Non-stationary in annual maximum rainfall series – is it an issue for Australian data?

F.M. Johnson[1] and J.H. Green[2]
[1]Climate & Water Division
Bureau of Meteorology
Sydney, NSW 2000
AUSTRALIA
[2]Climate & Water Division
Bureau of Meteorology
Melbourne, VIC 3000
AUSTRALIA

Abstract: *As part of the revision of Australian Rainfall and Runoff, the Bureau of Meteorology is updating Intensity-Frequency-Duration data. One of the motivations for the study is the additional data that are available through the Water Regulations 2008 and in addition the longer records available at all stations. However if climate change has caused non-stationarity in the recorded rainfalls, then possibly only a portion of the observed record should be used in deriving IFD information. This paper assesses the degree of non-stationarity present in the historic record at rainfall stations across Australia. It is concluded that although some stations show strong trends in annual maximum time series, particularly for short durations and more frequent events, the magnitude of these changes is within the expected accuracy of the fitted IFD relationships. Future changes are however likely to be larger and will have to be considered for the revision of AR&R.*

Keywords: *Rainfall, trends, non-stationarity, Australia*

1. INTRODUCTION

The Bureau of Meteorology (the Bureau) is currently undertaking a revision of the Intensity-Frequency-Duration (IFD) design rainfalls for the whole of Australia. The previous IFD estimates were developed over 20 years ago (Institution of Engineers Australia, 1987) and the current review is motivated by the quantity of additional data that is now available due expansions in the Bureau's network, the additional length of records available, and access, under the terms of the Water Regulations 2008, to daily-read and continuous rainfall data collected by other organisations which supplements the Bureau's network.

The earlier IFD estimates assumed that the period of record at each gauge was representative of long term conditions for that location. In addition it was assumed that by pooling site data together in the generalised data analysis, any impacts of this assumption would be minimised (Institution of Engineers Australia, 1987). Since then it has become evident that variability in Australia's climate which occurs on interannual to interdecadal time scales (e.g. Power et al., 1999, Ummenhofer et al., 2008), coupled with the impacts of anthropogenic climate change (Solomon et al., 2007) may mean that existing rainfall data, particularly short records, are not representative of the long term conditions at a site. In addition, future changes may be larger than previously experienced (Huntington, 2006). These changes to variability affect both the mean climate state and large rainfalls; the latter being of more importance for the IFD Revision Project.

The purpose of this study is to investigate if there is significant evidence of non-stationarity in Australian rainfall data. The results are important for answering the question of whether the full historic record should be used to derive updated IFD estimates. Two approaches, the first a point based approach and the second an area averaged approach, are used with daily and continuous rainfall data to test for trends and evidence of non-stationarity over a range of event durations. Section 2 discusses the methodology used for the study and available data. Section 3 provides results of the analyses and conclusions are presented in Section 4.

2. METHODOLOGY

Two broad methodologies have been used to establish if there are trends in rainfall extremes for Australia. The first is to examine the records at individual stations. Tests include assessing trends in the time series of annual maximum rainfalls and changes in the probability distributions fitted to the annual maximums to estimate design rainfall quantiles. The second methodology uses an area averaged approach to check for regional trends in annual maximum rainfalls. The approach is based on that carried out by Bonin et al.(n.d.) to assess trends in large rainfall events in the USA as part of the revisions by the National Oceanic and Atmospheric Administration (NOAA) to design rainfalls. This section describes the application of both methodologies to Australian rainfall data.

2.1. Point based assessment of the stationarity of annual maximums

The first part of the study involves analysing if there is significant evidence of changes in the annual maximum time series at individual stations. This can be tested in several ways. The first analysis involves dividing the time series into two (equal) periods and determining if there are differences between these two data samples. Formally this involves a hypothesis test on some statistic of the two samples to test for significant differences. Two hypothesis tests have been conducted, namely:

- The Wilcoxon rank-sum test (also called the Mann-Whitney rank-sum test) which is a non-parametric test for difference in location (i.e. median) between two data samples.
- The two-sample Kolmogorov – Smirnov test which tests if two sets of data were drawn from the same, unspecified, distribution.

