A Comparison of Alternatives for Daily to Sub-Daily Rainfall Disaggregation

A.Pui [1], A. Sharma [1] and R.Mehrotra [1]

[1] *School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia*

Email: a.pui@student.unsw.edu.au

Extended Abstract: Urban hydrology has traditionally used extreme rainfall information from Intensity – Frequency – Duration (IFD) curves to obtain design storms for the purpose of design flood analysis. Design storm approaches pose uncertainty in the estimated flood as the influence of catchment antecedent conditions cannot be properly accounted for, leading to an underestimation of the design flood. Hence, generating continuous series of rainfall using a stochastic procedure that can better reflect such antecedent characteristics presents an attractive alternative. Rainfall disaggregation is one such method. This paper evaluates three rainfall disaggregation methods with different theoretical underpinnings, namely Random Multiplicative Cascades (Microcanonical and Canonical versions), Randomized Bartlett Lewis Model (RBLM) coupled with Proportional Adjusting Procedure, and the Method of Fragments (MOF). These models were used to perform disaggregation from daily to hourly rainfall using 86 years (1916 – 2001) of continuous hourly rainfall observed rainfall data from Observatory Hill weather station, Sydney. A few statistics of interest from a flood design perspective including rainfall variability and intermittency, wet spells and the reproduction of extreme values were investigated. While all models reproduced standard statistics such as mean, variance, lag -1 – autocorrelation and dry proportions well, only the MOF was found to match observed IFD behaviour at hourly scale with the cascade models both underestimating (Canonical) and overestimating (Microcanonical) extreme rainfall. The RBLM poorly simulated low ARI events (see Fig 1). A study of within day wet spell distributions produced results consistent with IFD analyses: i.e. that the Microcanonical and Canonical models both underestimated and overestimated mean spell lengths respectively. The RBLM and MOF models characterized wet spell distribution fairly well. It is suggested that the poor performance of the microcanonical cascade model may be attributed to the unsuitability of assumed probability distribution function.
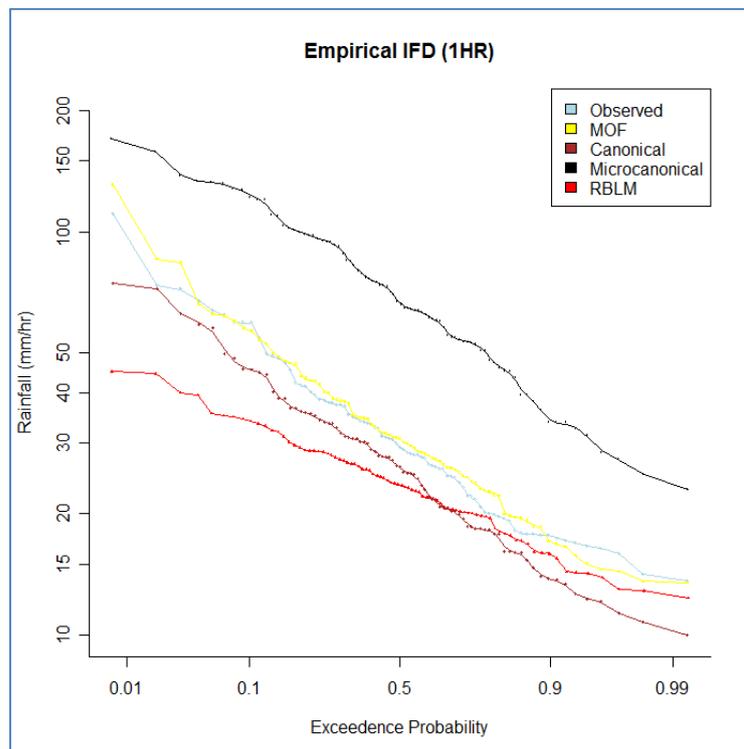
Figure 1: Empirical 1 hour IFD curves for observed and disaggregated rainfall

Keywords: *Rainfall; Stochastic; Disaggregation; Cascades; Bartlett- Lewis; Fragments; Sydney*

