University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Anderson, B;Siriwardena, L;Western, A;Chiew, F;Seed, A;Blöschl, G

Title:
Which theoretical distribution function best fits measured within day rainfall distributions across Australia?

Date:
2006

Citation:
Anderson, B., Siriwardena, L., Western, A., Chiew, F., Seed, A. & Blöschl, G. (2006). Which theoretical distribution function best fits measured within day rainfall distributions across Australia?. 30th Hydrology and Water Resources Symposium, 1, (1), pp.1-6. Conference Organising Committee.

Persistent Link:
https://hdl.handle.net/11343/258531

30th Hydrology and Water Resources Symposium
4 - 7 December 2006
Launceston, TAS

Which Theoretical Distribution Function Best Fits Measured Within Day Rainfall Distributions Across Australia?

Brett Anderson[1], Lionel Siriwardena[1], Andrew Western[1],
Francis Chiew[2], Alan Seed[3] and Günter Blöschl[4]
1 Department of Civil and Environmental Engineering, The University of Melbourne
2 CSIRO Land and Water, Canberra
3 Bureau of Meteorology Research Centre, Melbourne
4 Institut für Wasserbau und Ingenieurhydrologie, Technical University of Vienna, Austria

Abstract: Rainfall data at high temporal resolutions is required to accurately model the dynamics of surface runoff processes, in particular sediment entrainment. These processes respond to rainfall intensity variations over short intervals, yet measurement of rainfall intensity at sufficient resolution is available only at a limited number of locations across Australia. On the other hand there is good coverage of rainfall data registered at a daily time step, thus it is desirable to establish a means to estimate within-day distributions of rainfall intensity given the daily rainfall depth and other readily available hydrometeorological data (e.g. temperature, pressure). As a first step towards such a method, an investigation was conducted into the shape of the temporal distribution of high-resolution (6 minute) rainfall intensity within the wet part of rainy days (total rainfall depth > 10mm). This paper quantifies the skill of nine different theoretical distribution functions (TDFs) in fitting characteristics of measured rainfall that are most likely to drive sediment entrainment and transport on hillslopes. Skill is reported by two goodness-of-fit statistics: the Root Mean Square Error (RMSE) between the fitted and observed within-day distribution; and the efficiency of prediction of the 30 minutes of highest rainfall intensity (average intensity of the 5 highest intensity intervals). Four TDFs provided relatively poor fits to higher intensity rainfall (two and three parameter *lognormal*, two parameter *Generalized Pareto* and *Gumbel*), and also showed higher RMSE values. The remaining five TDFs performed equally well for both goodness-of-fit measures. Two of these TDFs are extreme value distributions (*Generalized Extreme Value* and *Weibull*) and in a strict statistical sense should not be applied to within-day rainfall intensity data. On this basis, the remaining three TDFs (*gamma*, *exponential* and the three parameter *Generalized Pareto*) were selected as suitable candidates to represent within-day rainfall distributions in Australia, in particular for hydrological models seeking to estimate runoff and erosion.

Keywords: rainfall intensity, pluviograph, distribution function, disaggregation

1. INTRODUCTION

Rainfall data at high temporal resolutions is required to accurately model the dynamics of surface runoff processes, in particular sediment entrainment. These processes respond to rainfall intensity variations over short intervals. However, measurement of rainfall intensity at sufficient resolution is available only at a limited number of locations across Australia. On the other hand there is good coverage of rainfall data registered at a daily time step, consequently many models used to inform water managers are based on this data. A number of studies have demonstrated that significant improvements in predictive skill can be achieved when higher resolution rainfall data are made available to models (e.g. Dodov and Foufoula-Georgiou, 2005; Kandel et al., 2005; Mertens et al., 2002). The overall goal of this research is to establish a means of estimating the within-day distribution of rainfall intensity given the daily rainfall depth and other readily available hydrometeorological data (e.g. temperature, pressure).

Our intention is to disaggregate a rainy day into wet and dry periods, and to estimate for the wet period the distribution of rainfall intensities. As a first step, we analysed high resolution (6 minute) rainfall data recorded at Bureau of Meteorology pluviometer installations around Australia. This work is the subject of this paper. In particular we focus on the shape of the temporal distribution of rainfall intensities within the wet part of the day.

