Defining Independence of Rainfall Events with a Partial Duration Series Approach

Karin Xuereb
Senior Meteorologist, Climate and Water Division, Bureau of Meteorology, Melbourne, Australia

Janice Green
IFD Revision Project Manager, Climate and Water Division, Bureau of Meteorology, Australia

Abstract

The Intensity-Frequency-Duration (IFD) design rainfall revision project has explored the use of partial duration series (PDS) with rainfall frequency analysis. For PDS analysis to be valid, the PDS rainfall values need to be independent necessitating the need for establishing criteria for independence. This paper discusses this process of determining independence criteria for Australian rainfall data.

Events that occurred close together in time were deemed independent if they arose from different meteorological systems. Meteorological analysis was carried out at selected locations in Australia and this showed that short-duration multiple rainfall events in the PDS occurring on a sub-daily time scale could result from a larger scale meteorological system on a daily time scale. Thus, small scale thunderstorms, whilst appearing to be independent and separated by several hours with no rainfall could be part of a larger scale system, such as a trough, which renders multiple events more likely. In such cases only the larger event should be included in the PDS.

The effect of independence on design rainfall estimates was explored using the PDS with and without independence factored in. It was found that having non-independent peaks in the PDS affected the design rainfall the most for sub-hourly durations in tropical areas. For durations longer than 1 hour, the effect on the design rainfall was minimal and within the 95% confidence intervals.

Analysis of one-day events in the daily PDS showed that the minimum inter-event separation time (MIT) required between PDS values varies with latitude, between two days in the southern regions and six days in the north. For longer durations, more work is required to define independence rigorously. Until such work has been carried out, the MIT adopted decreases with increasing duration.

1. INTRODUCTION

The Australian Rainfall and Runoff Revision project will present revised estimates of design rainfall over Australia. An approach which has been considered for the project is rainfall frequency analysis using partial duration series (PDS).

The advantage of PDS analysis compared with annual maximum series (AMS) analysis is that the PDS contains more observations at the higher end of the scale, because sampling is not limited by the calendar year as is the case with the AMS and periods of missing data do not compromise the extraction of the PDS. Nonetheless, frequency analysis using the AMS is much more common in hydrologic studies compared with PDS analysis because extracting the PDS is not straightforward. One drawback involves the necessity of selecting a threshold above which PDS values are extracted. Another involves ensuring that peaks included in the PDS are independent.

Cunnane (1973) recommended an average number of exceedances per year greater than 1.65 for a PDS approach. Madsen *et al* (1997) found that with Probability Weight Moments (PWM) estimation, for negative shape parameter, the PDS with Generalised Pareto model is preferable to an AMS approach with a Generalised Extreme Value (GEV) model. For Australia the shape parameter is generally negative for the 1-day duration, therefore the PDS/GP approach may be appropriate.

Madsen *et al* (1997) found that with PWM the number of exceedances per year required for a PDS with GPA to obtain equal performance to AMS with GEV ranged from less than 0.4 to 6.5 for shape parameter values between -0.3 to 0 (which is the case for most 1-day data for Australia). Trefry and Watkins (2001) set the average number of exceedances to 2.

For this study, it was decided to set the average number of peaks per year to 3 because it is adequate most of the time for shape parameters between -0.3 and 0 (Madsen *et al*, 1997) and it is more practical to have a constant average number of exceedances. In the initial creation of the PDS data sets for each station, the average number of events was set to 5 because some of the events will not be independent and will need to be removed from the data set. Following the application of the independence criteria, developed in this work and reported in this paper, the final PDS will have an average of 3 peaks per year.

For frequency analysis to be valid, it was necessary to devise a method to remove non-independent peaks by establishing criteria that set the minimum separation time between events such that the events extracted are independent. These criteria need to be applied across Australia spanning different climatic regimes, therefore the criteria may vary geographically. The aim is for such criteria to be easily applicable to PDS datasets.

