Revision of the Short Duration Intensity-Frequency-Duration (IFD) Design Rainfall Estimates for Australia

J.H. Green[1], F.M. Johnson[2] and C. The[2]
[1]Climate and Water Division
Bureau of Meteorology
Melbourne Vic 3008
AUSTRALIA
[2]Climate & Water Division
Bureau of Meteorology
Sydney, NSW 2000
AUSTRALIA

Abstract: *The current Intensity-Frequency-Duration (IFD) estimates were developed by the Australian Bureau of Meteorology 20 years ago using information from the Bureau's network of rainfall stations, adopting statistical techniques considered appropriate at the time and with a focus on the longer duration IFDs.*

The Bureau is currently revising the IFD estimates with considerable emphasis being placed on the derivation of IFD estimates for sub-daily durations. Firstly, the Bureau's network of pluviograph stations has been significantly augmented by data from the networks operated by large urban water utilities. Secondly, the revision is adopting more statistically rigorous techniques including the use of partial duration series analysis for the estimation of IFDs with Average Recurrence Intervals more frequent than ten years and the adoption of a Generalised Least Squares Regression approach for deriving sub-daily rainfall statistics from daily read data. This paper will present the work being undertaken in revising the short duration IFD estimates.

Keywords: *Design rainfalls, Generalised Least Squares Regression, rainfall data, IFD data*

1. INTRODUCTION

Design rainfall estimates are an essential component of the design of infrastructure including gutters, roofs, culverts, storm water drains, flood mitigation levees and retarding basins. The current Intensity-Frequency-Duration (IFD) estimates for Australia, which are contained in Australian Rainfall and Runoff (AR&R) (Institution of Engineers Australia, 1987) were developed by the Australian Bureau of Meteorology (the Bureau) over 20 years ago using a database comprised primarily of information from the Bureau's network of daily read and continuous rainfall stations and adopting techniques for the statistical analysis of the data that were considered appropriate at the time.

For the current AR&R, the focus of the IFD estimates was primarily for the design of structures on relatively large rural catchments. As a consequence, the extent of the durations considered was from 5 minutes to 72 hours. However, the requirements of the end-users have changed with a significant shift in focus to urban design for small catchments. This has resulted in a corresponding focus on IFDs for shorter durations, in particular, sub-hourly durations down to and including one minute, for adoption with urban drainage design tasks where catchment response times are short.

However, perhaps *the* major problem in deriving the short duration design rainfalls is that Australia suffers from a lack of continuously-recording rain gauges from which to calculate sub-daily rainfall statistics – both in terms of the paucity of the networks and the short record length at individual sites. This paper presents the work being undertaken by the Bureau to address these limitations in revising the short duration IFD design rainfall estimates. Section 2 summarises the approaches that have been adopted previously; Section 3 presents the database of continuous rainfall stations that will be used and Section 4 outlines the proposed approach and discusses the advantages of it over previously adopted approaches.

2. PREVIOUS APPROACHES

The most common approach to deal with insufficient data in rainfall analyses is the application of regional frequency analysis which pools data from many stations in a region. The index rainfall method is a widely adopted regional approach which is based on a regional growth factor and prediction equations for the mean rainfall. The use of Ordinary Least Squares Regression (OLSR) and Partial Least Squares Regression (PLSR) is common with the index rainfall approach. In addition, a technique to derive information on sub-daily durations from daily read gauges is required.

In the estimation of the current IFDs, a Principal Component Analysis (PCA) followed by regression was used to derive equations for predicting design rainfalls for durations shorter than 24 hours for a range of average recurrence intervals (ARI) using information from the 24, 48 and 72 hour durations for the corresponding ARIs. Sub-daily durations down to one hour were estimated by regression of rotated principle components of 1, 2, 3, 6 and 12 hour intensities against principal components of 24, 48 and 72 hour intensities. Maps were subsequently derived for durations of 1 and 12 hours and ARIs of 2 and 50 years.

