Australian Application of a Statistical Downscaling Technique for Multi-Site Daily Rainfall: GLIMCLIM

Frost, A. J.

Hydrology Unit, Australian Bureau of Meteorology
eWater Cooperative Research Centre
Email: a.frost@bom.gov.au

Keywords: *Stochastic downscaling, daily rainfall, generalised linear models*

EXTENDED ABSTRACT

Climate variability, both natural and that introduced by anthropogenic activity, is of major concern to water resource planners within Australia. The current long lasting drought affecting Australia is causing scarcity of rainfall to such an extent that many regional potable water supplies are reaching critical lows. Given the importance of water, models which predict the influence of natural and anthropogenic factors on rainfall – the dominant hydrological driver - on a regional basis are urgently required for adequate assessment of water supply risk.

Large spatial scale predictions of (typically 300 to 500 km grids) global scale climate scenarios output by General Circulation Models (GCM) are inadequate for use in such studies as they do not capture the large degree of spatial variability over smaller distances which is inherent in rainfall. Multi-site daily rainfall – a common requirement within many hydrological models - is a required input for modelling complex multi-catchment systems, as small scale spatial variability due to factors such as topography has a large bearing on how much rainfall falls in a given area. Statistical downscaling is a technique which can produce such fine spatial scale rainfall pattern predictions conditional on the larger scale climate scenarios output by a GCM.

This paper details the application of the GLIMCLIM (Generalised Linear Model for daily Climate time series) software package over a set of 30 selected sites spread over an area of 200,000 km^2 in south-eastern Australia. This model has been used in much smaller scale downscaling studies in the UK (Frost et al., 2006; Leith, 2005), but as yet has not found application to Australian rainfall.

NCEP reanalysis data is used to provide predictors for model simulation. Statistics relating to the 'occurrence' and 'amounts' models showed satisfactory performance given the simplicity of the models specified – with wet-spell/dry-spell mean/standard deviation reproduced well (Figure 1), along with wet day mean/standard deviation. The correlation decay with distance between site rainfall occurrences were shown to be reproduced poorly (Figure 2), with further work on model specification required. Correlation between site amounts on the other hand was reproduced well.

On the whole, given the degree of heterogeneity displayed across the area trialled, the model provides promising results as a method for providing multi-site daily rainfall simulation conditioned on large scale atmospheric outputs.

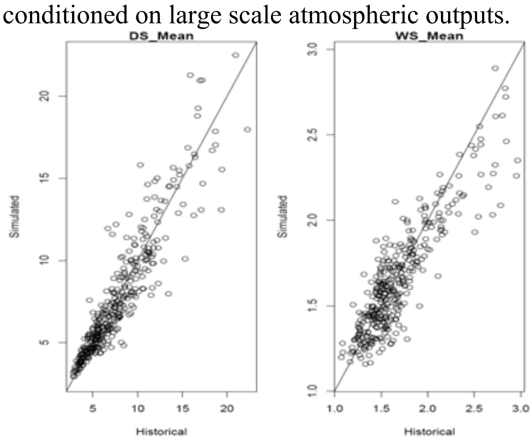

Figure 1. Historical vs. simulated monthly site mean a) dry-spell duration and b) wet-spell duration (days).

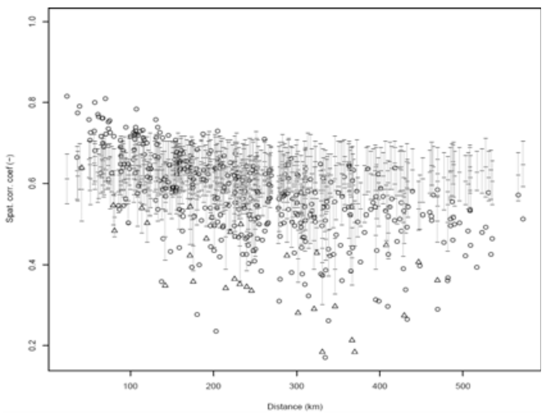

Figure 2. Correlation of occurrences (July) vs. distance between sites. Historical-circles/ triangles, Simulated-Bar (2.5%, 50% and 97.5%).