1. INTRODUCTION

A wide range of applications involving the planning, design and management of small (especially urban) water resources systems rely on the proper estimation of a design flood. A conventional approach is to ascertain the design rainfall intensity (or design storm hyetograph) from the available Intensity –Frequency – Duration (IFD) curves developed for the region, and convert it to a corresponding flood hydrograph using a rainfall runoff model. However, the IFD based design storm poses uncertainty in the estimated flood as the influence of catchment antecedent conditions cannot be properly accounted for. Additionally, IFD relationships are typically developed using a single maximum rainfall event per year which may not necessarily translate to the corresponding annual maximum flow event for that year depending on catchment antecedent conditions. This non-consideration of antecedent conditions may subsequently lead to an underestimation/overestimation of the design flood (Sharma and Srikanthan, 2006), especially for low return periods. For example, if the catchment is wet before a design event, the resulting flood will be greater than otherwise. Hence, an alternative to this design storm based approach is to use continuous series of rainfall that can better reflect such antecedent characteristics. However, the existence of such data in reality is scarce due to the fact that its measurements are costly and time-consuming.

To address this issue, previous studies have suggested the use of high resolution data generated from available low-resolution (daily) data through a data transformation procedure (Sivakumar and Sharma, 2008; Molnar and Burlando, 2005)**.** A wide variety of models have been proposed in the literature to obtain fine temporal scale rainfall data from a coarser scale, namely from a daily to sub-daily level by way of rainfall disaggregation. Some common rainfall disaggregation models include Random Cascade Models based on scale invariance theory (Menabde and Sivapalan, 2000; Molnar and Burlando, 2005) as well as the Bartlett Lewis or Neyman Scott rectangular pulse models based on point process theory (Rodriguez -Iturbe et al., 1987). Given that there are a host of possible disaggregation methods achieving the same objective of converting lower resolution rainfall to higher resolution rainfall available in the literature with different theoretical underpinnings, it would be useful to compare the results produced by different models in light of the application of different theory behind these models. As such, an attempt is made in the present study to evaluate the utility and suitability of different approaches for disaggregation of daily to sub-daily rainfall.

Here, we evaluate the performance of two variants of a simple disaggregation model based on random cascade theory (canonical and micro canonical versions of the discrete multiplicative random cascades (Molnar and Burlando, 2005; Over and Gupta, 1994) the Randomized Bartlett Lewis Model (RBLM) (Koutsoyannis and Onof 2001), as well as the non-parametric method of fragments based on observed fractions of rainfall (MOF) (Sharma and Srikanthan, 2006), with each model assessed in the context of disaggregation. Cascades based models are chosen for study here in light of the encouraging results earlier studies have reported (Molnar and Burlando, 2005), and also their appeal from a practical viewpoint because they are parameter parsimonious. The RBLM, being a widely applied stochastic rainfall generator, has also formed the basis for the development of other variants. Therefore, possible complications (if any) that arise from this model when tested in the disaggregation context will have far reaching ramifications that extend to other models which apply the RBLM logic. Lastly, the MOF is included to test the performance of a non-parametric model against its parametric counterparts. It is helpful to note that while the random cascades and method of fragments models have been exclusively used as 'disaggregators', the RBLM in its original form was applied strictly as rainfall simulator. The RBLM has since been modified and accorded with an appropriate adjusting procedure to enable it to be applied as a rainfall disaggregator (Koutsoyannis and Onof, 2001).

In this study, we choose our target disaggregation time step as hourly as it is appropriate for flood design purposes even in small catchments with response times or 'times of concentration' less than a day. As this study is primarily conducted to obtain fine resolution rainfall for the purposes of flood design, we are especially interested in knowing how realistically the models simulate rainfall variability and rainfall intermittency as well as the extremes. In addition, a measurement of the sensitivity of estimated parameter values is conducted. This will help to improve our understanding of internal storm structure, the importance and influence of individual parameters, and hence the rainfall phenomenon in general. Gauging each model's ability to realistically simulate wet spells at an hourly time step would provide important information regarding rainfall intermittency properties, as well as reproduction of other standard statistics such as mean, variance, lag -1 –autocorrelation and dry proportions. IFD type analyses will also be performed to ascertain each model's ability to simulate extreme rainfall occurrences. The study is conducted using 86 years (1916 – 2001) of continuous hourly observed rainfall data from Observatory Hill, Sydney with negligible missing data.

