Alternative Probabilistic Exponential Distributions For Modelling Rainfall Intensity In Australia

[1]Surawski, N. and [2]B.Yu

[1]CRC for Catchment Hydrology, [2]Faculty of Environmental Sciences, Griffith University, E-mail: n.surawski@student.qut.edu.au

Keywords: *Rainfall intensity; soil erosion; modelling; goodness of fit.*

EXTENDED ABSTRACT

Most runoff and soil erosion models take into account rainfall intensity, since higher storm intensities lead to greater runoff and losses of soil due to erosion. Accurate prediction of soil erosion requires rainfall intensity data with a high temporal resolution. This places a restriction on the simulation of soil erosion, since high temporal resolution data are usually not available. One way to overcome this problem is to synthesise storms using a stochastic rainfall intensity distribution. It has been common for past authors to employ an exponential relationship between rainfall intensity and the frequency of occurrence of that rainfall intensity. The motivation for using an exponential relationship between rainfall intensity and its frequency is based on the log-normal distribution for the recurrence interval between various rainfall depth and duration measurements.

For those authors assuming an exponential relationship between rainfall intensity and its frequency, it has been traditional to formulate the distribution in terms of the amount of time for which rain falls at a particular intensity. The drawback of this approach is that the parameter for this distribution is strongly influenced by rain falling at a low intensity over long periods of time. Clearly this type of rainfall has little impact on runoff and soil erosion rates. An alternative distribution for simulating rainfall intensity has been proposed that considers the depth of rain that falls at a particular intensity, rather than the duration for which rain falls at a particular intensity. The advantage of this formulation for simulating the temporal pattern of rainfall is that this distribution's single parameter is strongly influenced by high rainfall intensity. In addition, parameter values for the proposed distribution can be readily estimated using intensity data collected with tipping bucket rain gauges. The question that motivated this research was which of the two distributions better simulates rainfall intensity data over a range of Australian climates?

It thus follows that the aim of this paper was to test two exponential probabilistic rainfall intensity distributions, one in the time domain and the other in the rainfall domain. Ten sites were considered in this work, namely; Adelaide, Alice Springs, Brisbane, Cairns, Canberra, Darwin, Hobart, Melbourne, Perth and Sydney. The two distributions were compared in terms of goodness of fit via the Kolmogorov-Smirnov test statistic. The Kolmogorov-Smirnov test was used to perform all goodness of fit assessments, since it does not assume a distributional form for its test statistic. In addition, the Kolmogorov-Smirnov test possesses a higher power relative to the chi-square test.

Based on comparison of the p-values for each of the distribution fits, the rainfall domain distribution provided a superior fit than the duration domain distribution for almost all of the three hundred fits. A paired t-test testing for equality between the mean number of null hypothesis acceptances for each site and for both distributions was found to be highly significant.

Storm depth at high rainfall intensity tended to be predicted more accurately with the rainfall domain distribution then was storm duration with the duration domain distribution. It should be mentioned however, that a rigorous test of goodness of fit at high rainfall intensity was outside the scope of this study and hence was not considered. The results of this study have implications for the simulation of runoff and soil erosion, since this new distribution can replicate rainfall intensity data with greater accuracy and should lead to improved runoff and soil erosion models.

1. INTRODUCTION

The study of relationships between storm depth and rainfall intensity is of hydrological interest, particularly in the modelling of runoff and soil erosion (Van Dijk *et al* 2005). Nicks *et al* (1995) and Brown and Foster (1987) are examples of two authors who have used an exponential relationship between rainfall intensity and the frequency of occurrence of that rainfall intensity. The problem with the approach used by Nicks *et al* (1995) and Brown and Foster (1987) is that both distributions involve total storm duration, which biases low intensity rainfall.