A pilot study to explore methods for deriving design rainfalls for Australia undertaken by the Bureau (Jakob et al. 2005) used a Partial Least Squares Regression (PLSR) approach to infer statistics for sub-daily durations from statistics at the daily duration based on 190 stations with a minimum record length of 30 years. A PLSR approach can be seen as two linked Principal Component Analyses (PCA), one on the set of independent variables, the other on the set of dependent variables. Partial Least Squares Regression is an extension of PCA whereby a set of linear combinations of the predictors and the predicants is constructed to use for the regression. The difference from PCA is that PCA only creates the set of linear combinations for the predictors and uses the predictant as is. The rationale for adopting PLSR was that it can deal with a large number of independent / dependent variables.

The approach taken in the pilot study attempted to infer the index rainfall, the L-CV and the L-skewness at sub-daily time steps from these statistics at longer durations (together with additional independent variables - predictors). In determining the optimum set of predictors, the Predictive Residual Error Sum of Squares (PRESS) values were used – the selected predictors for each of the statistics are as follows:

- Index rainfall – latitude, longitude, distance from the coast & index rainfall at 24, 48 & 72 hours
- L-skewness – latitude, distance from the coast & L-skewness at 24 hours
- L-CV – latitude, distance from the coast & L-CV at 24 hours.

Some of the limitations, however, of the PLSR are that it ignores the inter-station correlation and variation in record lengths from site to site. In contrast, use of Generalised Least Squares Regression (GLSR) avoids some of these problems (Stedinger and Tasker, 1985). GLSR accounts for possible cross-correlation and unequal error variance, by constructing an error covariance matrix rather than the assumption in OLSR that the errors have equal variance and are uncorrelated. For example, Tasker and Stedinger (1989) used GSLR procedures to estimate empirical relationships between streamflow and physiographic basin characteristics that account for the "actual length of the gauged records employed in the parameter estimation step, the differences in variances of flows at different sites and possible cross-correlation among concurrent streamflows at the various gauged sites".

In Australia, Haddad et al. (2006, 2008) have applied the Quantile Regression Technique based on GLSR to the development of regional flood estimation models. They found that the approach outperformed other approaches because the GLSR estimators account for differences in the variance of streamflows from site to site due to different record lengths, correlation between concurrent flows, correlation between the residuals and the fitted quantiles, and the model error in the regional model.

More specific to the revision of the IFDs, Madsen et al. (2002, 2009) have derived regional estimates of rainfall intensity duration curves for Denmark using GLSR of partial duration series statistics. A comparison was undertaken to assess the whether PLSR or GLSR was the most suitable for predicting sub-daily rainfall statistics as part of the IFD revision (Haddad and Rahman, 2009).

3. AVAILABLE DATA

The current IFD estimates for Australia were based primarily on the Bureau's network of daily read and continuous rainfall stations which consisted of approximately 7500 daily read rainfall stations with greater than 30 years of data and 600 continuous rainfall stations with greater than 6 years of data.

In the intervening years, the Bureau's network of rainfall stations has been expanded and more than 20 years of additional data have been collected resulting in an increase in the number of stations with sufficient length of record to be included in the analyses. In 2009, the Bureau of Meteorology's Australian Data Archive for Meteorology (ADAM) contained:

- approximately 20 000 daily read rainfall stations (both open and closed) from 1800
- nearly 1500 continuous rainfall stations – using both Dines tilting syphon pluviograph (DINES) and Tipping Bucket Rain Gauge (TBRG) instrumentation.

The location of these rain gauges and the period of record are shown in Figure 1(a) for the daily read stations and Figure 1(b) for the continuous rainfall stations.