2. MODEL DESCRIPTION

2.1 Random Multiplicative Cascades (RMC)

Multiplicative random cascade models originate from turbulence theory and have been increasingly used in rainfall modelling in recent years (Menabde et al., 1997). Although these models are purely phenomenological because the exact relationship between turbulence and rainfall remains unclear (for example, physical processes that produce large areas of uniform rain and small convective showers are intrinsically different and therefore are unlikely to be simply different realizations of a RMC with the same parameters) their ability to reproduce observed structures in rainfall pattern analysis and its statistical properties have justified their continued application (Menabde et al., 1997). According to general multi-fractal theory, once fluctuations at a given scale are understood, those at other scales are deduced from scale invariance (via connecting a common thread through moments at different scales) and need not be independently specified. In particular, this study applies two versions of a larger group of models which operate based on the theory of scale invariance. These models were chosen due to good results reported in Molnar and Burlando (2005) when applied to temperate climate of Zurich, Switzerland. The section below summarizes the basic methodology of the models:
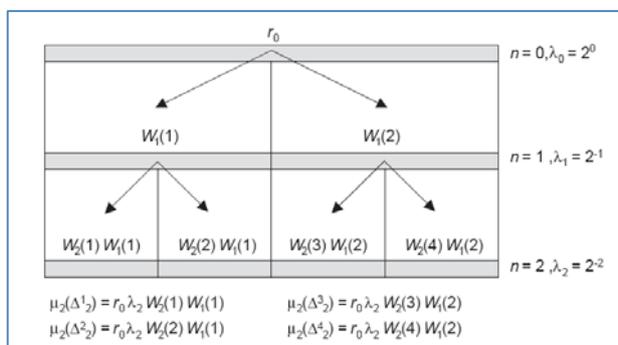

Figure 2: Framework of canonical RMC with branching number $b = 2$ and cascade generator W for scales $n = 0, 1$, and 2. r_0 denotes , rainfall amount at scale 0, (i.e. daily rainfall) (Molnar and Burlando, 2005)

The canonical RMC model distributes rainfall on successive sub divisions (see Fig 2) with b as the branching number. As such, the ith interval after n levels of subdivision is denoted as Δ_n^i. The dimensionless scale is defined as $\lambda_n = b^{-n}$. The distribution of mass then occurs via a multiplicative process through all levels, n of the cascade, such that the mass, μ , in sub division Δ_n^i is:

$$\mu_{(\Delta_n^i)} = r_0 \lambda_n \prod_{j=1}^{n} W_j(i) \quad for \quad i = 1, 2, \dots b^n ; \quad (1)$$

Where, r_0 is the initial rainfall depth (in our case, daily rainfall) at n = 0, and $W_j(i)$ (hereafter denoted as just W for 'weights') is a range of weights that essentially forms the cascade generator (Molnar and Burlando, 2005). W is treated as an independent and identically distributed (iid) random variable, with the important condition that $E(W) = 1$ so that mass is conserved, on average through all levels of the cascade although weights at each time step are not constrained to add up to unity. Properties of W can be estimated from the moment scaling function behaviour across all scales of interest. In addition, this study assumed a Bernoulli distribution for intermittency and lognormal distribution for controlling rainfall amounts. As such, the intermittent lognormal - β – model (Gupta and Waymire, 1993) was applied. The intermittent lognormal - β – model weights generator can be expressed as $W = BY,$ where B (the intermittency factor) controls intermittency in rainfall and Y controls rainfall amounts. In order to divide rainfall into rainy and non-rainy portions, the intermittent section of the model is based on the following probabilities derived according to the Bernoulli distribution with parameter β. Variability in the positive part of the generator (i.e. where $B > 0$) is derived from the lognormal distribution with parameter σ^2 to characterize the variance of Y.

The micro canonical cascade model by definition conserves mass exactly at each cascade level. Hence, for a cascade with branching order (b) of two, the micro canonical weights (M) must either equal to a combination of [0, 1] when intermittency arises (hereafter intermittent part), or [M, 1-M], where $0 < M < 1$ when there is no intermittency (hereafter variability part). The microcanonical cascade also differs from its canonical counterpart because its weights (M) are scale dependent and not iid.