Van Dijk *et al* (2005) proposed a new one-parameter exponential distribution for modelling the relationship between storm depth and rainfall intensity, and tested this new model using data from one site in central Java, Indonesia. However, they did not test this distribution across a range of climates, nor did they test whether any previously existing distribution provided a significantly better fit. This paper tests the performance of two rainfall intensity distributions across a range of Australian climates, one in the time domain and the other in the rainfall domain. This work was conducted with a view to determine which of the two exponential distributions was superior for modelling the relationship between storm depth (or storm duration) and rainfall intensity.

2. METHODOLOGY

This section presents a methodology for testing goodness of fit for the two rainfall intensity distributions. The method of storm selection for the distribution fitting is discussed, as well as the test used for assessing the goodness of fit for both of these distributions. The analytical forms for the two rainfall intensity distributions are also presented.

2.1 Storm selection method

In Australia, rainfall intensity data were archived and maintained by the Bureau of Meteorology at 6 minute intervals with 0.1 *mm* resolution. Ten sites were selected around Australia to represent the full range of climatic conditions. The sites selected cover a latitudinal range from 12 degrees to 42 degrees, with mean annual rainfall ranging from 270 *mm* to 2000 *mm*. Table 1 documents the ten sites selected in this study.

Thirty storms were selected for each of the ten sites so that the goodness of fit could be compared for both distributions. The storms selected were required to be statistically independent. Use of the term statistically independent refers to a situation where the selected storms are not correlated in any way, and hence are not part of a single larger storm. A dry day preceding and following a wet spell was the criterion used to achieve statistical independence for each of the storms selected.

Only data of good quality was used in this study. For data to be judged as having good quality the extracted pluviograph must contain no missing data, whilst additionally, the total amount of rainfall accumulation as a result of pluviometer malfunction must be less than the resolution at which rainfall was recorded (*i.e.* 0.1 *mm*).

This study selected the thirty largest storms in terms of total rainfall depth that were statistically independent and met the selection criteria.

Table 1. Station number (Station No), Station name, and mean annual rainfall (MAR) for the ten sites selected for this study.

Station No	Station name	(MAR*) mm*
09021	Perth Airport	795
14015	Darwin Airport	1654
15590	Alice Springs Airport	274
23034	Adelaide Airport	449
31011	Cairns Airport	2000
40223	Brisbane Airport	1175
66037	Sydney Airport	1109
70014	Canberra Airport	634
86282	Melbourne Airport	573
94008	Hobart Airport	516

2.2 Rainfall distributions fitted

As was mentioned previously, this project attempted to determine which of the two exponential distributions considered fitted rainfall intensity data better for the ten sites across Australia from a range of climatic zones. The first distribution investigated the duration of storms greater than some threshold rainfall intensity, whilst the second distribution investigated the depth of rain that falls at greater than some threshold rainfall intensity.

The storm duration domain distribution was the first distribution considered and can be written as:

$$T(>R) = T_t \exp(-\frac{R}{\bar{R}}), \quad (1)$$

where:

- T_t is the total storm duration (*hr*),
- R is the rainfall intensity (*mmhr*$^{-1}$),
- and $T(>R)$ is the length of time for which the rainfall intensity is greater than R (*mmhr*$^{-1}$).

$\bar{R}$ is the only parameter for the distribution and can be calculated as:

$$\bar{R} = \frac{\sum i\Delta t}{\sum \Delta t}, \quad (2)$$

where:

- i is the average rainfall intensity (*mmhr*$^{-1}$) over a six minute period,
- and Δt (*hr*) is the time elapsed during each period (NB: this is always fixed at 6 minutes or 0.1 hours).

It can be observed from (2) that $\bar{R}$ is merely the storm depth (*mm*) divided by the storm duration (*hr*), and thus has an intuitive interpretation. The parameter $\bar{R}$ can be interpreted as the average storm intensity. A problem with this parameterisation scheme is that $\bar{R}$ is strongly influenced by low rainfall intensity over long periods of time which theoretically limits this distributions effectiveness for simulating runoff and soil erosion.

The storm depth based distribution can be written as:

$$P(>R) = P_t \exp(-\frac{R}{\check{R}}), \quad (3)$$

where:

- P_t is the total storm depth (*mm*),
- and $P(>R)$ is the amount of rainfall falling at an intensity greater than R (*mm*).