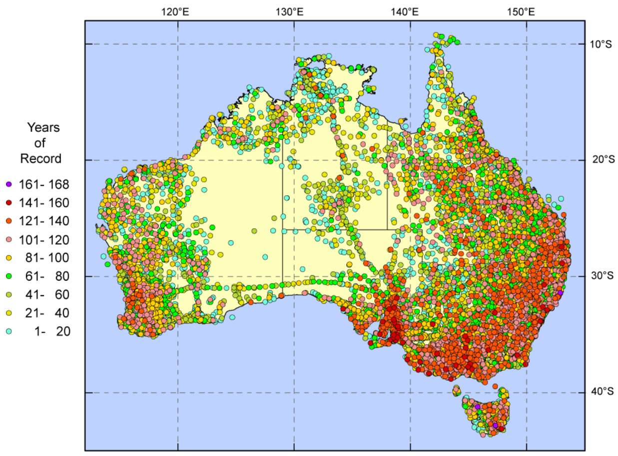

Figure 1(a) Location of Bureau daily read rain gauges and period of record.

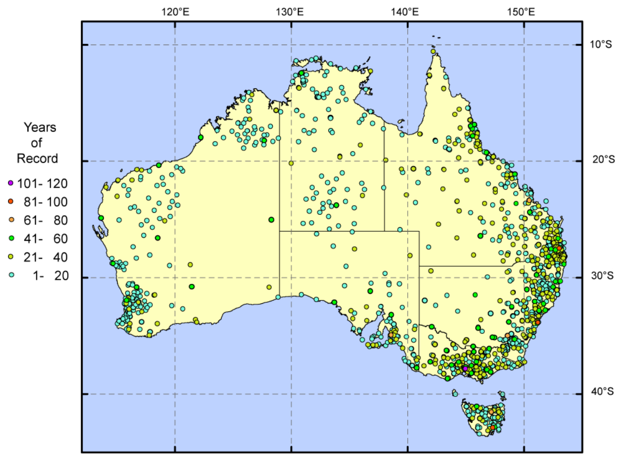

Figure 1(b) Location of Bureau continuous rain gauges and period of record.

In addition, the Bureau now has ready access, under terms of the Water Regulations 2008, to daily-read and continuous rainfall data collected by other organisations. These additional stations supplement the Bureau's network particularly in areas of steep rainfall gradients, for example along the Illawarra escarpment near Sydney in New South Wales.

The Water Regulations 2008 identified approximately 260 'persons' who are required to give to the Bureau, water information that they have in their possession, custody, or control. Of the 260 data custodians, 74 have indicated that they possess data from daily read rainfall stations and 45 that they have data from continuous rainfall stations. Based on the information from an inventory of data providers it is expected that the following additional rainfall data will be received via the Water Regulations:

- ~ 350 daily read rainfall stations
- ~ 2175 continuous rainfall stations.

The effect of more than doubling the number of continuous rainfall stations upon which the IFD estimates are derived will be a significant improvement in the accuracy and representativeness of the IFD estimates, especially for sub-daily durations and in urban areas.

The location of these rain gauges and the period of record are shown in Figure 2(a) for the daily read stations and Figure 2(b) for the continuous rainfall stations.

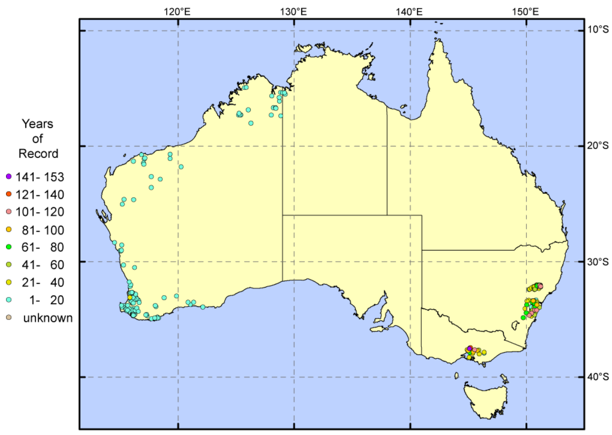

Figure 2(a) Location of Water Regulations daily read rain gauges and period of record.