The intermittency parameter (p), and variability parameter (a) are estimated from the breakdown coefficients, which are defined as the ratio of rainfall of a random field averaged over different scales to account for decrease in variance with decrease in timescales (see Menabde and Sivapalan, 2000). More specifically, p represents the probability that one of the intervals in disaggregation is dry for scales between n and $n + 1$ and therefore is simply estimated from historical rainfall. a is estimated from the single parameter (symmetrical) *Beta* distribution (Molnar and Burlando, 2005):

$$f(M) = \frac{1}{B(a)} \cdot M^{1-a} (1-M)^{1-a} \qquad\qquad (2)$$

where $B(a)$ is the *Beta* function with mean $E(M) = 0.5$ and $Var(M) = 1/(8(a+0.5))$

Since both a and p have been known to be scale dependent (Menabde and Sivapalan, 2000), we have also fitted the scale dependent behavior of these parameters by power functions with scale (not shown here).

2.2 Randomized Bartlett Lewis Model (RBLM)

The Bartlett Lewis Model (BLM) is a point process model originally developed by Rodriguez-Iturbe et al (1987) that represents rainfall events as clusters of rain cells where each cell is considered a pulse with a random duration and random intensity. The logic of the original BLM model produces a sequence of storm cells each having its own intensity and duration. The original BLM process was characterized by 5 parameters $(\lambda, \beta, \gamma, \eta, \mu_x)$ as follows: The generation of a "storm origin" is assumed to follow a Poisson distributed rate process having an expected value of λ. Thus the expected value of the time between storm origins is $1/\lambda$. Each storm origin will generate a variable number of "storm cells" until a certain time from the storm origin is exceeded, at which time cell generation ceases (see Fig 3). In the original BLM, this time from the storm origin, often called the "generation duration" is taken to be exponentially distributed with an expected value of $1/\gamma$ with $\gamma = \phi * \eta$, where ϕ is a dimensionless parameter. The duration of each cell's contribution is taken to be exponentially distributed with expected value $1/\eta$. Although the η value was a constant in the original BLM, in the RBLM, η values for distinct storms were assumed to be independent variables (hence the name "randomized") from a gamma distribution with shape α and scale $1/v$. In essence, the η parameter was now replaced by α and v. This meant that the mean and variance of cells from different storms now changed randomly from storm to storm. In the RBLM, the first cell is formulated to occur at the time of the storm origin, and additional cells are generated (until the generation duration is exceeded) as a second Poisson process with expected value β, where $\beta = \kappa * v / \phi$. As such, mean time between storm cells is $1/\beta$ and number of storm cells generated within generation duration has a mean of $1 + \beta/\gamma$. The intensity of each cell's contribution (or cell depth) is taken to be exponentially distributed with expected value $1/\mu_x$. While the RBLM has been a widely applied rainfall generator, here it is used to disaggregate based on proportional adjusting procedures (see Koutsoyannis and Onof, 2001).
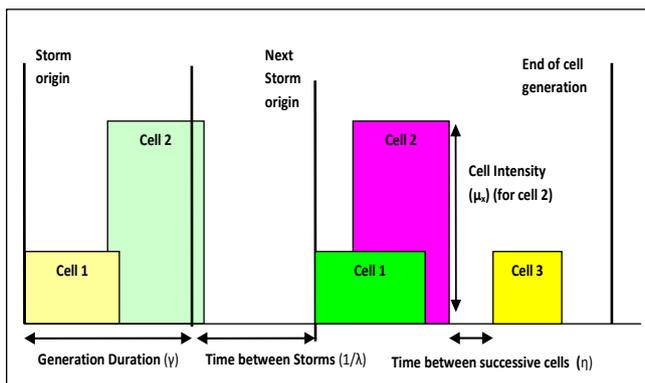