$\check{R}$ is a parameter for this distribution that can be calculated via:

$$\check{R} = \frac{\sum i^2 \Delta t}{P}, \quad (4)$$

The parameter $\check{R}$ from (4) can be interpreted as rainfall-weighted average intensity and as a result, high-intensity rainfall contributes more to this parameter than to the average intensity in (2). $\check{R}$ is also a theoretically superior parametric formulation for simulating runoff and soil erosion. The reason for this is that $\check{R}$ is influenced by storms with high rainfall intensity, the type of storms that would lead to greater runoff and soil losses.

2.3 Assessing goodness of fit

This sub-section outlines a methodology for determining whether the duration domain distribution (1), or the rainfall domain distribution (3) fitted the observed rainfall intensity data better. The Kolmogorov-Smirnov test was used for each storm to determine which distribution provided a better fit to the observed break-point data based on the p-value for the respective distribution fits.

The Kolmogorov-Smirnov test comes from a family of empirical distribution function tests (Wang *et al*, 2004) and was used in this paper to determine whether the observed and simulated data came from the same distribution. The Kolmogorov-Smirnov test was preferred over the chi-square test because of the higher power of the former test relative to the latter (Wang *et al*, 2004), and also because the Kolmogorov-Smirnov test does not assume a distributional form for its test statistic.

The Kolmogorov-Smirnov test was used to determine the degree of fit between a theoretical and empirical distribution, the test statistic for which can be written as (Hogg and Tanis, 2001):

$$D_n = \sup_x \{|F_n(x) - F_0(x)|\}, \quad (5)$$

where:

- D_n is the Kolmogorov-Smirnov test statistic,
- $F_n(x)$ is the empirical distribution function,
- $F_0(x)$ is the theoretical distribution function,
- and $\sup_x$ represents the *supremum* of the pointwise differences.

Two versions of the Kolmogorov-Smirnov test exist (Press *et al*, 1992). Namely;

1. if you are comparing the goodness of fit between an empirical and theoretical distribution, you must use the two-sample Kolmogorov-Smirnov test.

2. Alternatively, if you are attempting to discover whether two sets of empirical data follow the same distribution you must use the one-sample Kolmogorov-Smirnov test.

Since this study was addressing the degree of fit between a theoretical and empirical distribution, the two-sample Kolmogorov-Smirnov test was used for all goodness of fit assessments. A methodology for performing the two-sample Kolmogorov-Smirnov is outlined in Press *et al* (1992).

All goodness of fit assessments were performed with a null hypothesis stating that the empirical and theoretical distributions came from the same distribution, along with a significance level of 5%. If the Kolmogorov-Smirnov test statistic was greater than the critical value for this distribution, the null hypothesis was rejected. In such a case, we can reject the hypothesis that the empirical and theoretical distributions came from the same distribution. Thus, the two-sample Kolmogorov-Smirnov test will therefore establish whether or not the observed pluviograph data follow the duration domain or rainfall domain rainfall intensity distributions.

Subsequent to performing the Kolmogorov-Smirnov test, the number of storms at each site where the null hypothesis was accepted was recorded for both distributions. To further compare the fit for both distributions, the *p*-value was compared between the two distributions for each of the three-hundred storms.

3. RESULTS AND DISCUSSION

Results are presented in this section for the fitting of two rainfall intensity distributions. A duration domain distribution and a rainfall domain distribution were considered. The two-sample Kolmogorov-Smirnov test was employed to determine which distribution provided a superior fit to the observed pluviograph data.

3.1 Storm duration distribution fits

Figure 1 shows observed data and model predictions for the duration domain distribution for an Adelaide storm from the 8-10th February 1969. In this and subsequent figures the dotted line represents the observed pluviograph data, whilst the solid curve represents model predictions. For both the duration and rainfall domain distributions, the left panel of each figure displays the original distributions whilst the right panel displays the cumulative distributions. According to the Kolmogorov-Smirnov test, the observed pluviograph data for the Adelaide storm displayed in Figure 1 do not follow the duration domain distribution, since the null hypothesis was rejected. Figure 1 displays one of the better fits; however, the duration domain distribution predicts storm durations poorly at both low and high intensities.