1. INTRODUCTION

Multi-site downscaling of rainfall is a maturing field with many recently proposed statistical methods (eg. Fowler et al., 2005; Haylock et al., 2006; Vrac and Naveau, 2007; Wetterhall et al., 2006) with some also finding Australian application (Charles et al., 2004; Hope et al., 2006; Mehrotra and Sharma, 2006; Timbal, 2004). The methods of interest in this study are those which can be used to provide multiple simulations for hydrological risk estimation which explicitly account for the uncertainty and variability of precipitation given large scale conditioning variables.

Downscaling methods can usually be classified as either relying on division into climate states and/or reliance on non-parametric distributional assumptions. Typically these models are calibrated and validated on a 6 month seasonal basis (using large scale test indices), with little monthly at-site validation presented. Admittedly validation on such a fine-scale is difficult given the length constraints in the literature, and the varying possible contexts in which the output can be used. However, it is unclear from the literature whether these models adequately reproduce the total monthly variations evident throughout the year and from site-to-site, as may be required for hydrologic models reliant on such input.

The Generalised Linear Model (GLM) for daily Climate time series provides an alternative conceptualisation of the rainfall process and has been used to analyse and simulate spatial daily rainfall given natural climate variability influences in the UK (Chandler and Wheater, 2002; Yang et al., 2005) and further to predict the influence of various future climate scenarios on regional rainfall by downscaling larger spatial scale GCM simulations (Frost et al., 2006). Rather than relying on division into 'states' the GLIMCLIM relies on a linear regression like structure (inherent to GLMs) for both rainfall occurrence and the amounts falling on wet days. This work details the application of this method of downscaling NCEP reanalysis data (Kalnay et al., 1996) under Australian conditions.

2. MODEL DESCRIPTION

A brief description of GLM methodology for daily rainfall is given here following that given within Chandler and Wheater (2002) and Yang et al. (2005). The methodology broadly follows a two stage approach to modelling daily rainfall relating to 1. occurrence and 2. amount associated with wet days.

The pattern of wet and dry days at a site is modelled using logistic regression. Let p_i denote the probability of rain for the ith case in the data set, conditional on the covariate vector $\mathbf{x}_i'$; then the model is given by

$$\ln\left(p_i/(1-p_i)\right) = \mathbf{x}_i'\boldsymbol{\beta} \qquad (1)$$

for some coefficient vector $\boldsymbol{\beta}$.

Gamma distributions are fitted to the amount of rain on wet days. The rainfall amount for the ith wet day in the database is taken, conditional on a covariate vector $\boldsymbol{\xi}_i'$, to have a gamma distribution with mean μ_i where

$$\ln \mu_i = \xi_i'\boldsymbol{\gamma} \qquad (2)$$

for some coefficient vector $\boldsymbol{\gamma}$.

These two models are referred to as 'occurrence' and 'amounts' models respectively. The right hand sides of (1) and (2) are called 'linear predictors'. Estimation of coefficient vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, and selection of such predictors can be carried out using likelihood measures – see Chandler (2002) for details.

Table 1. Daily rainfall sites used in study and proportion of record missing.

Site #	Site Name	Altitude (m)	% Missing
049048	Balranald	58	0.6
070014	Canberra Airport	578	0.0
070028	Yass	595	0.3
070054	Cooma	870	0.1
072150	Wagga Wagga	212	0.0
073007	Burrinjuck Dam	390	0.1
074087	Urana	115	0.1
075049	Maude	75	0.4
072019	Holbrook	345	0.8
072023	Hume Reservoir	184	0.1
073051	Murringo	420	0.3
074008	Grong Grong	159	2.8
074025	Burrumbuttock	240	3.1
075012	Wakool	84	0.0
075054	Conargo	97	0.1
075067	Carrathool	104	2.1
076044	Nyah	80	1.2
077001	Barraport North	111	0.1
080044	Patho West	90	0.5
080053	Tandarra	110	0.7
081019	Goulburn Weir	122	1.1
082002	Benalla	170	0.0
082018	Gibbo R Park	550	0.9
082127	Peechelba East	140	2.2
083010	Eurobin	275	0.6
083038	Tawonga	295	0.6
088011	Campbelltown	366	0.1
088042	Malmsbury Res.	450	0.0
088060	Wallaby Ck	488	0.2
088131	Narbethong	345	3.4