In addition to the split sample analysis approach, analysis has been carried out on the full data period at each station to test for trends. Two trend tests have been used:

- A t-test on the significance of the slope of a linear regression of the annual maximum time series, and
- The non-parametric Spearman's rank correlation to test for correlation of the annual maximum series vs time.

The tests on annual maximum time series have been carried out for both daily and continuous rainfall stations. Due to the large number of daily rainfall stations in Australia, it was considered that the most value from the analysis would be in determining if there are spatial patterns to trends or non-stationarity in the annual maximum rainfall series. As all stations have different starting and ending dates, a common period of analysis from 1901 to 2000 was adopted and stations with at least 40 years of data within the period (and at least 20 years of data in each half of the nominated period) were used for the analysis. For the continuous rainfall stations a common 50 year analysis period (1956-2005) has been used for all stations. This approach allows for field significance tests on the likelihood of obtaining the observed number of significant trends across Australia. The field significance tests are undertaken by considering the results from each gauge as samples from a binomial distribution, where the probability of success for any one trial is p, the significance level (here set as $\alpha = 5\%$) of each of the individual tests.

2.2. Point based assessment of the stationarity of probability distributions and estimated rainfall quantiles

Using the daily and continuous rainfall data for each station and partitioning the record into two separate periods, it is possible to test if the data in the two periods belong to the same or different probability distributions. As for annual maximum trends discussed above, common analysis periods were used for all stations to allow for spatial comparison of the results.

To test for difference in the distributions, a GEV distribution is fit to one of the data samples using the method of L-moments (Hosking and Wallis, 1997) and then a one sample Kolmogorov – Smirnov test is used to test if the data from the second sample is consistent with this distribution. Using the fitted GEV distributions for both periods, estimates of rainfalls quantiles with Annual Exceedance

Probabilities (AEPs) of 1 in 1.25 to 1 in 100 have been calculated. Confidence limits on the quantiles were estimated by creating 100 bootstrap samples of the data in each period. The difference in the quantile estimates for each bootstrap was calculated and confidence limits on the differences estimated. If the confidence interval on the difference does not cover zero, then the two quantile estimates are considered to be significantly different.

2.3. Regional assessment of the stationarity of annual maximums

Bonin et al. (n.d.) describe a methodology used by NOAA to determine if there are regional trends in the number of exceedances of design rainfall thresholds in the historical record. Two large regions of United States were considered using the most recent analysis in NOAA Atlas 14 (Bonin et al., 2006) of frequency, depth and duration estimates of design rainfall. Atlas 14 covers the semi-arid southwest US and the Ohio River Basin and surrounding areas.

For the NOAA analysis the historic record at each station in a region is treated as a partial duration series and for each duration a count is made of the number of times that the recorded rainfall exceeded the relevant design rainfall at that station, for different AEPs. The average number of exceedances across all stations in the region is then calculated (the sum of all exceedances divided by the number of stations operating in that year). A time series is then constructed of the average number of exceedances per year and tested for non stationarity using a number of non-parametric and parametric tests – namely the Wilcoxon rank sum test, Spearman's rank correlation test and parametric t-test.

The NOAA methodology has been applied for both the continuous and daily rainfall stations in Australia. The analysis requires regional averaging of the exceedance time series; a way of grouping stations together is therefore required. For the daily stations, there is sufficient station density such that the analysis could be carried out Australia wide. Rainfall districts were used to group the stations with each district containing multiple gauges. The average exceedance time series has only been calculated for years when at least five gauges in a region have records. For the continuous rainfall stations, there are far fewer stations available for analysis, so an Australian wide analysis of trends is not possible. The analysis has instead been carried out for Districts 31, 40, 39 & 40 and 85 & 86, which each contain between 3 and 7 stations with long term records. The locations of the districts are shown in Figure 1b.

2.4. Data sources

Data are available for 17247 Bureau operated daily read rainfall stations across Australia. The data from each station have been quality controlled with infilling of missing data from nearby stations and disaggregation of any aggregated multi-day rainfall totals. For the non-stationarity assessment, annual maximum time series are required at each station. Annual maxima were calculated for each year that had at least 10 months of data with at least 75% of data values available. Annual maxima were also checked for independence over consecutive years. For the analysis period of 1901-2000 it was found that 4191 stations had sufficient quality controlled data for the analyses.