A wide array of rainfall disaggregation schemes exist, each designed to suit different applications or to reproduce different characteristics of rainfall fields (e.g. spatial structure, temporal event profile). These range from early stochastic approaches, exemplified by point process models (Rodriguez-Iturbe et al., 1987), to the current state-of-the-art models based on multiplicative cascades (Seed et al., 2000). All of these approaches are based in some way on a random field generated according to some theoretical distribution function (TDF). The aim of this investigation was to quantify how well the many available TDFs fit the measured within-day rainfall intensity data, and in particular fit the characteristics of rainfall that are most relevant to runoff generation and erosion.

2 METHOD

High resolution rainfall data from pluviograph stations across Australia was obtained and a detailed analysis conducted to explore the distribution of within-day intensities. There were three stages to the analysis. First, the raw rainfall intensity records were filtered to ensure data quality and to exclude days of small rainfall depth (not of interest for runoff or erosion). Second, seven different theoretical distributions were fitted to the measured cumulative density function (CDF) of rainfall intensity. Multiple methods for estimating the distribution parameter values were employed. Third, a number of objective functions were employed to assess the goodness of fit of the different distributions. Data processing and analysis was achieved via custom routines written in Fortran-90. Each stage is described in more detail in the following sections.

2.1 Data

Pluviograph records were obtained from the Australian Bureau of Meteorology (BOM) from the 42 sites shown in Figure 1. This set of sites (similar to those identified by Lu and Yu (2002)) provides a broad spatial coverage across Australia, record lengths span at least 20 years and the range of mean annual rainfall totals range from 271mm at Giles to 2431mm at Koombooloomba (Lu and Yu, 2002, p.890).

2.2 Quality control and censoring

Rainfall intensity data for each station was supplied with each 24 hour period divided into 240 6-minute intervals (hereinafter referred to as *pluviograph data*). For this analysis, a day was designated as the period starting and finishing at 0900 hours (as per BOM standard). This investigation was concerned only with intra-day characteristics; therefore inter-day relationships could be neglected and periods of record where data was missing were not used rather than being in-filled. A *valid day* for this analysis was defined as one for which the pluviograph record is complete (i.e. 240 values, including zeroes, starting at 0900 hours).

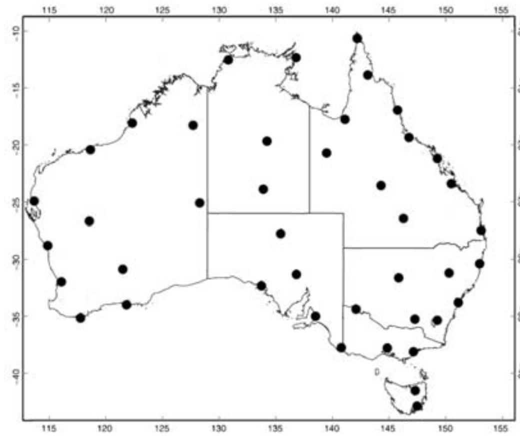

Figure 1. Map showing the location of pluviograph stations.

Records of rainfall intensity measured using tipping bucket technology incur errors at very low rain rates (see review by Nystuen, 1999). As this work was concerned with the upper end of the rainfall intensity spectrum, the valid day pluviograph records were censored in two ways to eliminate low intensity data from consideration. First, only days where the total rainfall depth (*P*) equalled or exceeded 10mm were considered. Second, only those intervals where intensity exceeded a threshold minimum (R_{min}) of 1 mm/hr (0.1mm/6min) were considered in fitting the CDF with each TDF. Finally, in order to numerically resolve the higher order moments (ignoring for the time being the problem of algebraic boundedness (Kirby, 1974)) at least four (distinct) above-threshold intervals (n) were required.

The results of this censorship regime in terms of the number of rainy days on the record and the percentage of the rainfall depth that fell within the various categories is summarised in Table 1. The bottom line describes the data analysed by this investigation, showing for example that in Darwin 12.3% of valid days were sufficiently wet but that on these days 88.2% of the total rainfall depth was received. In contrast, almost half of Melbourne's rainfall depth is delivered on days where the total accumulation is less than 10mm. However, over all stations, the average rainfall depth retained in the data after censorship includes 74.7% of the total rainfall depth which was considered a reasonable result given our interest in processes sensitive to larger events.

Table 1 Summary of data broken into categories based on daily total rainfall depth (P) and the number of 6 minute periods (n) where rainfall intensity exceeded the threshold (1 mm/hr). The number of days and daily rainfall depth (mm) are listed for the stations in Melbourne and Darwin individually, and for all stations combined.