Figure 3: Schematic of the BLM process

2.3 Method of Fragments

This method stands in contrast to the RBLM and Cascades models as it is non-parametric. As such, it makes no major assumptions about the nature of the relationship between continuous and aggregate rainfall. The MOF generates sequences of rainfall that exhibit persistence attributes similar to those observed by maintaining temporal dependence at a daily time scale, and then using non-parametric disaggregation logic to impart dependence to sub-daily time steps (Sharma and Srikanthan, 2006). The methodology of MOF reflects how it represents daily temporal dependence by using high frequency rainfall predictors and longer term attributes such that distributional and seasonality characteristics. This is done by way of resampling a vector of fragments representing the ratio of the sub-daily to daily rainfall. Resampling is performed by via a modified *K*-nearest neighbour algorithm (Sharma and Srikanthan, 2006; Lall and Sharma, 1996) on the basis of the two criteria, the first being that 'within day' fractions are sampled from an 'observation window' that spans 15 days on either side of our day of interest. This would increase the sample size available for sampling purposes and also account for effects of seasonality. For example, if we want to disaggregate daily rainfall on 15th Jan of our daily time series, we look into entire month of January (1-14, 16-30) to find the nearest neighbor in rainfall amounts. The second criterion requires rainfall on our day of interest to satisfy a pre-specified configuration of dry or wet 'states'. This is important because it will help to identify whether our 'wet' day of interest lies at the start, the end or in the middle of a storm. As such, our 'selected'

day from history is expected to be accorded with a more realistic distribution, compatible with our future wet day of interest (Sharma and Srikanthan, 2006).

3. PARAMETER ESTIMATION

3.1 Cascade Models

The canonical model parameters for intermittency and variability, β and σ^2, were estimated based on the relationship between log moments about the origin of order 0 to 4, versus time scale for the entire rainfall record. For further details on this procedure, see Molnar and Burlando (2005). To assess the stability of the parameters, estimation was performed on a seasonal (monthly) basis as well as separately for each year of the record without consideration of seasonal variations. Interestingly, this study found that both parameters β and σ^2, when estimated on a yearly basis fluctuated significantly (see Fig 4). These annual scale fluctuations could possibly be attributed to low frequency signals, something we aim to investigate in future studies. On a monthly basis, the same parameters also hinted at a strong seasonal structure which was curious since Sydney monthly is fairly uniform. The beta distribution produced a poor fit for the variability parameter of the micro canonical model (results not included here). In particular, while the variance of the observed cascade weights was well simulated, its distribution that was highly skewed towards the mean at fine time scales (i.e. 120mins to 60 mins) could not be reproduced.
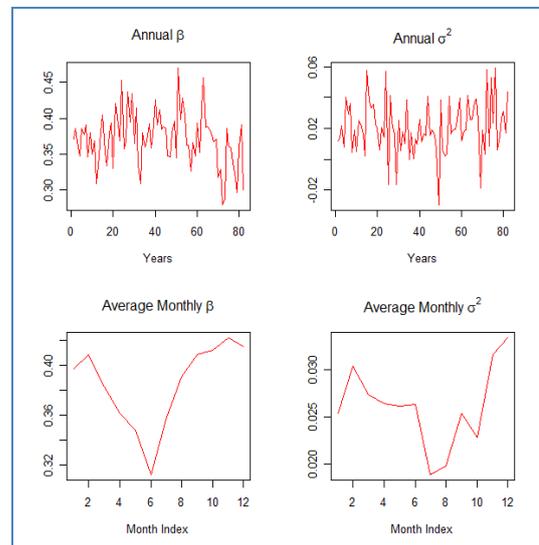

Figure 4: Canonical cascade parameters β and σ^2 showing significant fluctuations when estimated at both annual and monthly basis.

3.2 RBLM

The parameters for RBLM were estimated on a monthly basis, assuming local stationarity within the month according to procedures set out in Koutsoyiannis (2000) where method of moments was used (Rodriguez-Iturbe et al., 1987). The set of parameters to be fitted is given by the set Θ, where $\Theta = \{\lambda, \mu_x, \kappa, \phi, \alpha, \nu\}$. Parameters that control storm generation duration, α and ν were unstable, and resulted in generation of storms with physically unfeasible, excessively long durations. As such, we used an acceptable storm threshold, based on observed storm lengths, to prevent the trawling of 'useless' parameter space that also is flexible enough to allow for possibility of a varying number of acceptable parameter combinations.