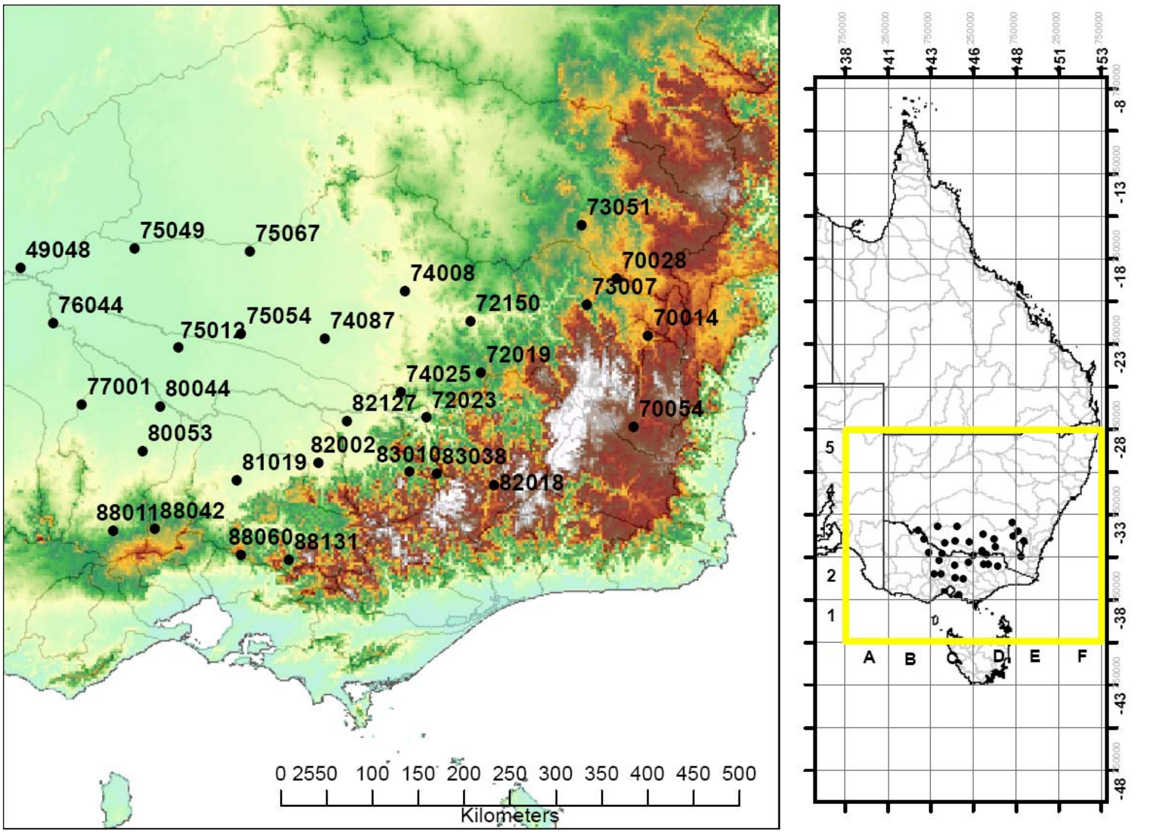

Figure 3. a) Daily rainfall site locations and b) NCEP reanalysis grid. The yellow rectangle corresponds to area over which predictors are used in this study following Charles (pers. comm., 2007b).

Table 2. NCEP reanalysis predictors.

Atmospheric Predictor	Description	NCEP reanalysis GRID (refer to Figure 3(b))
Summer.MSLP	Mean sea level pressure	(B2+B3+B4+C2+C3+C4)/6
Summer.DTD700	Dew Point temperature depression at 700 hPa	(C2+C3+C4+D2+D3+D4)/6
Summer.E-W GPH500	East-West Geopotential Height Gradient at 500 hPa	((C3+C4+D3+D4)-(E3+E4+F3+F4))/4
Winter.N-S MSLP	North-South Mean seal-level Pressure gradient	((A5+B5+C5+D5)-(A4+B4+C4+D4))/4
Winter.DTD700	Dew Point temperature depression at 700 hPa	(B2+B3+B4+C2+C3+C4+D2+D3+D4)/9
Winter.DTD850	Dew Point temperature depression at 850 hPa	(A5+A4+B3+B4+C3+C4)/6
Winter.N-S GPH700	North-South Geopotential Height Gradient at 700 hPa	((A5+B5+C5)-(A4+B4+C4))/4

2.1. Specification of predictors

The choice of predictors within GLIMCLIM is required for both the 'occurrence' and 'amounts' models. Typically model specification might include the fitting of a smooth Legendre polynomial (LP) for site effects (to accommodate the non-homogeneity displayed across a region not explained by the input predictors), a sinusoidal seasonal component, an altitude component, and terms relating to previous days rainfall (to account for autocorrelation of the process). The fact that the effect of one predictor may depend on the values of others can be accommodated within the GLIMCLIM with the introduction of 'interaction' parameters between two specified components. Terms are accepted into the model following the procedure as specified within Chandler (2002).