Station:	Melbourne				Darwin				All Stations			
Criteria	Days	(%)	Depth	(%)	Days	(%)	Depth	(%)	Days	(%)	Depth	(%)
P < 0.2mm	10269	(68.3)			11664	(75.5)			473196	(80.9)		
0.2 ≤ P < 10	4056	(27)	12118	(47)	1886	(12.2)	8069	(11.6)	81416	(13.9)	274371	(25.2)
10 ≤ P, n < 4	2	(0.01)	42	(0.16)	7	(0.05)	85	(0.1)	59	(0.01)	841	(0.08)
10 ≤ P, n ≥ 4	700	(4.7)	13626	(52.8)	1894	(12.3)	61156	(88.2)	30418	(5.2)	813781	(74.7)

2.3 Theoretical Distribution Functions

Nine different theoretical distribution functions (TDFs) were fitted to the data (as per list below). The list indicates a short name, the name of the TDF, and the number of distribution parameters (in parentheses). Definitions for each TDF can be found in Stedinger et al. (1993).

- LOGN Lognormal (2)
- LGN3 Lognormal (3)
- EXP Exponential (2)
- GAMA Gamma (2)
- GPT2 Generalized Pareto (2)
- GPT3 Generalized Pareto (3)
- GEV* Generalized Extreme Value (3)
- WEBL* Weibull (2)
- GMBL* Gumbel (2)

The final three TDFs (*) are Extreme Value Distributions (EVD), that specifically represent the distribution of the largest observation drawn from a large sample. The rainfall intensity data is not an extreme value data set. However, recent evidence suggests that heavy rainfall is in fact 'heavy tailed' (Wilson and Toumi, 2005). This is a characteristic of an EVD, and we are specifically interested in the accurate prediction of high intensity rainfall. Therefore, the three EVDs were included to at least determine whether they provide a good fit to the data.

Parameter values for each of the distributions were computed from the pluviograph data in two ways: first via the method-of-moments (product moments); and second by the computation of L-moments (Stedinger et al., 1993). The utility of two other parameter estimation methods (Wang's (1997) LH-moment method and a maximum likelihood estimation technique) were tested on a subset of TDFs. Due to length constraints, in this paper we report only results obtained via the method-of-moments (which gave better results than the L-moment method).

2.4 Assessment of fit

Measures of goodness-of-fit were selected that quantify the rainfall characteristics most likely to drive sediment entrainment and transport on hillslopes. For this paper two measures are reported. First, the Root Mean Square Error (RMSE) between the fitted TDF and the observed rainfall intensity data (for the entire distribution of rainfall above the 1 mm/hr threshold) was computed to show how well the TDF fits all the within-day data. There is therefore one RMSE value per day analysed.

Sediment entrainment is sensitive to high intensity rainfall, so a second statistic was computed to look at how well each TDF reproduced the high end of the distribution. This statistic, called the *30 min intensity*, was the average of the 5 intervals of highest rainfall intensity across the day (i.e. not the highest intensity in a contiguous 30 minute period). Following Legates and McCabe (1999) the goodness-of-fit between the predicted 30 min intensity (from the fitted TDF) and the observed data was quantified using the Modified Coefficient of Efficiency (mCOE). The mCOE is formulated on the magnitude of the difference between observed and fitted data, rather than inflating the importance of outliers by squaring the difference as per the traditional COE. One mCOE value is calculated per station.

3 RESULTS

Indicative results for these two measures for Melbourne and Darwin are shown in Figure 2. Plots are drawn for three TDFs (LGN3, GAMA, and GEV). A number of additional statistics are provided with these plots as described in the figure heading. These charts provide an opportunity to compare the fitting skill of different TDFs at different stations and provide an indication of the merit of using an EVD. Figure 3 summarises the fitting results across all stations, with the skill of each TDF indicated by mCOE and the 90th percentile RMSE value.