Studies that have addressed this issue so far commonly define a minimum inter-event time (MIT) between one event and another to maintain independence. Such definitions are usually somewhat subjective and the criteria adopted to define independent events vary from one study to another depending on the type of research undertaken. MITs ranging from 3 minutes up to 24 hours have commonly been used (Dunkerley, 2008). Trefry and Watkins (2001), in their study of design rainfalls from PDS, specified gaps between peaks equal to the rainfall duration. Therefore peaks for the 2-hour duration, for example, are separated by a period of at least 2 hours. Manfroi *et al*, (2004) adopted an MIT of 6 hours based on the diurnal double peak rainfall pattern in the study area in Borneo in their research on the stemflow of trees.

This paper defines the MIT as the minimum separation time between the end of one rainfall event and the start of the subsequent rainfall event. The MIT is to be specified for geographic regions in Australia and, when applied to time series of rainfall, will generally result in the selection of independent rainfall events. The method used to establish the criteria for MIT involves identifying multiple rainfall peaks in the PDS database for the subdaily durations, calculating the separation time and identifying such peaks as independent or not independent through meteorological considerations to arrive at the MIT for the sites. The method is then extended to the daily rainfall dataset to establish criteria that will result in generally independent rainfall peaks from time series.

2. STATISTICAL DEFINITION OF INDEPENDENCE

In probability theory, two events are termed independent if the occurrence of one event does not affect the occurrence of the other.

More formally: $\Pr(A \cap B) = \Pr(A)\Pr(B)$, i.e. the joint probability of both events A and B occurring equals the product of the probabilities of events A and B occurring individually. More generally for n events:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr(A_i). \tag{1}$$

Criteria that define independence in this study are consistent with equation (1) qualitatively.

3. METHOD

Subdaily data was used initially to examine whether the MIT could be defined on a subdaily time scale. Sites at several locations in Australia representative of a range of different climates were selected and events occurring close together in time in PDS identified as being independent or 'not independent' through meteorological analysis whilst satisfying equation (1) in a qualitative manner. If, given certain meteorological conditions, the occurrence of two or more events is more likely than a single event, equation (1) is considered to be violated.

The analysis was extended to the daily dataset to provide criteria for independence of rainfall events for durations from 1 to 7 days.

3.1. Definition of independence using subdaily data

Nine sites were selected for analysis to test the adopted approach and are shown in Figure 1. The sites are: Melbourne Regional Office (086071), Kalinga Bowls Club (040222) (near Brisbane), Sydney (Observatory Hill) (066062), Townsville Aerodrome (032040), Darwin Airport (014015), Alice Springs Airport (015590), Perth Airport (009021), Adelaide Airport (023034) and Hobart Airport (094008). The sites were selected on the basis of availability of satellite image data and geographic location to capture a variety of climate regimes that would impact on the rainfall patterns that would affect criteria for independence. Kalinga Bowls Club was selected rather than Brisbane because of the availability of more recent rainfall data for which satellite imagery can be extracted.

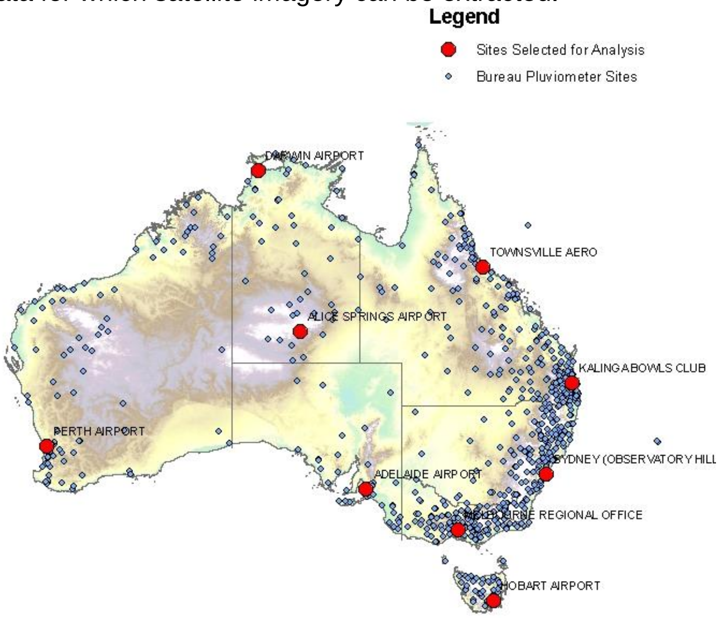

Figure 1 Sites selected for analysis.