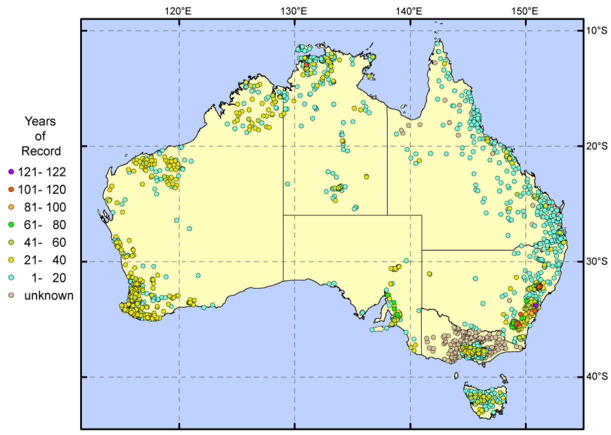

Figure 2(b) Location of Water Regulations continuous rain gauges and period of record.

4. PROPOSED APPROACH

The GLSR is the preferred regression method to be adopted for deriving sub-daily statistics from daily data. Comparison with previously adopted approaches such as PCA & PLSR has found that the GLSR has the following advantages:

- GLSR accounts for possible cross-correlation and unequal error variance, by constructing an error covariance matrix;
- The regression model based on GLSR can explicitly account for sampling uncertainty and intersite dependence; and
- GLSR distinguishes between model and sampling error which will yield a more reliable assessment of uncertainty in the revised IFD estimates.

The GLSR will be adopted for durations of 1, 2, 3, 6 and 12 hours. For sub-hourly durations, where available data are scarcer, both the GLSR and a ratio approach will be trialed. The predictor set and approach to be adopted are presented in the following sections.

4.1. Predictor set

The potential predictors selected to be used for the GLSR model are listed below and comprise both station and geographic data:

- Location (latitude and longitude)
- Elevation
- Slope
- Aspect
- Facet (slope orientation)
- Distance from the coast
- Mean annual rainfall
- Rainfall statistics for daily durations - index rainfall (mean), L-CV and L-skewness.

Daily read and continuous rainfall stations were chosen to be included in the GLSR modeling if there was at least eight effective years of rainfall data available. This resulted in a total of 12960 daily stations and 777 continuous stations being used for the analysis. The background and sources for the predictors at each station locations are discussed below.

4.1.1. Station data

For each station, the Bureau sites metadata database (SitesDB) includes details of the station's location in latitude, longitude and elevation. For stations that are no longer operational, the station metadata may not be as reliable as that available for open stations; in these instances the station elevation was extracted based on the latitude and longitude from the Geoscience Australia 9 second DEM.

Gross error checks on station locations and elevation were performed by comparing DEM derived elevations to those recorded in the station metadata. Checks of latitude and location were also carried out by plotting the latitudes and longitudes in GIS, although the only stations that will be clearly wrong are those that plot off the coast of mainland Australia. In all 308 stations were found to have either locations or elevations that needed to be revised. Revisions to station locations or elevations were carried out using Google Earth and information on the station provided in SitesDB.

Mean annual rainfall was calculated at each station by averaging daily rainfalls over the entire period of record at each station. Finally the rainfall statistics were calculated at each site for daily durations (24 hour, 48 hour and 72 hour). The rainfall statistics were based on the Partial Duration Series (PDS) for each duration, with the index rainfall (mean of the PDS), L-CV and L-skewness calculated. The PDS from the daily rainfall stations were first converted to unrestricted rainfall totals using a conversion factor of 1.15 (Jakob et al. 2005). The PDS threshold and number of years data used in the PDS were also recorded. For all stations, the PDS contained three times the effective station length in years.

4.1.2. Geographic data

The other predictors required for the GLSR model are geographically based and were derived using GIS. The distance to the coast was found in decimal degrees using Geoscience Australia's Framework Boundary to define the coastline.