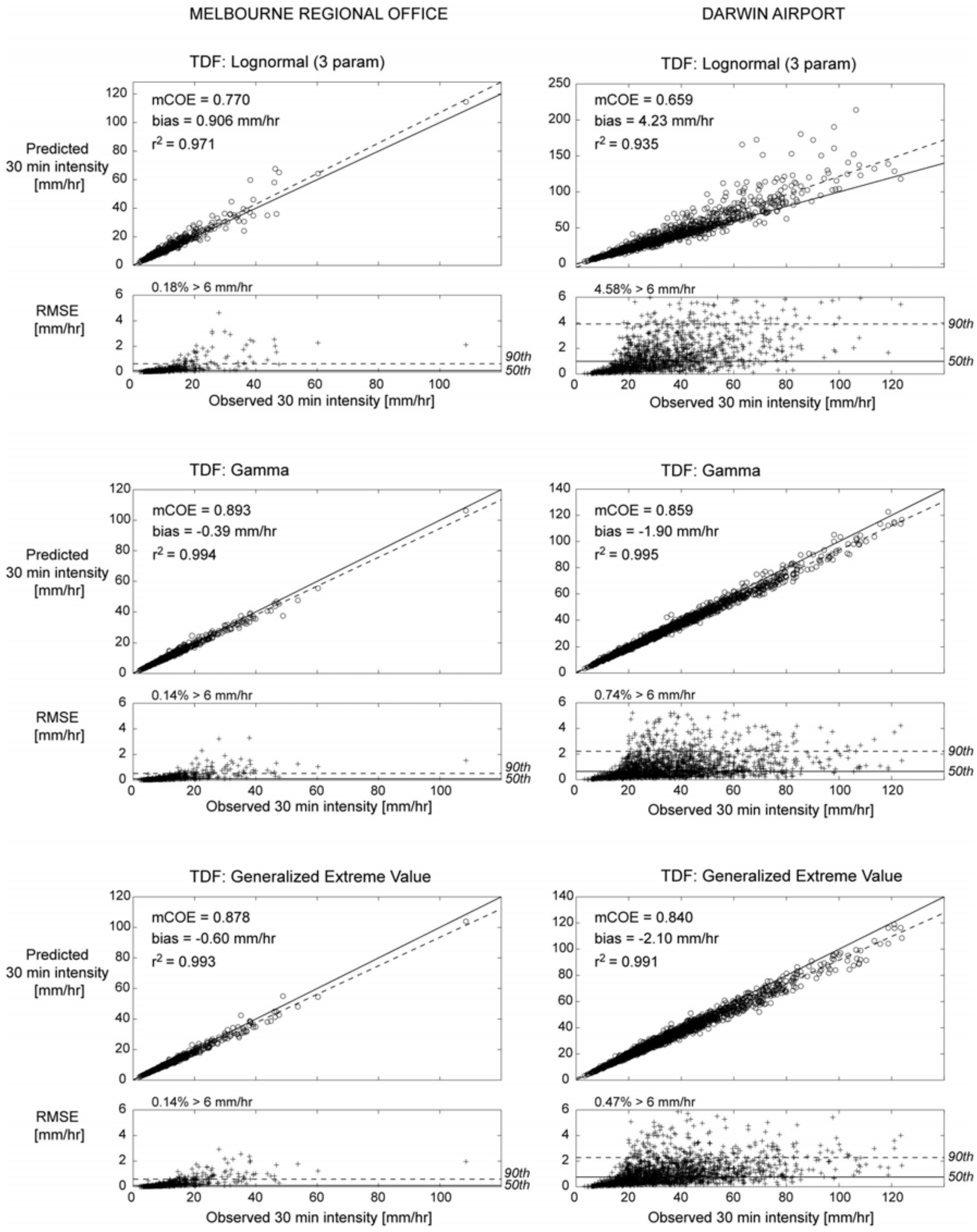

Figure 2. Six paired scatterplots are shown to illustrate the skill of three selected TDF's: three parameter Lognormal (top); Gamma (middle); and the Generalized Extreme Value (lower). The data is for two pluviograph stations: Melbourne on the left and Darwin on the right. The plots are paired to show both the predicted versus observed 30 minute rainfall intensity and the RMSE for each event (also plotted against the observed 30 minute rainfall intensity). Noted on each of the first-mentioned plots are values for mCOE, bias and the square of Pearson's correlation coefficient (r^2), as well as the line-of-perfect-agreement (solid) and the linear regression line (dashed). The latter plots indicate the value of the 50th (solid line) and 90th (dashed line) percentile RMSE values, and also indicate the percentage of RMSE values greater than 6 mm/hr and are hence outside the vertical scale of the plot.