The time series containing PDS events that occurred close together in time were examined. An example from Townsville Aerodrome is shown in Figure 2.

The time series of rainfall often exhibited a noisy pattern with many peaks separated by little or no rainfall as in Figure 2 Time series for events between 14 and 16 February 2002 at Townsville Aerodrome. The coloured lines indicate the time span over which an event was significant enough to be included in the partial duration series.. While it may seems obvious that peaks D and E should not be considered as separate events because there is no clear dry period between them, it is not that clear whether events A and B or E and F are separate. Quite often the peaks were bursts within a large rainfall event that was caused by a meteorological system such as a cut-off low, a front, or a tropical cyclone. Rainfall peaks that are linked through a meteorological system are defined as being 'not independent'. For events to be considered independent, rainfall peaks would need to have been caused through different meteorological processes. Conversely, rainfall peaks that are due to a common meteorological system could not have occurred in isolation and they are considered 'not independent'. The difference between meteorological events that give rise to a single peak as opposed to multiple peaks depends on several meteorological conditions and atmospheric processes such as: moisture availability, atmospheric instability, mobility and structure of meteorological systems, location of the rain gauge relative to the rainfall generating system and cloud microphysical processes that convert water vapour into rain droplets. Given the correct atmospheric conditions, multiple rainfall peaks may be more likely than single rainfall peaks within a rainfall event. If multiple rainfall peaks are more likely, then the assumption of independence in Equation 1 is violated.

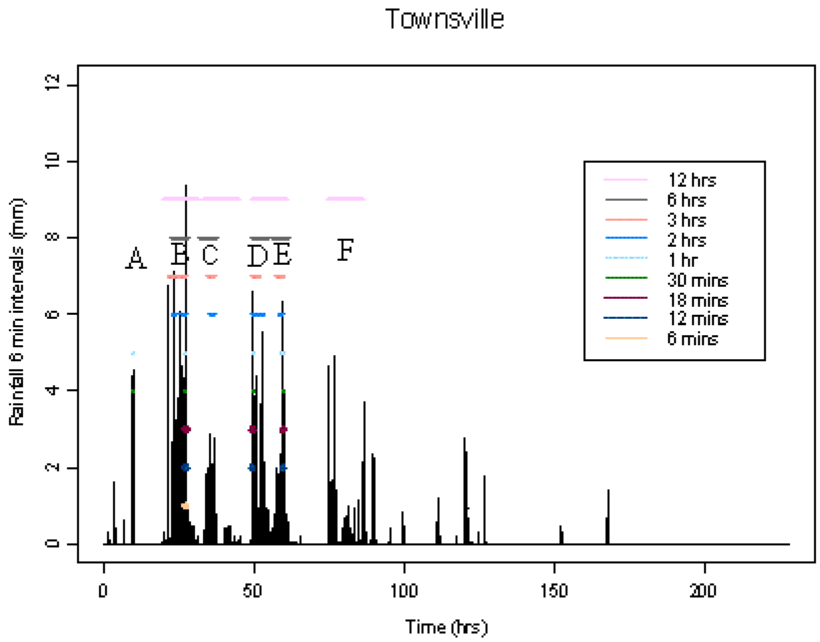

Figure 2 Time series for events between 14 and 16 February 2002 at Townsville Aerodrome. The coloured lines indicate the time span over which an event was significant enough to be included in the partial duration series.

The approach tested involved finding the separation time between rainfall peaks that occurred close together in time at the same location. By identifying which peaks were 'independent' or 'not independent' through meteorological analysis, the MIT for that location could be specified. Multiple rainfall events were classified according to the meteorological features that gave rise to the events. The meteorological categories for rainfall events were: fronts, thunderstorms in a trough, low pressure systems and tropical cyclones. Table 1 shows the MIT for 9 study locations. The MIT is the same for events of all durations up to the MIT.