Data have been analysed for 58 Bureau operated continuous rainfall stations. Stations were chosen to ensure that they have at least 40 years of data (with each year having at least 10 months with at least 75% of data values available). To determine the common period to be used for the field significance tests and distribution fitting, the starting and finishing dates for each station were compared and a 50 year period chosen that maximised the number of stations with records in the period. A common period of analysis of 1956 to 2005 was chosen. Twelve event durations were examined ranging from 6 minutes to 24 hours. The distribution of the daily and continuous rainfall stations is shown in Figure 1.

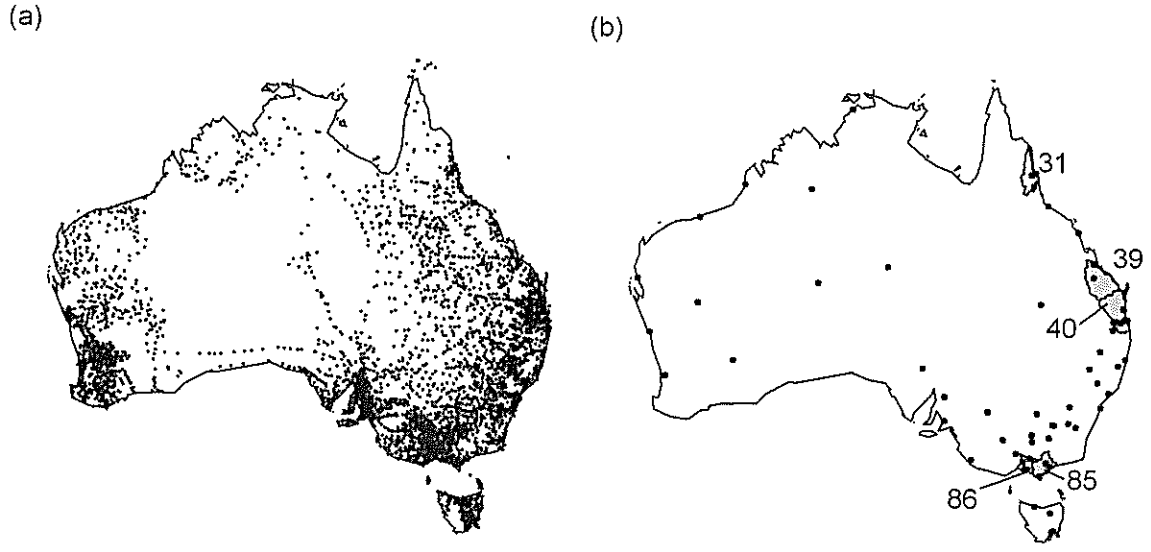

Figure 1 Distribution of (a) daily rainfall stations and (b) continuous rainfall stations used for the analyses. Districts used for regional assessment of sub daily durations are also shown.

3. RESULTS

3.1. Point based assessment of stationarity

Trend analysis for both the continuous and daily rainfall stations shows that there are more increasing trends (2733 stations) than decreasing trends (1458). However the majority of these trends are not significant at the 5% level. Figure 2 shows the directions of the trends in the daily annual maximum time series and their significance based on a t-test. There is reasonably good consistency between the three tests (i.e. Kolmogorov-Smirnov, Wilcoxon Rank Sum test and t-test) on the direction and significance of the trends. Approximately 6-9% of stations show significant increasing trends while approximately 4% of stations have significant decreasing trends. There are no clear spatial patterns in the increasing/decreasing trends or in the significance of those trends.