Exogenous atmospheric variables such as the NCEP reanalysis dataset can also be added as predictors to GLIMCLIM - see Leith (2005) and Frost et al. (2006) for examples. Once it is shown that the model sufficiently captures past variability using past site rainfall and the reanalysis data, GCM outputs on the same spatial scale can be used to provide future rainfall scenarios.

3. METHODOLOGY AND DATA

GLMs were fitted to a set of 30 Bureau of Meteorology daily rainfall stations spanning the period 01/01/1986-31/12/2005 – see list of sites in Table 1. These stations were the same as those used in a previous downscaling study by Charles (pers. comm., 2007b) in the SEACI project (www.mdbc.gov.au/subs/seaci) so as to facilitate future comparison. These stations had less than 3.4% missing data over the 20 years, at any one station, and given previous investigation can be considered to be of reasonable quality. Site locations (mapped within Figure 3) show the large area over which downscaling is attempted in this study, with marked distances and changes in altitude between sites. This area is much greater than has been attempted previously using this model for daily rainfall. Daily NCEP atmospheric predictors were selected according to a correlation analysis between atmospheric variables and site rainfall as undertaken by Charles (pers. comm., 2007a). These predictors were calculated from the 2.5˚ by 2.5˚ NCEP reanalysis grid area as plotted in Figure 3(b) according to the formulas presented in Table 2. It is noted that Charles used differing predictors for April-September (Winter) and October-March (Summer). Here, all of these identified predictors are used over the entire year.

4. RESULTS

GLIMCLIM was calibrated using a simple structure for both the occurrences and amounts – see Appendix A. In summary there are LP site effects, an altitude term, an annual sinusoidal seasonal term, some at-site auto-correlative effects and all of the 7 atmospheric predictors previously mentioned. All predictors were found to be significant with the exception of Winter.N-S GPH700 for the occurrence model, and the Summer DTD700 and N-S MSLP for the amounts model. Notably, there were no interaction terms included apart from those relating solely to the site effects LP. This model thus serves as a basis on which to build a more meaningful model where there are interactions between these parameters – depending on the results of validation.

Once the model was calibrated, it was then used in simulation mode to produce 100 replicates with the same length as the observed series. A set of annual, monthly, and validation statistics were calculated from each of the simulations and compared to the observed statistics for the full fitting period (1986-2005).

4.1. Occurrence model

Figure 4 plots the historical versus median simulated number of wet days (calculated for each site and month), and is reasonably well produced across most sites. The site and monthly pooling to a singular plot can mask specific sites and or seasons where the model may be performing poorly. Two individual sites plots of the mean number of wet days per month are presented in Figure 5 (including 2.5%, 25%, 50%, 75% and 97.5% confidence limits). While the model performs well for Balranald, reproducing the seasonality, the seasonality is not reproduced well for Canberra airport. The majority of sites reproduced the seasonality well. However, approximately 10 of the 30 either showed some consistent bias, or reproduced the seasonality poorly. It is noted that sites at higher altitudes tended to produce the poorest results in terms of seasonality (e.g. Canberra Airport, Cooma, Narbethong, Wallaby Creek).

Wet-spell/dry-spell mean (Figure 1) and standard deviation (not presented) are also reproduced satisfactorily. Monthly dry-spell skewness (Figure 6a) was reproduced reasonably with the majority of historical values falling within the 95% confidence limits (not shown). However, seasonality was not matched closely by the simulated values, resulting in the spread of historical versus median simulated in Figure 6. The wet-spell skewness (Figure 6b) historical values do not match the median simulated closely, with a large degree of variability in the simulated values (not shown). It is hypothesised that the high degree of variability of simulated skewness is produced due the quality of the seasonal fit for lower order statistics (mean, standard devation). It is predicted that the seasonality of skewness would be identified better with a closer fit for those statistics.