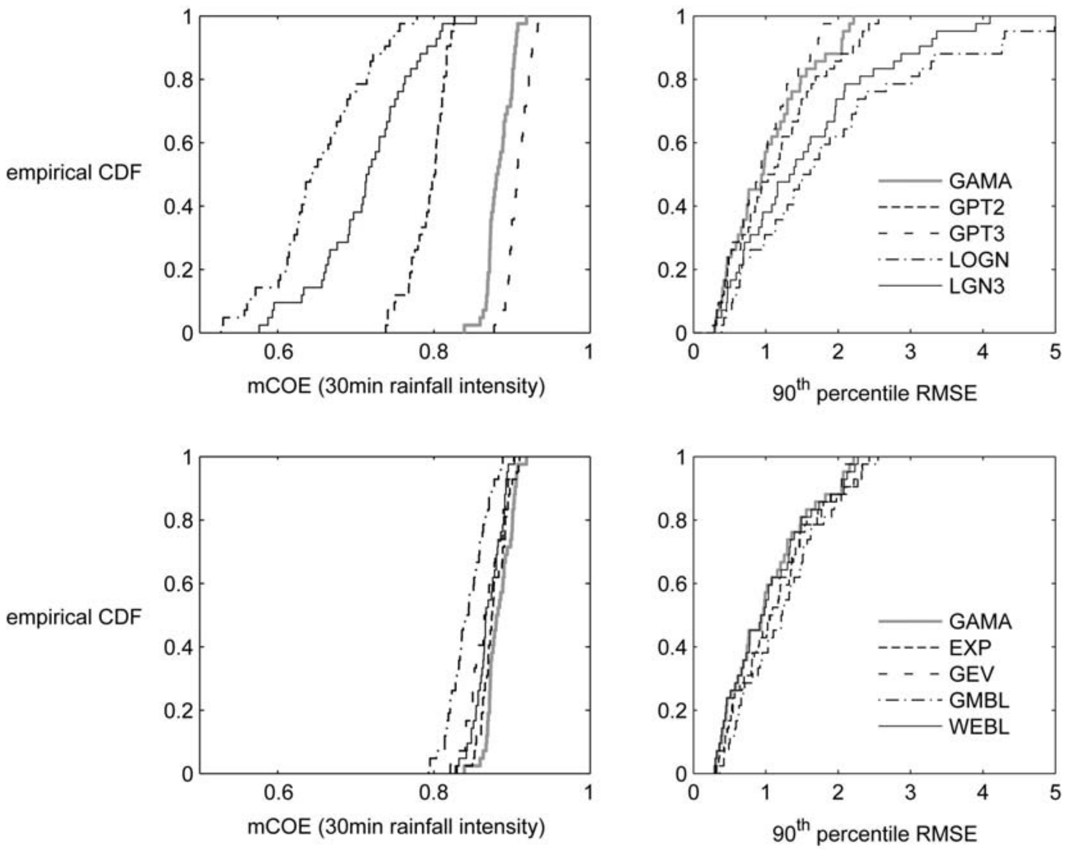

Figure 3. Summary fitting results for all TDFs at all stations. Half of the nine TDFs are shown on the upper two charts, the remainder on the lower charts (with the Gamma DF shown on both). Charts on the left indicate the efficiency (mCOE) of the TDF in fitting the observed 30 minute rainfall intensity, with a good fit indicated by high values. The charts on the right show the 90th percentile RMSE (i.e. 10% of events at the station exhibit a larger RMSE), and the best fitting TDFs have the lowest value. The vertical axis is the estimated (empirical) cumulative distribution function (CDF) constructed from the list of fitting statistics (just described) for each of the 42 stations.

4 DISCUSSION

In Figure 2 the degree of scatter around the line-of-perfect agreement is greater for the lognormal fit than either the gamma or GEV distributions, and this is the case for both Melbourne and Darwin. The mCOE statistics support this observation, with the lognormal statistic more than 10% lower than either other TDF. It is interesting to note however, that the fit given by the RMSE data for Melbourne is not dissimilar to the RMSE fit of the other two TDFs, indicating that the lognormal distribution is only deficient in reproducing the higher rainfall intensities. By contrast, the RMSE fit of the lognormal TDF in Darwin is poor.

The degree to which Darwin events can be fit is poorer for each of the three TDFs shown than the equivalent fit to Melbourne storms. The mCOE values are around 4% lower (gamma and GEV) while the RMSE errors are substantially higher (90th RMSE greater than 2.0 for Darwin but less than 1.0 in Melbourne).