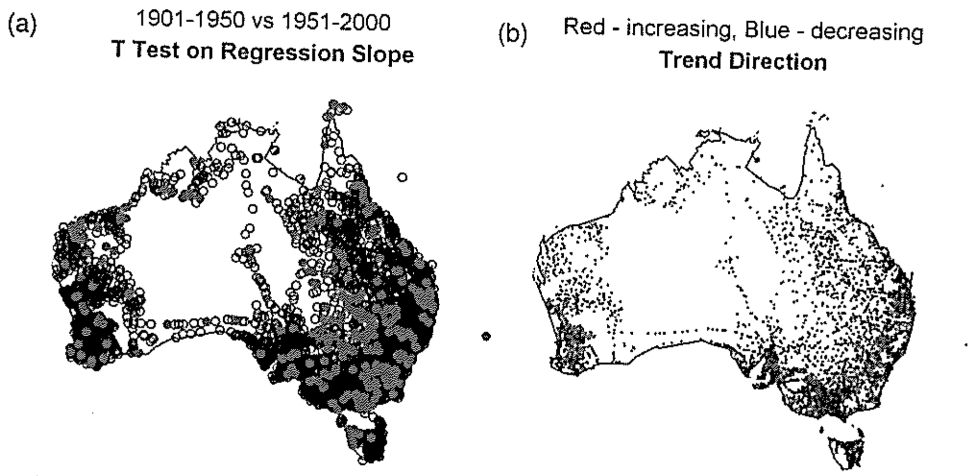

Figure 2 (a) Locations of significant trends (in red) and (b) direction of all trends, increasing trends in red and decreasing trends in blue.

With fewer continuous rainfall stations available for analysis, it is possible to examine the results at individual locations. Figure 3 shows the trends found at each station for each of the 12 durations analysed for the period 1956 to 2005. Approximately 65% of the calculated trends are positive (i.e. increasing annual maximum rainfall), although only 12% are significant at the 95% level. Very few of the negative trends at individual stations or for particular durations were found to be significant (only 1% of the total number of trends overall). Stn040004 stands out having 7 durations with significant negative trends.

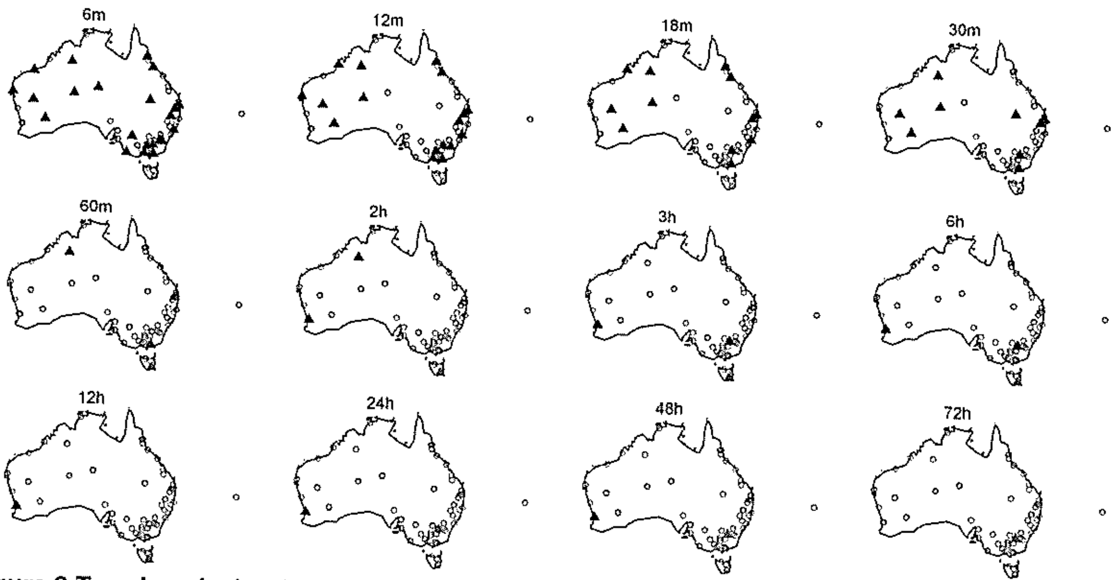

Figure 3 Trend analysis of continuous rainfall stations - sites with decreasing trends are shown in red, sites with increasing trends are shown in blue. Sites with significant trends are shown as filled triangles.

Generally significant trends are found for durations less than 1 hour, with very few stations reporting significant trends in the longer duration annual maximum series. There do not appear to be any clear spatial patterns in the locations of the positive or negative trends (both significant and not significant). Results from the Wilcoxon rank sum test and the Kolmogorov – Smirnov test are similar to those from the t-test on the trend.