Table 1 – Minimum Inter-Event Time (MIT) for study locations

Station	MIT (hours)
Perth Airport	36
Darwin Airport	96
Alice Springs Airport	36
Adelaide Airport	36
Townsville Aerodrome	72
Kalinga Bowls Club	48
Sydney Observatory Hill	36
Melbourne Regional Office	36
Hobart Airport	36

3.2. Effect of the presence of non-independent PDS values on design rainfall estimates

The Generalised Pareto distribution was fitted to the partial duration series using the independent and non-independent datasets for Melbourne Regional Office (086071), Kalinga Bowls Club (040222), and Townsville (032040) and percentage differences calculated between using the independent and non-independent datasets (Figure 3). It was found that the largest impact on PDS by the presence of non-independent events was at the shorter durations up to 1 hour, quite likely to be due to the presence of contiguous events which occurred quite often at the short durations. Contiguous events occurred much less frequently for the longer durations and less sensitivity to the presence of non-independent events in the dataset was manifest.

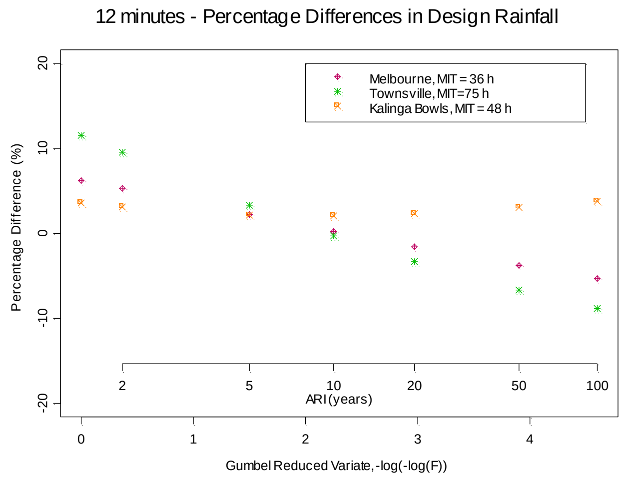

Figure 3 Percentage differences in design rainfall between independent and non-independent datasets.

3.3. Definition of Independence using the daily rainfall dataset

The analysis from the pluviometer data reported in the previous section shows an MIT of 36 hours across the southern part of Australia. The criteria for the MIT need to be applicable to daily data and thus needs to be rounded to a whole number of days, suggesting an MIT of 1 or 2 days. Setting the MIT is a balance between a short MIT which introduces non-independent events and a too long MIT which will unnecessarily reject events that are truly independent. It was decided to apply an MIT of 2 days because, from the analysis of the study locations, this reduced the chance of having non-independent events in the data more often than losing an event through the application of a longer MIT.

To extend the analysis to the daily data, the relationship between rainfall events separated by only a few days was considered. If two rainfall peaks occur close together in time, how likely is it that the secondary peak is significant enough to be ranked highly in the PDS? If the secondary peak is not likely to be a high-ranking rainfall depth, its presence in the PDS will not affect design rainfall estimates whether it is independent or not.

The focus of the extended daily analysis was the area north of latitude 30 degrees south where it is not clear how to define the MIT. A subset of the daily sites in this region consisting of 627 sites was selected to reduce computation time. The sites were selected at random but all had at least 20 years of data. When the selected sites were close together spatially, those with the longer record lengths were selected because these would be more likely to have higher rainfall depths than neighbouring sites with shorter record lengths. Pairs of events occurring with 2, 3, 4 and 5 day separations were extracted from the 1-day PDS. For each site and for each pair, the ratio of the highest 1-day rainfall to the highest PDS value in the 1-day dataset was calculated. Ratios of the rainfall depths of each event pair were then calculated by dividing the smaller rainfall depth by the higher for each pair. The higher this ratio the more similar the two events are and it is more likely that the second peak affects design rainfall for the site. Furthermore, if the pair also happens to be ranked highly in the PDS series then the secondary peak is even more likely to affect design rainfall estimates. Graphically, this is depicted by plotting the standardised one day rainfall against the ratio of the smaller of the 1-day rainfall to the higher 1-day rainfall. The plots were prepared using separations of 2, 3, 4 and 5 days in the PDS dataset for each consecutive event pair.