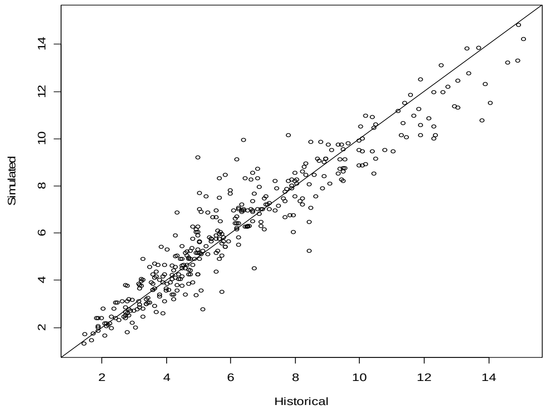

Figure 4. Historical vs. simulated mean number of wet days per month.

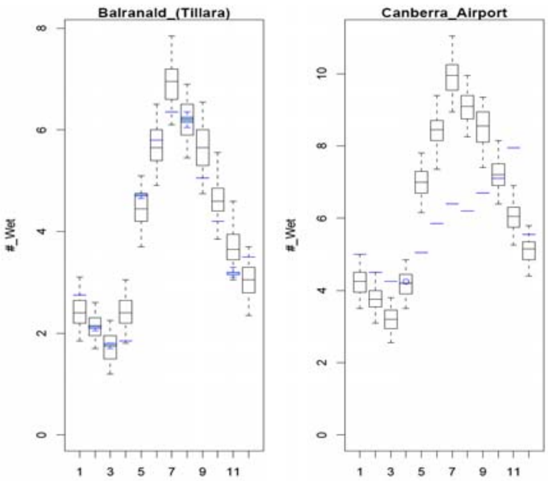

Figure 5. Monthly mean number of wet days for Balranald and Canberra Airport. Observed (Blue) vs. simulated (Black).

The site-to-site correlation between occurrences was also calculated, and is plotted versus distance in Figure 2. July - A month representative of the winter dominated occurrences is shown and is similar to all other months (yet there is variability in slope of correlation decay from month to month). There are many observed values lying outside of the simulated confidence limits, with general underestimation at distances below 200km, and overestimation from 350km upwards. It is noted that the model correctly forecasts dry and wet days 88.7% and 71.6% of the time respectively, which is quite high considering the effects of the poor reproduction of spatial correlation of occurrences.

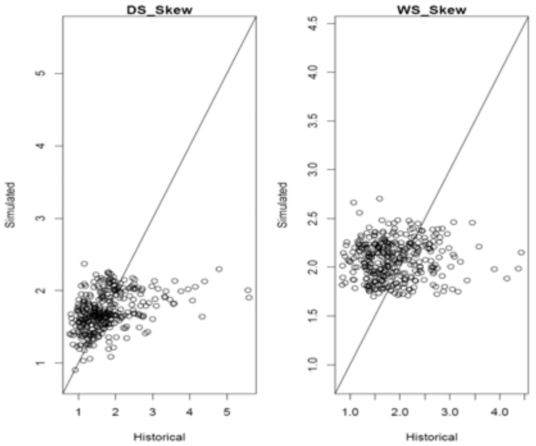

Figure 6. Historical vs. simulated skewness of a) dry-spell duration and b) wet-spell duration.

While the general statistics relating to wet and dry day occurrences such as dry-spell/wet-spell mean and standard deviation are reasonable, there remains clear evidence that the GLIMCLIM does not account for the decrease in correlation of occurrences over distance. This effect could in turn cause the model to be producing poor wet-spell/dry-spell characteristics. The model specified here is very simple, and it appears that spatially varying effects included here such as the altitude term and the LP are insufficient to induce greater spatial coherency between occurrences. Furthermore, the model has been specified such that the seasonality of occurrences is assumed to be constant throughout while this is not the case for some sites in the alpine regions. Also, the ‘occurrence’ correlation between all sites and Wallaby Creek (088060) was markedly lower than all other sites (see triangles in Figure 2), and there is currently no mechanism to incorporate the probable effect of differing weather systems (coupled with topographic features) affecting different areas.

It is noted that a further attempt at fitting a more complex model including interactions between parameters provided improved results in terms of seasonality and bias across all sites. However, whilst also improved to some degree, the problems in reproducing spatial occurrence remain. There is some evidence of poor data quality (rainfall recorded on the wrong days when compared to nearby stations) for some sites (eg. Narbethong). This issue requires further investigation to negate the possible biasing effect it may have on the model calibration.