The second test that was carried out on the point data was to examine the stationarity of the fitted probability distributions. The GEV distribution has been fit using L moments; fitting the distribution parameters separately to each half of the period. The Kolmogorov-Smirnov test was then used to test if the data from one half period was consistent with the distribution fit to the other half period as the null hypothesis. It was found that 819 stations had data in the period 1951-2000 that were inconsistent with the probability distribution fitted for the period 1901-1950, whilst 758 stations had data from the first period inconsistent with the GEV distribution fitted to the later period. Thus the percentage of stations where the assumption of stationarity is questionable is higher than the results from the trends in the annual maximum series, and is around 20%.

The final test on the stationarity that was carried out involved testing the quantile estimates from the fitted GEV distributions are consistent. Confidence limits on the quantile estimates were estimated by bootstrapping the annual maximum series 100 times and re-fitting the GEV distribution on each bootstrap sample and calculating the quantile estimates for AEPs of 1 in 1.25, 2, 5, 10, 20, 50 and 100. The difference in the quantile estimates from the two periods was then calculated for each bootstrap sample and the 2.5% and 97.5% quantiles on these differences estimated to construct the 95% confidence interval on the differences. If this 95% confidence interval does not cover zero, then we can conclude that the quantile estimates are significantly different from each other. Approximately 10% of daily rainfall stations have significant differences in the rainfall quantile estimates. Figure 4 shows the spatial pattern of these significant differences.

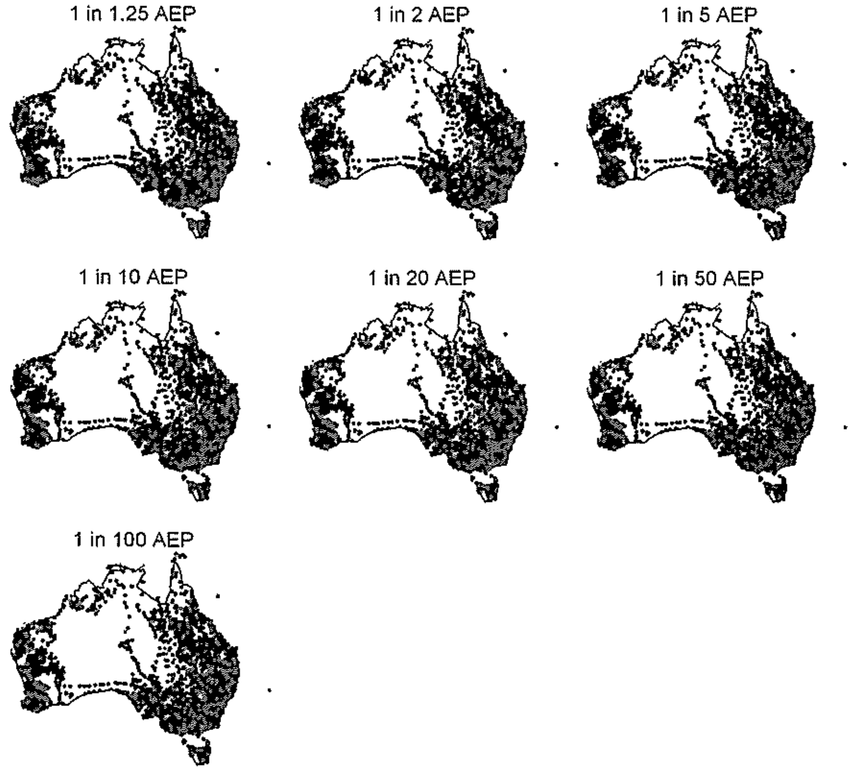

Figure 4 Maps showing locations (in red) where data rainfall quantile estimates do not overlap for a range of AEP events for 1901-1950 vs 1951-2000.

For the continuous rainfall stations, there are many more durations to consider than for the daily stations. The results of the probability distribution stationarity tests are summarized in Table 1. The first two columns summarise the results of the Kolmogorov-Smirnov test. The percentage of stations showing differences in the probability distributions decreases for longer durations. Up to 50% of stations have different probability distributions for the first and second 25 year periods. The final four columns of the table show how these differences translate into rainfall quantile estimates. It is clear that although the probability distributions may be different, the rainfall quantile estimates are generally similar. This is due to the wide confidence limits resulting from only using 25 years of data to fit the probability distribution.