4. RESULTS AND DISCUSSION

4.1 Reproduction of Standard Statistics

All the models were generally able to reproduce the mean hourly rainfall well (see Table 1). However, for the hourly variance, the micro canonical model performed poorly compared to the other models. The micro canonical model over estimated the observed hourly variance by a significant margin. It is contended that the single parameter beta distribution that was used to estimate variability in non-intermittent weights is not well suited to Sydney rainfall, thus

Table 1: Reproduction of standard statistics by different models

1 HR	Mean	Variance	lag-1-autocovariance	Dry Proportion
Observed	0.133	0.920	0.484	0.908
RBLM	0.127	0.871	0.501	0.954
MOF	0.127	0.907	0.439	0.913
CANON	0.132	0.865	0.218	0.894
MICRO	0.128	2.398	0.278	0.913

highlighting the importance of choosing an appropriate probability distribution for fitting the data in any parametric model. Dry proportions were well reproduced by all models (with a slight over estimation by the RBLM) except the canonical model.

4.2   Extreme Value Analysis

IFD curves for the 1 and 24 hour durations were constructed based on empirical cumulative distribution of the rainfall series from observed and disaggregated rainfall series. As evidenced in Fig 1, at hourly timescale, MOF most closely resembles observed IFD. The canonical cascades model slightly underestimates extreme rainfall characteristics by underestimating the magnitude of the highest of rainfall intensity for short duration storms of 1 hour. The micro canonical cascades model on the other hand has its IFD curve overestimating rainfall intensities and as such also poorly captures extreme rainfall characteristics. This may be attributed to the inflated variance in rainfall disaggregated by this model. The RBLM also produced an underestimate of the empirical IFD, and was especially poor at simulating events at lower Annual Recurrence Intervals (ARI). All models preserved 24 hour maxima well, which is to be expected since the disaggregation exercise was performed from a daily to hourly scale.
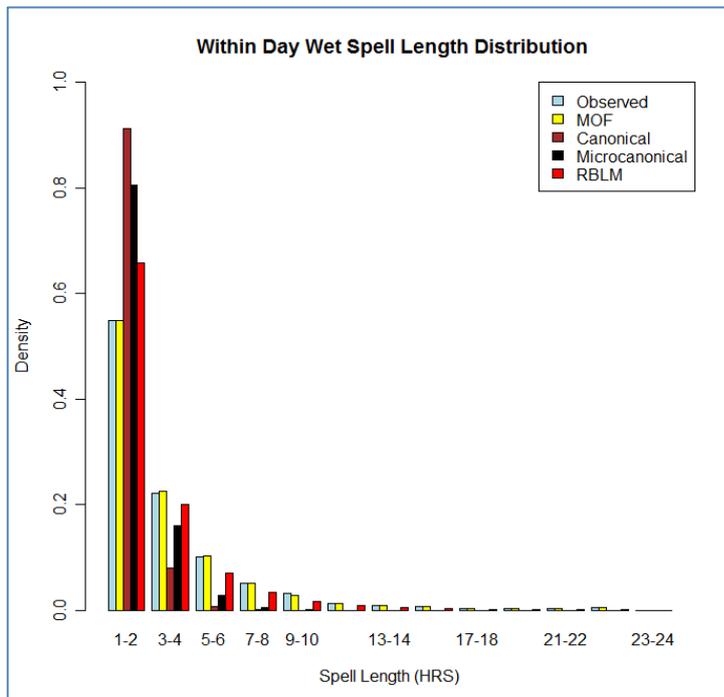
4.3   Wet Spells

Figure 5: Within Day Wet Spell Distribution for disaggregation models compared to observed wet spells.