Estimates of design rainfall would be most affected by data plotted in the far upper right hand corner of

Figure 4 (the pink shaded region) because these suggest the presence of a pair of events close together in time that are of the same magnitude and are also ranked near the top of the PDS. The small number of points in the shaded region indicates that it is unlikely that two events of similar magnitude occur close together in time suggesting that the presence of peaks that are not independent even with a two-day separation is unlikely. Examination of the points that plot in the far upper right corner showed they were all not independent or could not be assessed due to lack of satellite imagery. With longer event separations the number of anomalies decreases until with event separations of 5 days (Figure 4(d)) there is only one anomaly. This event was associated with a tropical cyclone so it was clearly not independent. A six-day separation would eliminate all anomalies. The geographic locations of the anomalies are shown plotted in Figure 5.

The analysis suggests a varying MIT of 2 to 6 days is appropriate for Australia. A gradually varying MIT would be ideal, however, it is not possible to define the MIT accurately enough. For ease of application and considering the uncertainty of about 1 day in attributing an MIT across a zone, the adopted MIT is defined in 2-day time steps. The zones of MIT have been drawn by plotting the MIT from the pluviometer data along with the anomalies from the daily data at 2- and 4-day separations and are shown in Figure 5. The MIT is slightly longer for the same latitude in eastern Australia compared with Western Australia. This has been drawn to reflect the persistence of a deep trough in the east coast in the warmer months which prolongs rainfall events. The delineation of MIT zones is not rigid but is more of a guide to a reasonable MIT to adopt in Australia.