The values of the remaining predictors to be used in the GLSR model (slope, aspect and facet) are dependent on the scale of the DEM used for the calculations. Very little research appears to have been carried out on the prediction of sub-daily rainfall statistics from daily data. However there has been a significant amount of work carried out around the world on the mapping of longer duration rainfall data (e.g. daily, annual or climatological means) and the influence of topographic features such as slope or aspect on reducing prediction errors.

Based on the literature reviewed, it was decided to use a DEM resolution of 0.05°, which is equivalent to a grid of approximately 5 km to calculate the slope and aspect. For the initial investigations, traditional definitions of slope and aspect were used. However sensitivity studies are proposed on the use of the continuous position functions from Hutchinson (1998).

The final geographic predictor used as an input to the GLSR model is facet. The Parameter-elevation Regressions on Independent Slopes Model (PRISM), developed by Oregon State University to produce spatial climate products, uses point climatic data combined with a DEM and other spatial data sets to produce mapped climate data at monthly or annual timescales or for events (PRISM Climate Group, 1998). In PRISM, "facets" are defined such that the slope orientation is reasonably constant over a region. It is believed that the relationship between elevation and precipitation will be strengthened by controlling for factors such as the slope orientation, predominant wind direction and the steepness of the terrain (Daly et al. 1994). The definition of facet allows for stations to be grouped via these properties. For the current study, a DEM resolution of 6 minutes was adopted for the facet calculations, which was chosen because it is an even multiple of the base 9 second DEM used to the project and is close to the PRISM DEM resolution. Up to 120 filter passes were used, although there is little difference in aspect between the 40 filter pass DEM and 120 filter pass DEM. A minimum number of stations per facet were set at 15.

4.2. GLSR Method

The predictor set extracted as described above will be used to derive sub-daily statistics from daily data using the method summarised below.

Ordinary Least Squares Regression (OLS) summarises the relationship between two or more variables by using linear combinations of one or more variables to predict the value of the remaining variables through the solution of Equation 1.

$$y = X\beta + \varepsilon \qquad (1)$$

However, OLSR requires some strong assumptions regarding the independence of the variables and the variance of the errors. GLSR accounts for possible cross-correlation and unequal error variance, by constructing an error covariance matrix rather than assuming that the errors are equal and uncorrelated. For GLSR, Equation 1 is solved as for OLSR; however, in this case the solution for β includes model errors that have possibly different variances and allows for cross correlations. The GLSR solution for β is:

$$\beta = \left(X^T \Lambda^{-1} X\right)^{-1} X^T \Lambda^{-1} y \qquad (2)$$

where the error covariance matrix is Λ and the errors have a mean expectation of zero. The question with GLSR then is to find Λ, in theory independently of the data which corresponds to the assumption that Λ is known. The error covariance matrix is composed as:

$$\Lambda = \Gamma + \Sigma \qquad (3)$$

where Γ = diag(γ_i^2) and γ_i^2 is the model error variance for site *i* due to an imperfect model and Σ is the sampling covariance matrix for the estimator. The variance estimates of the model will therefore reflect the regional uncertainty that cannot be explained by the regression model (Γ) as well as sampling uncertainty corrected for inter-site correlation (Σ) (Madsen et al. 2009)

Stedinger and Tasker (1985) proposed an approximation of the true covariance matrix by using a constant value of γ^2, and smoothing the sampling variances to approximate Σ, assuming an average value of cross-correlation between concurrent flows. Further work (Tasker and Stedinger, 1989) allowed for varying model error variance by site (i.e. γ_i^2 can be different for each site *i*), and the cross-correlation used in estimating Σ was smoothed by distance rather than selecting an average value.

If only β_0 is used in the regression model (i.e. the only predictor is a constant term) then the GLS regression model provides an estimate of the regional average parameter value and its associated uncertainty. In general this will be different from the arithmetic average because the GLS algorithm weights the parameter estimates according to the covariance matrix of the errors (Madsen et al. 2002).