The generation of realistic wet spells is important from a design flood perspective because rainfall intermittency and persistence need to be quantified to gauge antecedent catchment soil conditions. This cannot be achieved simply by an undertaking an IFD analysis of extreme values. For purposes of this study, a wet spell is defined by consecutive hours of rainfall within a rainy day since we are primarily concerned with gauging the model's performance exclusively as daily to sub-daily disaggregators. The MOF model best captures observed wet spell properties of interest such as the mean spell length per day (an average of all spells of different lengths occurring within a day) and mean number of spell occurrences per day(see Fig 5). Both the cascades models underestimate mean wet spell lengths. The canonical model in particular failed to generate within day wet spells longer than four hours which suggests that it poorly simulates rainfall persistence. This may be caused by its inherent model structure which does not require cascade weights to be conserved at each time step, thus resulting in physically unrealistic generation of zero rainfall values at scale *n + 1* where scale *n* contained rainfall.

5.   CONCLUSION

This paper evaluated the performance of some widely known daily rainfall disaggregators such as the RBLM, two versions of the random cascades model and MOF model using continuous rainfall data of Sydney. In particular, the models were gauged in the areas of reproduction of standard rainfall statistics, extreme values and reproduction of wet spells. In terms of reproduction of observed statistics, the MOF model generally outperformed the other models. This is not entirely unexpected given that the MOF logic operates based on resampling of observed rainfall fractions (at sub-daily time scale) and is therefore expected to produce statistics that bear the closest resemblance to observed data. This model, however, requires a long record of continuous rainfall data. In the absence of such data, parameter based approaches may be employed by estimating parameters from sites of similar climate that have continuous rainfall data. The RBLM also performed better on average than the cascade models albeit with a slightly inflated

reproduction of dry proportions at an hourly scale. However, complications were encountered during parameter estimation stage, and the choice of statistics (and the time scales at which these statistics are gauged) remains subjective. The microcanonical model did not perform satisfactorily when applied to Sydney rainfall, which can be explained by the beta distribution producing a poor fit. As a result, intermittency and rainfall variability were not well captured, and disaggregated rainfall produced shorter than observed wet spells with more intense rainfall. In general, as both cascade models are wholly reliant on cascade weights, the selection of appropriate distribution of the weights is of great importance. In addition, one has to apply parameter parsimonious models such as cascade models with caution: it is contended here that physically different mechanisms (i.e. contrast between temporal distribution of convective and frontal rainfall) are unlikely to be simply the result of drastically different realizations of the same parameters. Secondly, parameter uncertainty may arise from the failure to account for longer term, low frequency modes that influence South East Australian rainfall such as El- Nino Southern Oscillation type events as well as seasonality.

Concluding, with the advent of climate change, we believe that discerning the relative weaknesses and strengths of each model is especially important as there is anticipated to be more frequent applications of the disaggregation methods to downscaled daily rainfall series from global climate models in the future. As such, information regarding the performance of the models is expected to guide hydrologists in applying daily to sub-daily disaggregators with greater insight.

6. Acknowledgements

7. References

Gupta, V.K. and Waymire, E.C., 1993. A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteorol.* 32, 251– 267.

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

Koutsoyiannis, D. and Onof, C., 2000. A computer program for temporal rainfall disaggregation using adjusting procedures. XXV General Assembly of European Geophysical Society, Nice, *Geophysical Research Abstracts*, 2

Lall, U. and Sharma, A., 1996. A nearest neighbor bootstrap for time series resampling. *Water Resources Res*, 32(3): 679-693.

Menabde, M., Sivapalan, M., 2000. Modeling of rainfall time series and extremes using the bounded random cascades and Levy-stable distributions. *Water Resources Res.* 36 (11), 3293– 3300.

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

Molnar, P., Burlando, P., 2005. Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. *Atmos. Res.* 77, 137–151.

Over, T.M., Gupta, V.K., 1994. Statistical analysis of mesoscale rainfall: dependence of a random cascade generator on large scale forcing. *J. Appl. Meteorol.* 33, 1526–1542.

Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1987a. Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London* A 410, 269–288.

Sharma, A., Srikanthan.R, 2006, Continuous Rainfall Simulation: A Nonparametric Alternative, 30th Hydrology and Water Resources Symposium 4 - 7 December 2006, Launceston, Tasmania

Sivakumar, B. and Sharma, A., 2008. A cascade approach to continuous rainfall generation at point locations. *Stochastic Environmental Research and Risk Assessment (SERRA)(*DOI 10.1007/s00477-007-0145-y): 1-9.