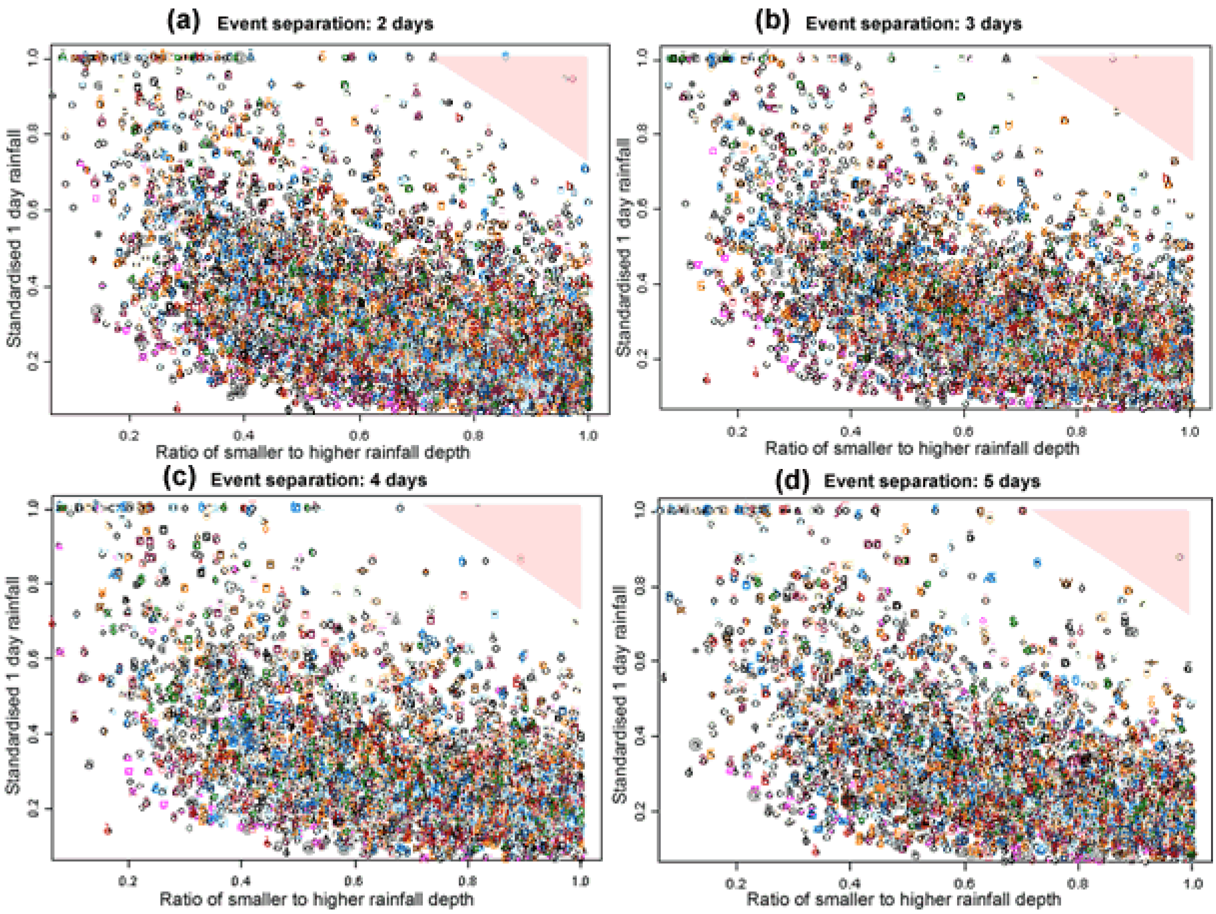

Figure 4 Standardised 1 day rainfall plotted against the ratio of the 1 day smaller rainfall to the higher rainfall for each pair of events separated by 2, 3, 4 and 5 days. Pink shading indicates the region where an event pair has a secondary peak that is significant enough to affect design rainfall estimation.

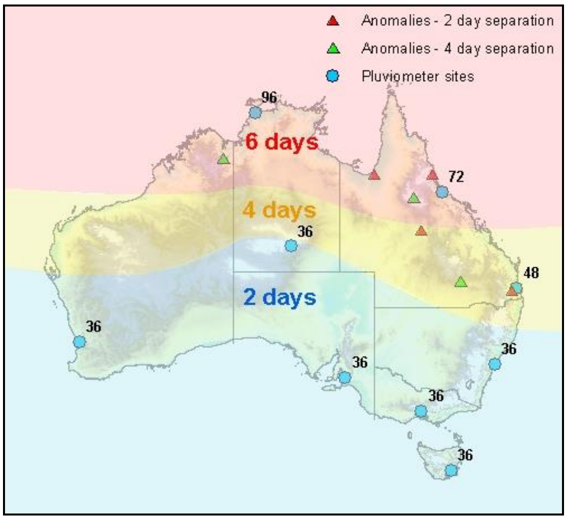

Figure 5 Criteria for independence for 1-day rainfall.

These criteria have been derived from meteorological analysis as well as the analysis from daily data above. The daily data analysis has been performed on events of 1-day duration. Therefore the MIT in Figure 4 is based on the separation of 1-day events. To define the criteria for independence for durations up to 7 days, further analysis could be undertaken. For example, the analysis reported above could be extended to longer durations. Until such work can be undertaken, a more pragmatic approach has been adopted.

One-day events are often contained within longer-duration events. Assuming that multi-day events contain the 1-day event in the PDS, the MIT for an event of duration longer than the MIT would be zero. This would suggest that two events can be contiguous and also independent. This can occur if two events are from different meteorological systems for example one is a tropical cyclone which decays and is followed by monsoon convective activity. Both systems could occur independently with any number of intervening days between them. A detailed study of how often this might happen has not been carried out. By setting an MIT of zero for the longer durations, there is less likelihood of rejecting events that may be independent and possibly including more events that are not independent. The effect on design rainfall estimates would only be significant if the two non-independent events were to be of similar magnitude and both were highly ranked in the PDS. It would not be expected that this would occur often enough to affect results.

The MIT for the 1-day duration shown in **Error! Reference source not found.** contains the event as well which can vary in duration from 1 to 7 days (7 days is the duration limit required). Suppose for simplicity, events are symmetric.