Figure 2 shows observed data and model predictions for the duration domain distribution for a Darwin storm from the 29-30th October 1992. According to the Kolmogorov-Smirnov test, the observed pluviograph data do not follow the duration domain distribution, since the null hypothesis was rejected. Figure 2 displays one of the poorer fits between the theoretical and empirical distributions; and once again, the model does not predict storm duration well at high and low rainfall intensities.

For the majority of the three hundred distribution fits, the duration domain distribution modelled storm duration poorly at both low and high rainfall intensities. It must be emphasised that this study made no attempt to explicitly test goodness of fit at high rainfall intensity. For this reason, this further work is recommended as future research.

At low rainfall intensity, the duration domain distribution overestimates the duration of storms. This result is intuitively obvious, since the most common rainfall amount over a six minute interval during a storm (for any site in Australia) is close to 0 *mm*. Since the storm duration distribution specifies a monotonically decreasing storm duration (with increasing intensity) this gives rise to the problem of the storm duration distribution overestimating storm duration at low rainfall intensity. Despite the poor fit at low rainfall intensities, this is not a problem of hydrological concern (especially in soil erosion modelling), since the aim is to model high intensity rainfall well.

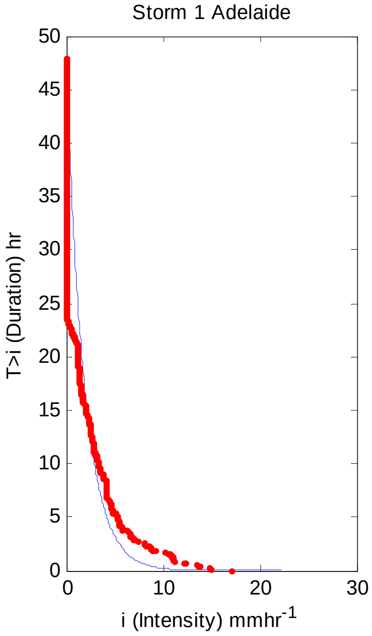

Figure 1. A plot of observed data and model predictions for the duration domain and cumulative duration domain distributions for an Adelaide storm from the 8-10th February 1969.

At higher rainfall intensities, the storm duration based distribution underestimates the duration of rainfall at high rainfall intensity. At this point in time, no sound explanation can be provided for this phenomenon.

It was noted that the storm duration based distribution was of little use in modelling pluviograph data, since it predicted storm duration poorly at both high and low rainfall intensity. Moreover, for only thirty-three of the three hundred distribution fits did the two-sample Kolmogorov-Smirnov test deem the theoretical and empirical distribution to be drawn from the same distribution at a five-percent level of significance.

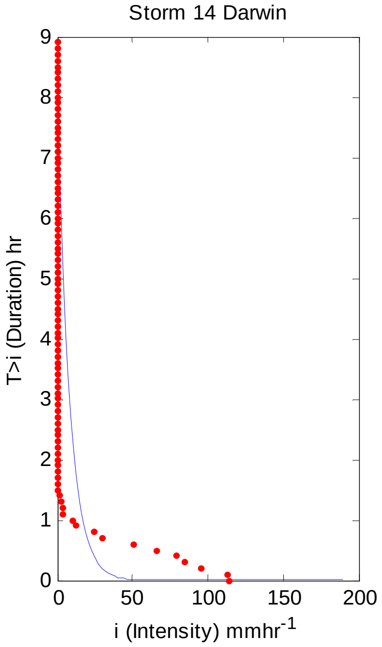

Figure 2. A plot of the observed data and model predictions for both the duration domain and cumulative duration domain distributions for a Darwin storm from the 29-30th October 1992.