Table 1 Summary of the percentage of stations with differences in GEV distributions and non overlapping rainfall quantile estimates

Duration	1st period data different from 2nd period GEV dist	2nd period data different from 1st period GEV dist	Confidence limits do not overlap	1 in 1 AEP	1 in 10 AEP	1 in 100 AEP
6 mins	48.3	39.7	29.3	1.7	15.5	10.3
12 mins	34.5	31.0	25.9	3.4	15.5	5.2
18 mins	36.2	25.9	22.4	3.4	13.8	3.4
30 mins	19.0	19.0	17.2	1.7	5.2	8.6
60 mins	17.2	10.3	6.9	0.0	5.2	1.7
2 hours	12.1	10.3	3.4	0.0	1.7	0.0
3 hours	20.7	17.2	3.4	0.0	0.0	3.4
6 hours	13.8	15.5	3.4	0.0	1.7	0.0
12 hours	8.6	15.5	1.7	0.0	0.0	1.7
24 hours	10.3	15.5	3.4	0.0	0.0	3.4
48 hours	12.1	13.8	5.2	0.0	3.4	1.7
72 hours	6.9	19.0	6.9	1.7	3.4	3.4

3.2. Regional assessment of the stationarity of annual maximums

Figure 5 shows the direction and whether the trend is significant at the 5% level of the trends in each district for the period 1909 to 2008 for the 1 in 1.25 AEP event and the 1 in 100 AEP events. Results are similar for other AEP events. Generally increasing trends (i.e. more exceedances of the AEP threshold over time) are found in the western and northern parts of Australia with decreasing trends generally limited to south-eastern Australia. However, in most cases the trends were found not to be significant. Trends in the rarer AEP events are also likely to be less significant, which is not unexpected as the time series will be noisier due the infrequent number of threshold exceedances. Results using Spearman's rank correlation to test for the significance of the trends and results for the other periods of record are broadly similar. The analysis was also repeated using different periods of analysis and the results (not shown here) indicate that the trend of average number of exceedances per district is very sensitive to the period of analysis. Generally larger trends (both increasing and decreasing) are found in the shorter periods. It is therefore possible that multi-decadal processes may be affecting the occurrence of large rainfall events.

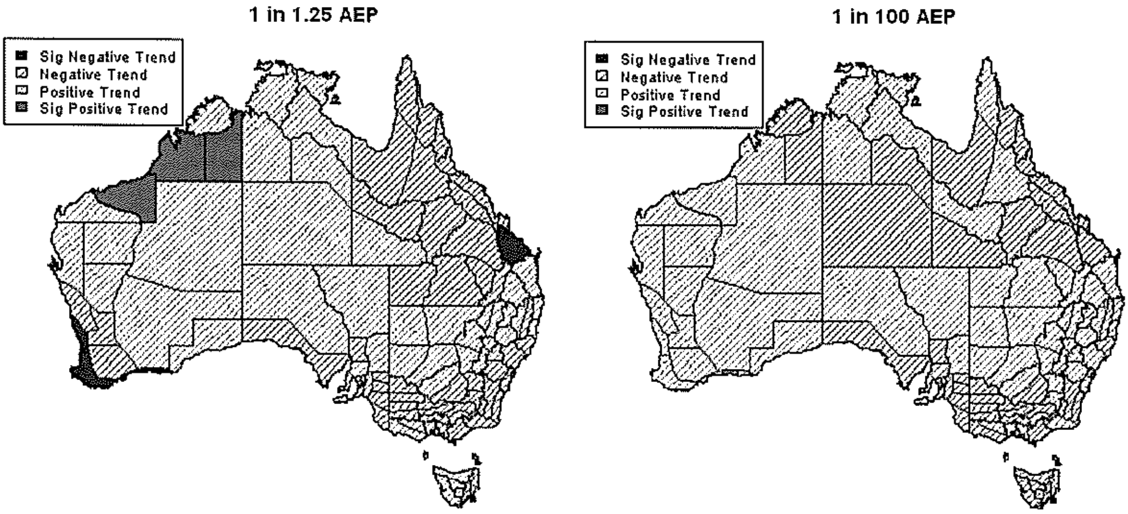

Figure 5 Trends in the average number of exceedance of the design rainfall threshold for the 1 in 1.25 AEP and 1 in 100 AEP event in each district, showing trend direction and if a t-test indicates that the trend is significant at the 5% level