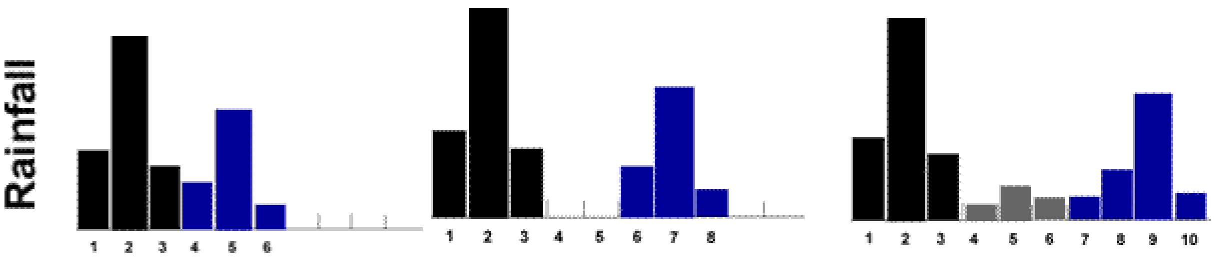

Figure 6 (a) MIT = 2 days **(b) MIT = 4 days** **(c) MIT=6 days**

Figure 6 (a) shows two main 1-day events occurring on day 2 and day 5. If these were parts of two-day events, and the rainfall on day 3 were bigger than on day 1 and that on day 4 bigger than on day 6, the MIT for 2 days would be zero (whilst the MIT for 1 day is 2 days). If the event were a 2-day event in a region where the MIT for a one-day event is 4 days, the minimum separation would be 2 days for the 2-day event. In a region where the 1-day MIT is 6 days, the 1-day events fall on days 2 and 9. For 2-day events, the minimum separation is 4 days, for 3-day events the minimum separation is 2 days, and for 4-day events the minimum separation is zero.

Of course real situations are not necessarily symmetric and longer-duration events do not always contain the one-day event. It would be preferable to undertake a more rigorous exercise for the long duration events, however, in the interim, it is considered that by adopting a pragmatic approach, the quality of results using the criteria adopted would not be compromised.

The criteria for MIT for durations from 1 to 7 days that have been adopted are set out in Table 2. These have been devised for extraction of PDS through a computational procedure and are not necessarily realistic.

Table 2 MIT for durations from 1 to 7 days in each MIT region.

Duration (days)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
MIT_D^2	2	0	0	0	0	0	0
MIT_D^4	4	2	0	0	0	0	0
MIT_D^6	6	4	2	0	0	0	0

4. CONCLUSIONS

Pluviometer and daily data has been used to define the MIT required between rainfall events that satisfy the requirement for events to be independent for valid frequency analysis. The results suggest a varying MIT ranging from 2 to 6 days depending on geographic location in Australia for event durations up to 3 days. For durations greater than 3 days the MIT would be zero.

The criteria can be easily applied to rainfall data resulting in rainfall events that can be considered independent and appropriate for use in the PDS for rainfall frequency analyses.

5. ACKNOWLEDGEMENTS

The valuable input provided by Dr Sri Srikanthan, Dr Fiona Johnson, Dr Robert Smalley and Dr Grace Mitchell is gratefully acknowledged.

6. REFERENCES

Madsen H., Rasmussen P.F., and Rosbjerg D., (1997). *Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events. 1. At-site modelling*., Water Resources Research, Vol. 33, No. 4, 747-757, April1997.

Dunkerley D., (2008). *Identifying individual rain events from pluviograph records: a review with analysis of data from an Australian dryland site*, Hydrological Processes, 22, 5024-5036.

Hosking J.R.M. and Wallis J.R.(1997), *Regional Frequency Analysis: An Approach based on L-moments*, Cambridge University Press.

Manfroi O.J., Koichiro K., Nobuaki T., Masakazu S., Nakagawa M., Nakashizuka T. and Chong L., (2004). *The stemflow of trees in a Bornean lowland tropical forest*, Hydrological Processes, 18, 2455-2474.

Trefry C.M. and Watkins Jr, D.W., (2001). *Application of a Partial Duration Series Model for Regional Rainfall Frequency Analysis in Michigan*, World Water and Environmental Resources Congress: May 20-24, 2001, Orlando, Florida, USA, edited by Don Phelps, Gerald Sehlka, American Society of Civil Engineers.