In estimating the regional average parameter value, an estimate of regional homogeneity can also be obtained. Madsen et al. (2002) assume that the model error variance (γ_i^2) is the same for all sites. If it is found that γ^2 is equal to zero the region can be considered homogenous; if γ^2 is greater than zero then there is regional heterogeneity which may be able to be explained using a GLS regression with additional predictors (Madsen et al. 2002).

5. SUMMARY

For the derivation of sub-daily statistics from daily data a comparison has been undertaken of the PCA, PLSR, and GLSR methods. The results showed that the GLSR-based prediction equations performed as well or better than the PLSR ones with respect to the coefficient of determination values (R^2) for 1 to 12 hour durations (Haddad and Rahman, 2009).

It has been concluded that the GLSR is the most appropriate as it can explicitly account for sampling uncertainty and intersite dependence. Station and geographic data have been extracted from both

Bureau of Meteorology rainfall stations and those operated by other agencies to produce the predictor set. These data will now be applied using the GLSR method.

6. ACKNOWLEDGMENTS

The revision of the IFD estimates that is being undertaken by the Bureau of Meteorology is a team effort and the input of the individual team members is gratefully acknowledged: Catherine Beesley, Shirrah Comeadow, Jeanette Meighen, Garry Moore, Brian Taylor, and Karin Xuereb. The input of Ataur Rahman and Khaled Haddad from the University of Western Sydney on the GLSR is greatly appreciated.

7. REFERENCES

Daly, C., Neilson, R.P. and Phillips, D.L. (1994). A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal of Applied Meteorology* **33**: 140-158.

Haddad, K., Rahman, A. and Weinmann, P.E. (2006). Design flood estimation in ungauged catchments by quantile regression technique: ordinary least squares and GLS compared. In *Proc. 30th Hydrology and Water Resources Symp.*, 4-7 Dec 2006, Launceston, 6pp.

Haddad, K., Rahman, A. and Weinmann, P. E. (2008). Development of a Generalised least squares based quantile regression technique for design flood estimation in Victoria, *31st Hydrology and Water Resources Symp.*, Adelaide, 15-17 April 2008, 2546-2557.

Haddad, K. and Rahman, A. (2009) *A Pilot Study on Design Rainfall Estimation Using Generalised Least Squares Regression*, University of Western Sydney.

Hutchinson, M. F. (1998). Interpolation of Rainfall Data with Thin Plate Smoothing Splines - Part II: Analysis of Topographic Dependence. *Journal of Geographic Information and Decision Analysis* **2**(2): 152-167.

Institution of Engineers, Australia (1987). *Australian Rainfall & Runoff - A Guide to Flood Estimation*, Institution of Engineers, Australia, Barton, ACT, 1987.

Jakob, D., Taylor, B. and Xuereb, K. 2005. *A pilot study to explore methods for deriving design rainfalls for Australia – Part 1*, Hydrology Report Series, HRS Report No, 10.

Madsen, H., Mikkelsen, P.S., Rosbjerg, D and Harremoes, P. (2002). Regional estimation of rainfall intensity duration curves using generalized least squares regression of partial duration series statistics. *Water Resources Research* 38 (11), 1-11.

Madsen, H., Arnbjerg-Neilsen., K., Mikkelsen., P.S. (2009). Update of regional intensity-duration-frequency curves in Denmark: Tendency towards increased storm intensities. *Atmospheric Research*, (92) 343-349.

PRISM Climate Group (1998). Climate Mapping with PRISM. http://www.prism.oregonstate.edu/pub/prism/docs/prisguid.pdf

Stedinger, J.R., and Tasker, G.D. (1985). Regional hydrologic analysis, 1.Ordinary, weighted, and generalised least squares compared, *Water Resources Research* 21(9):1421:1432

Tasker, G. D. and J. R. Stedinger (1989). An Operational GLS Model for Hydrologic Regression. *Journal of Hydrology* 111: 361-375.