3.2 Storm depth distribution fits

Figure 3 shows observed data and model predictions for the rainfall domain distribution for an Alice Springs storm from the 27-28th April 1968 and displays one of the better fits between the theoretical and empirical distributions. According to the Kolmogorov-Smirnov test, the observed pluviograph data follow the rainfall domain distribution, since the null hypothesis was accepted. The model predicts storm depths well at low to medium intensities. From Figure 3 it can be observed that the model predictions only begin to deviate from the pluviograph data over the last three data points. This distribution fit is indicative of the success of this new distribution for modelling pluviograph data. The model does not predict storm depth well at high rainfall intensity in Figure 3; however, the modelling discrepancy at high rainfall intensity tends not to be as large as for the duration domain distribution.

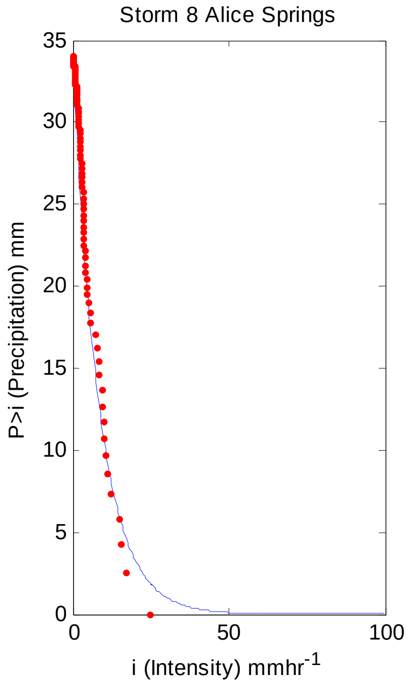

Figure 3. A plot of the observed data and model predictions for both the rainfall domain and cumulative rainfall domain distributions for an Alice Springs storm from the 27-28th April 1968.

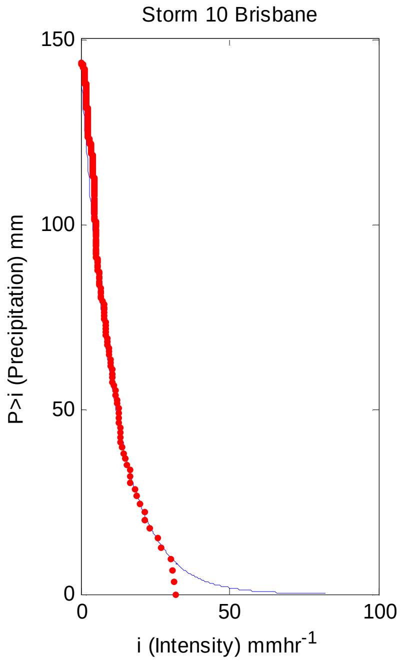

Figure 4. A plot of the observed data and model predictions for both the rainfall domain and cumulative rainfall domain distributions for a Brisbane storm from the 7-9th May 1985.

Figure 4 shows observed data and model predictions for the rainfall domain distribution for a Brisbane storm from the 7-9th July 1985 and once again displays one of the better fits between the theoretical and empirical distributions. In this case, the Kolmogorov-Smirnov test declares that the observed data and model predictions came from the same distribution, since the null hypothesis was accepted. Medium and low intensity rainfall is modelled quite well, however, the storm depth distribution has difficulty in replicating high intensity.

The rainfall domain distribution quite often provided a much better estimate of storm depth at high rainfall intensity compared to the duration domain distributions prediction of storm duration at the same intensity. Despite the better fit at high rainfall intensity, further research should be conducted to further improve this fit.

3.3 Comparative fit

Table 2 documents the comparative goodness of fit for both the duration domain (1) and rainfall domain (3) distributions at each of the ten sites. Table 2 records the number of storms (*i.e.* out of thirty) for which the null hypothesis was accepted at a five percent significance level for both distributions.

Table 2. Goodness of fit results at each site for the duration and storm depth distributions.