4.2. Amounts model

The wet (rainfall > 0.95mm as used by Charles, pers. comm., 2007b) monthly mean/standard deviation are reproduced satisfactorily (Figure 7). However, there is some degree of spread – indicating a poor fit for some sites/months. Similarly to the wet-spell/dry-spell skewness, the monthly skewness/correlation of daily amounts is reproduced reasonably by the model (Figure 8), with most values falling between the 95% confidence limits (not shown). However, seasonality was not matched closely by the simulated values, resulting in the spread of historical versus median simulated. Spatial correlation between sites on wet days is reproduced well (Figure 9), which is expected given the use of historical correlation structure in the simulations.

Overall the amounts model performs satisfactorily. There remain some sites where the seasonality is not reproduced well and/or there is constant bias with monthly mean/standard deviation. This result is again in part due to the simplicity of the model specified, with the consistent seasonality, and effects from other predictors (including NCEP atmospheric predictors) across all sites.

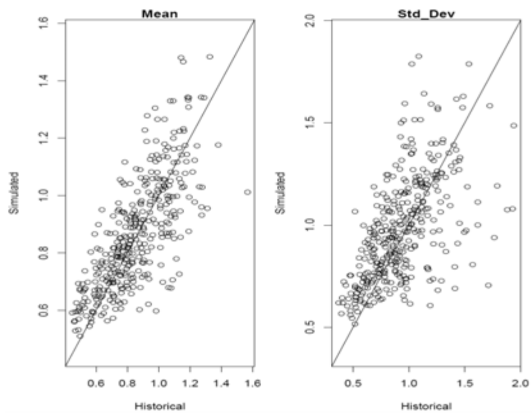

Figure 7. Historical vs. Simulated wet a) mean and b) standard deviation.

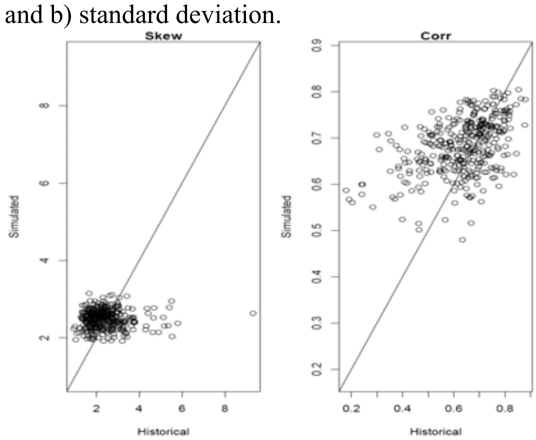

Figure 8. Historical vs. Simulated wet a) skewness and b) autocorrelation.

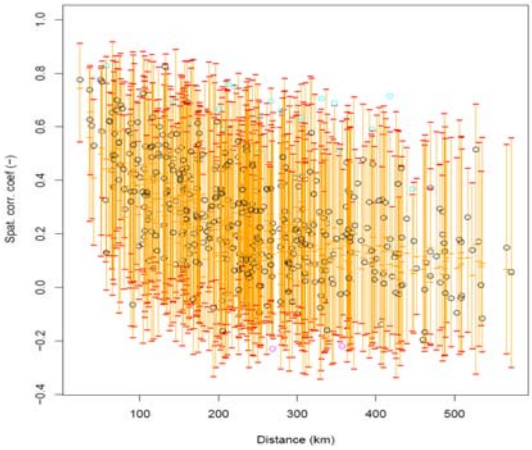

Figure 9. January correlation coefficient of amounts versus distance between sites.

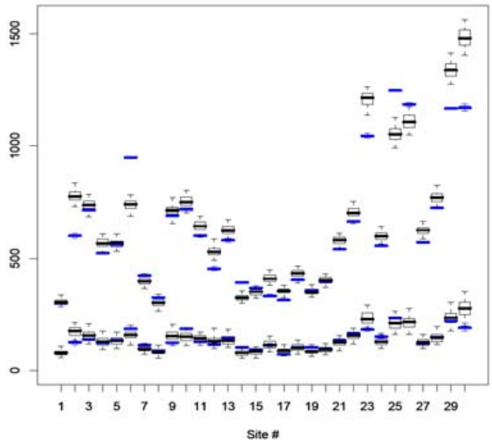

Figure10. Site annual mean (upper) and standard deviation (lower) statistics. Blue – Historical, Black – Simulated.