A final point to note from the predicted versus observed plots in Figure 2 is that both the gamma and GEV TDFs tend to slightly underestimate 30 min. intensity for higher observed values of 30 min. intensity. This is indicated by the negative bias and the position of the dashed regression lines being consistently below the line-of-perfect-agreement. Consequently, runoff and erosion would tend to be underestimated.

The data in Figure 2 demonstrate the utility of two principal goodness-of-fit measures: mCOE for 30 min. intensity and the 90th percentile RMSE for the general fit to within-day rainfall intensity. One value of each statistic can be computed for each TDF at each station, and it is these that are plotted in Figure 3 (as empirical cumulative distribution functions).

The lognormal mCOE data is clearly poor at all stations, as is the GPT2 distribution (see top left chart). Indeed the 90th RMSE data on the top left of Figure 3 also support this conclusion. In the lower plot the Gumbel distribution stands out as having the weakest fit for both mCOE and 90th RMSE. On the strength of this evidence these four TDFs were considered poor candidates for representing within-day rainfall intensity.

Of the remaining five TDFs, the three parameter Generalized Pareto distribution stands out for mCOE performance (top left), while the remaining distributions performed equally well (approximately), and did so for both goodness-of-fit measures. Two of these TDFs are extreme value distributions (GEV and Weibull) and in a strict statistical sense should not be applied to within-day rainfall intensity data. On this basis, the remaining three TDFs (gamma, exponential and the three parameter Generalized Pareto) were selected as suitable candidates to represent within-day rainfall distributions in Australia. It should be noted that the gamma and exponential TDFs have only two parameters. This may be a distinct advantage in the next phase of the investigation where relationships are sought between hydrometeorological data and the TDF parameters.

5 CONCLUSIONS

In applications where high intensity rainfall is particularly pertinent, this investigation suggests that the distribution of within-day rainfall intensity be represented by either a gamma, exponential or three-parameter Generalized Pareto distribution. The skill of the GEV and Weibull distributions were found to be of a similar quality, but are not recommended because they specifically pertain to extreme value data which within-day rainfall intensity is not.

6 ACKNOWLEDGEMENTS

This work was supported by an Australian Research Council Discovery Project Grant (DP-0557799).

7 REFERENCES

Dodov, B. and Foufoula-Georgiou, E., 2005. Incorporating the spatio-temporal distribution of rainfall and basin geomorphology into nonlinear analyses of streamflow dynamics. Advances in Water Resources, 28(7): 711-728.

Kandel, D.D., Western, A.W. and Grayson, R.B., 2005. Scaling from process timescales to daily timesteps: a distribution function approach. Water Resources Research, 41: 16pp, W02003, doi:10.1029/2004WR003380.

Kirby, W., 1974. Algebraic Boundedness of Sample Statistics. Water Resources Research, 10(2): 220-222.

Legates, D.R. and McCabe, G.J., 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35: 233-241.

Lu, H. and Yu, B., 2002. Spatial and seasonal distribution of rainfall erosivity in Australia. Australian Journal of Soil Research, 40(6): 887-901.

Mertens, J., Raes, D. and Feyen, J., 2002. Incorporating rainfall intensity into daily rainfall records for simulating runoff and infiltration into soil profiles. Hydrological Processes, 16: 731-739.

Nystuen, J.A., 1999. Relative Performance of Automatic Rain Gauges under Different Rainfall Conditions. Journal of Atmosphere Oceanic Technology, 16: 1025-1043.

Rodriguez-Iturbe, I., Cox, D.R. and Isham, V., 1987. Some models for rainfall based on stochastic point processes. Proceedings of the Royal Society London, Ser. A(410): 269-288.

Seed, A.W., Draper, C., Srikanthan, R. and Menabde, M., 2000. A multiplicative broken-line model for time series of mean areal rainfall. Water Resources Research, 36(8): 2395-2399.

Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E., 1993. Frequency Analysis of Extreme Events. In: D.R. Maidment (Editor), Handbook of Hydrology. McGraw-Hill, Inc.

Wang, Q.J., 1997. LH moments for statistical analysis of extreme events. Water Resources Research, 33(12): 2841-2848.

Wilson, P.S. and Toumi, R., 2005. A fundamental probability distribution for heavy rainfall. Geophysical Research Letters, 32(L14812): 1-4.