Site	T acceptances	P acceptances
Perth	1	23
Darwin	1	3
Alice Springs	5	23
Adelaide	4	17
Cairns	3	22
Brisbane	7	23
Sydney	5	23
Canberra	2	26
Melbourne	1	19
Hobart	4	18

It can be observed from Table 2 that far more acceptances of the null hypothesis occurred for the rainfall domain distribution compared to the duration domain distribution. For the duration domain distribution thirty-three acceptances of the null hypothesis occurred across all ten sites, whereas one hundred and ninety-seven acceptances of the null hypothesis occurred for the rainfall domain distribution across the same ten sites. At every site where the modelling was performed, the rainfall domain distribution achieved more acceptances of the null hypothesis than the duration domain distribution. In addition, across the ten sites considered, the rainfall domain distribution achieved nearly six times as many null hypothesis acceptances than the duration domain distribution.

A paired t-test was subsequently performed (with a significance level of five percent) to test whether the mean number of storms for which the null hypothesis was accepted was the same for both exponential rainfall intensity distributions. Not surprisingly, the paired t-test was highly significant with a p-value of 1.25×10^{-5}. Based on an equal sample size of thirty storms, this result suggests that the rainfall domain distribution provides a significantly better fit to the pluviograph data than the rainfall duration distribution. This result is achieved by virtue of a greater number of null hypothesis acceptances being achieved for the rainfall domain distribution compared to the duration domain distribution.

4. CONCLUSIONS

The rainfall domain distribution outperformed the duration domain distribution as evidenced by its goodness of fit results. The paired t-test was shown to be highly significant, whilst a null hypothesis acceptance occurred six times more frequently for the rainfall domain distribution compared to the duration domain distribution. Thus, the new rainfall domain distribution proposed by Van Dijk *et al* (2005) achieved more success in modelling pluviograph data than the duration domain distribution across all Australian climates considered. It can thus be concluded that this new rainfall domain distribution should be used instead of the previously existing duration domain distribution. The improved goodness of fit results for modelling pluviograph data is indicative of the potential of this new distribution for simulating runoff and soil erosion.

Despite the promising results for the new rainfall domain distribution, goodness of fit at high rainfall intensity was not explicitly tested. It is not always true that the maximum pointwise difference between the cumulative observed and modelled distributions (as required for the Kolmogorov-Smirnov test) always occurs at high rainfall intensity. Thus future research could focus on giving high weighting to goodness of fit at high rainfall intensity, given the hydrological significance of heavy storms.

Additionally, the simulation of storm depth (or duration) at high rainfall intensity continues to be a challenge. Improving the fit of a simulated distribution at high rainfall intensity is of particular relevance in the modelling of runoff and soil erosion. Thus future research should be directed towards better estimating storm depth or duration at high rainfall intensity.

5. REFERENCES

Brown, L.C. and G.R. Foster (1987). Storm erosivity using idealized intensity distributions. *Transactions of the American Society of Agricultural Engineers*, 30: 379-386.

Hogg, R.V. and E.A. Tanis (2001), Probability and Statistical Inference sixth edition, Prentice-Hall Inc, Upper Saddle River.

Nicks, A.D., L.J. Lane and G.A. Gander. (1995). Chapter 2 – Weather Generator. In: Flanagan, C., and M.A. Nearing (Eds). USDA – Water Erosion Prediction Project (WEPP). Hillslope Profile and Watershed Model Documentation. NSERL Report No. 10, USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, Indiana.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992), Numerical recipes in FORTRAN: the art of scientific computing second edition, Cambridge University Press, Cambridge.

Van Dijk, A.I.J.M., A.G.C.A. Meesters, J. Schellekens and L.A. Bruijnzeel (2005), A two-parameter exponential rainfall depth-intensity distribution applied to runoff and erosion modelling, *Journal of Hydrology*, 300 (1-4): 155-171.

Wang, Y., R.C.M. Yam and M.J. Zuo (2004), A multi-criterion evaluation approach to selection of the best statistical distribution, *Computers and Industrial Engineering*, 47: 165-180.