4.3. Monthly and annual statistics

Although not presented here in detail, as the primary concern here was the performance of the occurrence/amounts models, the annual mean (Figure 10) was reproduced satisfactorily for most sites, although was biased for sites/months where either the amounts and/or occurrence model were biased. The standard deviation was produced to a greater degree of satisfaction, with most sites falling between the simulation confidence limits. Furthermore, monthly time-series plots were produced for each site, to ensure that there were not any periods where the model markedly over- or underestimated rainfall. See Figure 11 for a single site example. Note the reproduction of the very dry drought period (post 1998 - especially January 2004) which is a consistent feature across all sites.

5. CONCLUSION

This paper has presented a first attempt at applying GLIMCLIM to Australian rainfall conditions. NCEP reanalysis data are used as predictors within model simulation. Statistics relating to the 'occurrence' and 'amounts' models showed satisfactory performance given the simplicity of the model specified. Correlations between occurrences was shown to be reproduced poorly, with further work on model specification required. Further investigation regarding the quality of data at some sites is also required. Overall the model shows some promise in providing simulations of multi-site daily rainfall for Australian conditions.

6. ACKNOWLEDGEMENTS

The helpful reviews and discussion of this work with Steve Charles and Bertrand Timbal, along with comments by Richard Chandler, Sri Srikanthan and Ross James are appreciated greatly.

7. REFERENCES

Chandler, R.E., 2002. GLIMCLIM: Generalized linear modelling for daily climate time series (software and user guide). Technical report, no. 227, Department of Statistical Science, University College London.

Chandler, R.E. and Wheater, H.S., 2002. Analysis of rainfall variability using Generalized Linear Models - a case study from the West of Ireland. *Water Resources Research*, 38(10).

Charles, S.P., 2007a. *Personal Communication*.

Charles, S.P., 2007b. *Personal Communication*.

Charles, S.P., Bates, B.C., Smith, I.N. and Hughes, J.P., 2004. Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. *Hydrological Processes*, 18(8): 1373-1394.

Fowler, H.J., Kilsby, C.G., O'Connell, P.E. and Burton, A., 2005. A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change. *Journal of Hydrology*, 308(1-4): 50-66.

Frost, A.J., Chandler, R.E. and Segond, M.-L., 2006. Spatial Temporal Rainfall Modelling with Climate Change Scenarios, DEFRA (UK).

Haylock, M.R., Cawley, G.C., Harpham, C., Wilby, R.L. and Goodess, C.M., 2006. Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology*, In press.

Hope, P.K., Drosdowsky, W. and Nicholls, N., 2006. Shifts in the synoptic systems influencing southwest Western Australia. *Climate Dynamics*, 26: 751-764.

Kalnay, E. et al., 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3): 437-471.

Leith, N., 2005. Using GLMs to simulate daily rainfall under scenarios of climate change. FD2113 project report, no. 2, DEFRA.

Mehrotra, R. and Sharma, A., 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. *Journal of Geophysical Research*, 111(D15101): doi: 10.1029/2005JD00637.

Timbal, B., 2004. Southwest Australia past and future rainfall trends. *Climate Research*, 26: 233-249.

Vrac, M. and Naveau, P., 2007. Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research*, 43(W07402): doi:10.1029/2006WR005308.

Wetterhall, F., Bardossy, A., Chen, D., Halldin, S. and Xu, C.-Y., 2006. Daily precipitation-downscaling techniques in three Chinese regions. *Water Resources Research*, 42(W11423): doi:10.1029/2005WR004573.

Yang, C., Chandler, R.E., Isham, V.S. and Wheater, H.S., 2005. Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research* , 41(W11415).

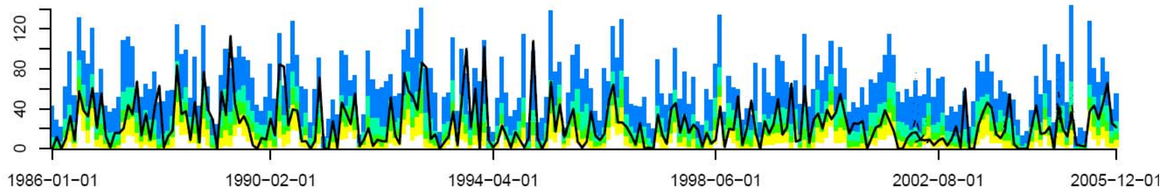

Figure11. Balranald monthly rainfall timeseries (mm): Observed (Black) vs. Simulated (Colours - 2.5%-Yellow-25%-green-50%-aqual-75%-blue-97.5%).

