements have been analysed to yield the moment scaling structure function over temporal ranges 10 s to 80000 s. At least two multi-scaling ranges are necessary to approximate the moment scaling function well.

A multiplicative cascade disaggregation algorithm has been developed from the rain gauge data. A random variable w has been defined, equal to the proportion of rain that falls in the first half of an accumulation period. After examination of the data, a beta distribution has been chosen as the generator distribution. Measured distributions have been shown to be slightly asymmetric due to non-stationarity of the rain rate time-series. The parameters of the beta distributions have been calculated, allowing rain rate time-series to be disaggregated within the scale range 10 s to 20000 s, while reproducing measured moments up to third order with reasonable accuracy.

The downscaling method developed in this work was aimed at taking the finest scale data available and introducing detailed variation down to 10 s. Although the method has been tested up to periods as long as six hours, the algorithm becomes site specific when the coarse data has an accumulation period much longer than event durations. The relative proportion of convective and stratiform events depends strongly on seasons and the local climate. When the initial rain accumulation data is sufficiently coarse as to have lost this distinction, then the incidence of short-intense and long-moderate events is determined by the downscaling algorithm. The algorithm described has been trained to Southern UK conditions. When the initial rain data is sufficiently fine to characterise rain events, such as five-minute accumulations, then the downscaling algorithm is likely to have much wider applicability.

Future work will investigate the affects of asymmetric generator distribution and seasonal variation. The assumption that rain height fractions w are independent needs to be tested and, possibly, a correlation structure built into the method. Data from different sites and climates will be tested to determine the applicability of the algorithm and parameters.

8 ACKNOWLEDGEMENTS

We would like to thank the British Atmospheric Data Centre and the Radio Communications Research Unit of Rutherford Appleton Laboratory for acquiring and providing access to the rain gauge data.

9 REFERENCES

J. Nelder and Mead R., 1965, A simplex method for function minimization. *Computer Journal*, **4**, pp. 308,313.

K. S. Paulson, 2002, The second order statistics of the spatial-temporal rain rate random fields, *Radio Science*, **37** (5), pp. 21-1,21-8.