As for the analysis of the annual maximum time series, it was found that the distributions fitted to shorter duration rainfall events are more likely to be different for the two periods than for the longer durations. However even when data from one period are inconsistent with the distribution fit to the other period, the estimated rainfall quantiles are often not significantly different for the two periods, due to the width of the confidence limits on the estimates. The confidence limits on the rainfall quantiles are relatively wide due to the short record lengths used to fit the GEV distributions (at best 25 years of data).

Table 2 Summary of trend significance (p values for t-test on regression line slope)

AEP (1 in x)	District 31	District 40	Districts 39 and 40	Districts 85 and 86
6 minutes				
1	0.25	0	0	0.31
2	0.03	0	0	0
5	0.76	0.06	0.4	0
10	0.5	0.95	0.41	0.02
20	0.44	0.57	0.23	0.04
50	0.58	0.19	0.44	0.3
100	0.85	0.13	0.43	0.15

AEP (1 in x)	District 31	District 40	Districts 39 and 40	Districts 85 and 86
60 minutes				
1	0.44	0.43	0.61	0.37
2	0.84	0.09	0.38	0.08
5	0.28	0.33	0.11	0.05
10	0.64	0.59	0.48	0.85
20	0.6	0.48	0.29	0.67
50	0.71	0.46	0.05	0.16
100	0.34	0.17	0.03	0.9

4. DISCUSSION AND CONCLUSIONS

The stationarity analysis has found that although records at some stations show significant changes over time in both annual maximum time series and fitted GEV distribution and estimated design rainfall quantiles, there are no clear patterns to the changes spatially, with decreasing and increasing trends occurring in similar areas. Significant changes were found at between 2% and 30% of stations analysed depending on the period and method of analysis. Considering the results from both the continuous and daily read rainfall stations, it would seem that significant trends are more likely to occur for shorter duration events (sub-hourly). In general trends are also smaller than the uncertainty associated with the fitted probability distributions.

This study has also applied the NOAA methodology to continuous and daily read rainfall gauges in Australia, grouped by rainfall district. Trends in average exceedances of design rainfall thresholds over time are generally not significant particularly for the daily duration events. Trends at the district level are strongly influenced by the period of analysis, indicating that multi-decadal climate variability may affect extreme rainfalls. Conclusions are difficult to draw for the sub-daily rainfall events due to the sparse spatial distribution of continuous rainfall stations with long records. Overall it is concluded that the full historical record should be used in deriving IFD estimates.

5. REFERENCES

Bonin, G., Maitaria, K. & Yekta, M. (n.d.) Trends in Heavy Rainfalls in the Observed Record in Selected Areas of the U.S.

Bonin, G., Martin, D., Lin, B., Parzybok, T., Yekta, M. & Riley, D. (2006) *NOAA Atlas 14 Precipitation-Frequency Atlas of the United States Volume 1 Version 4.0: Semiarid Southwest (Arizona, Southeast California, Nevada, New Mexico, Utah)*, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.

Hosking, J. R. M. & Wallis, J. R. (1997) *Regional frequency analysis: an approach based on L-moments,* Cambridge, Cambridge University Press.

Huntington, T. (2006) Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology,* 319, 83-95.

Institution of Engineers Australia (1987) *Australian Rainfall and Runoff,* Canberra, The Institution of Engineers Australia.

Power, S., Tseitkin, F., Mehta, V., Lavery, B., Torok, S. & Holbrook, N. (1999) Decadal climate variability in Australia during the twentieth century. *International Journal of Climatology,* 19, 169-184.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignore, M. & Miller, H. (2007) *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change 2007,* Cambridge, Cambridge University Press.

Ummenhofer, C., Gupta, A., Pook, M. & England, M. (2008) Anomalous rainfall over southwest Western Australia forced by Indian Ocean sea surface temperatures. *Journal of Climate,* 21, 5113-5134.