APPENDIX A

Table A. 'Occurrence' and 'amounts' model specification. See Chandler (2002) for explanation of terms.

OCCURRENCES

	Value	C1	C2	C3	Description	#
0	-2.0541				Constant	
1	1.3255	1	31		LP1 for Eastings	1
1	-1.4436	2	31		LP1 for Northings	2
1	0.9422	1	32		LP 2 for Eastings	3
1	0.6698	2	32		LP 2 for Northings	4
1	0.011	1	33		LP 3 for Eastings	5
1	-0.5021	2	33		LP 3 for Northings	6
1	-1.4467	3			Altitude	7
4	-0.2177	21			Daily seasonal effect, cosine component	8
4	-0.2294	22			Daily seasonal effect, sine component	9
4	0.7801	1	2	3	Ln(1+Y[t-1])	10
4	0.2746	3	5		I(Y[t-k]>0: k=1 to 3)	11
4	-0.9287	51	0		Summer.MSLP	12
4	-0.8734	52	0		Summer.DTD700	13
4	-0.2755	53	0		Summer.E-W GPH500	14
4	0.327	54	0		Winter.N-S MSLP	15
4	-0.2413	55	0		Winter.DTD700	16
4	-0.6868	56	0		Winter.DTD850	17
4	0.0114	57	0		Winter.N-S GPH700	18
5	-1.7257	1	2		Interactions	
5	1.3904	1	4		Interactions	
5	-1.3219	1	6		Interactions	
5	-0.693	3	2		Interactions	
5	1.5986	3	4		Interactions	
5	-0.7823	3	6		Interactions	
5	-3.2077	5	2		Interactions	
5	0.437	5	4		Interactions	
5	-3.6283	5	6		Interactions	
7	136	1	1	0	Easting bound 1	
7	702	1	2	0	Easting bound 2	
7	5840	2	1	0	Northing bound 1	
7	6290	2	2	0	Northing bound 2	
8	0.095	1	2		`Soft' threshold for +ve values	
10	2.4889	21	1		Parameter 1 in spatial dependence model	

AMOUNTS

	Val	C1	C2	C3	Description	#
0	-1.3979				Constant	
1	-0.285	1	31		LP 1 for Eastings	1
1	0.595	2	31		LP 1 for Northings	2
1	-0.8376	1	32		LP 2 for Eastings	3
1	0.1672	2	32		LP 2 for Northings	4
1	-0.6273	1	33		LP 3 for Eastings	5
1	0.4327	2	33		LP 3 for Northings	6
1	1.1027	3			Altitude	7
4	-0.1576	51	0		Summer.MSLP	8
4	-0.0975	52	0		Summer.DTD700	9
4	-0.1724	53	0		Summer.E-W GPH500	10
4	0.0329	54	0		Winter.N-S MSLP	11
4	-0.3462	55	0		Winter.DTD700	12
4	-0.1703	56	0		Winter.DTD850	13
4	-0.0704	57	0		Winter.N-S GPH700	14
4	0.2155	21			Daily seasonal effect, cosine component	15
4	0.1439	22			Daily seasonal effect, sine component	16
4	0.1959	1	2	3	Ln(1+Y[t-1])	17
5	0.8398	1	2		Interactions	
5	-0.7581	1	4		Interactions	
5	0.3763	1	6		Interactions	
5	0.1337	3	2		Interactions	
5	-1.8987	3	4		Interactions	
5	0.0971	3	6		Interactions	
5	0.762	5	2		Interactions	
5	-0.767	5	4		Interactions	
5	0.3696	5	6		Interactions	
7	136	1	1	0	Easting bound 1	1
7	702	1	2	0	Easting bound 2	1
7	5840	2	1	0	Northing bound 1	2
7	6290	2	2	0	Northing bound 2	2
8	0.095	1	2		`Soft' threshold for +ve values	
9	0.7305				Dispersion parameter	
10	0	1			Observed residual correlation structure	