

# Simulation of extreme rainfall from CMIP5 in the Onkaparinga catchment using a generalized linear model

**M. M. Rashid<sup>a</sup>, S. Beecham<sup>a</sup> and R. Chowdhury<sup>b</sup>**

<sup>a</sup> *School of Natural and Built Environments, University of South Australia, Mawson Lakes, SA 5095, Australia.*

<sup>b</sup> *Department of Civil and Environmental Engineering, United Arab Emirates University, United Arab Emirates*

*Email: [mdmamunur.rashid@mymail.unisa.edu.au](mailto:mdmamunur.rashid@mymail.unisa.edu.au)*

**Abstract:** Due to the changes in global climate, the intensity, frequency and magnitude of heavy rainfall events are changing and this has been documented in many recent studies. Increasing trends in extreme rainfall directly affects infrastructure, agriculture as well as public and ecosystem health. So, projection of changes in extreme rainfall events is useful for policy making associated with climate change adaptation. General Circulation Models (GCMs) are the most important tool for climate change impact studies. But due to their coarse spatial resolution and their inability to capture local rainfall processes, GCMs cannot be used directly in hydrological impact studies. To bridge this gap downscaling has often been applied to transform GCM information to a finer resolution. There are broadly two types downscaling, namely dynamic and statistical methods. The latter is inexpensive and readily implementable compared to dynamic downscaling. Among several statistical downscaling techniques, generalized linear model (GLM) based downscaling techniques incorporate the spatial-temporal structure of rainfall. Because of this, GLMs have been used in several recent studies. However, application of this technique for downscaling of extreme rainfall events is relatively new.

In this study, a GLM based multi-site downscaling technique has been applied using the GLIMCLIM software package for downscaling extreme rainfall in the Onkaparinga catchment at nine rainfall stations. A GLM was fitted to the observed rainfall conditioned to several large scale atmospheric and circulation variables from NCEP reanalysis data for the calibration period 1991 to 2005. This relation was used to simulate the daily rainfall for the validation period (1981 to 1990) using NCEP reanalysis and CSIRO MK3.6 historical data and for the future period 2041 to 2060 using RCP4.5 and RCP8.5 scenarios of CSIRO MK3.6. These daily rainfall series were used to estimate extreme rainfall indices such as consecutive dry days (CDD), rainfall events greater than 10mm/day (R10) and annual maximum daily (AM) rainfall. As far as we are aware, this is the first attempt where a GLM technique has been applied for downscaling extreme rainfall events using large scale data from CMIP5 GCMs at least for the Australian climate. The study reveals that the model performed reasonably well in reproducing the CDD and AM rainfall whereas it underestimated the R10 statistics in most of the months of the year when driven by NCEP reanalysis data. Although the R10 was underestimated, the trend and variability were simulated well. Performance of the model deteriorates when driven by CSIRO MK3.6 historical data. Simulation differences between NCEP reanalysis and MK3.6 can be attributed by the bias in the large scale atmospheric and circulation variables.

AM daily rainfall was reasonably downscaled for NCEP reanalysis data over the period 1981 to 2005, whereas it was overestimated most of the time at all rainfall stations in the simulation driven by MK3.6. AM rainfall magnitudes for different Average Recurrence Intervals (ARIs) when fitted to Log Pearson Type III distribution were significantly larger for the period 2041 to 2060 under both RCP4.5 and RCP8.5 scenarios of MK3.6 compared to the observed data over the period 1981 to 2005. But the reduced accuracy in the simulation run using MK3.6 data may be due to bias in the large scale atmospheric and circulation variables relative to NCEP reanalysis. The study concludes that the GLM can be used to downscale extreme rainfall events. Also non-stationarity in relation to local rainfall and large scale climate variables is considered as a source of uncertainty in climate change impact studies. Most importantly adequate bias correction in the GCM data is essential before any projection is made.

**Keywords:** *Generalized linear model, general circulation model, CMIP5, Extreme rainfall events*

## 1. INTRODUCTION

An increasing trend in the concentration of carbon dioxide and other greenhouse gases is raising the Earth's temperature and changing weather conditions around the world. These global climate changes have intensified the frequency, magnitude and persistence of extreme events of various climate variables, including rainfall, which has been identified in observational records. For example, significant changes in the rainfall percentiles and extreme events have been observed over the period 1910 to 1995 across Australia (Hennessy *et al.*, 1999; Plummer *et al.*, 1999). Suppiah and Hennessy (1998) observed an increasing trend in the 90<sup>th</sup> and 95<sup>th</sup> percentile rainfalls over most regions of Australia. Moreover, according to the IPCC Fourth Assessment Report (AR4), the projection of extreme precipitation frequency increased almost everywhere in the world over the 20th century and the trend will continue in the 21st century (IPCC, 2007). Increased extreme rainfall events can cause failure of existing urban drainage management systems through frequent flooding and erosion. It can also affect the design parameters of different Water Sensitive Urban Design (WSUD) technologies such as bioretention basins and permeable pavements (Beecham and Chowdhury, 2012). Society, economy and the environment are vulnerable to drought and flooding which are caused by changes in the frequency and intensity of extreme rainfall events. This leads to the requirement for assessing changes of such extreme events in the future. Accurate estimation of local changes in extreme rainfall in the future is invaluable for policy making.

To date the most reliable source of climate change information are the General Circulation Models (GCMs) which have very coarse spatial resolution which is not suitable for hydrological impact studies. Moreover, GCM outputs have uncertainty and also have limited or no ability to capture local rainfall processes which results in inaccurate simulation of rainfall. In order to overcome these limitations, downscaling methods have been employed for transforming this coarse information to a finer spatial resolution. In statistical downscaling, the relationship between local rainfall (predictand) and large scale variables (predictor) have been developed and this is then applied to equivalent predictors from climate model data.

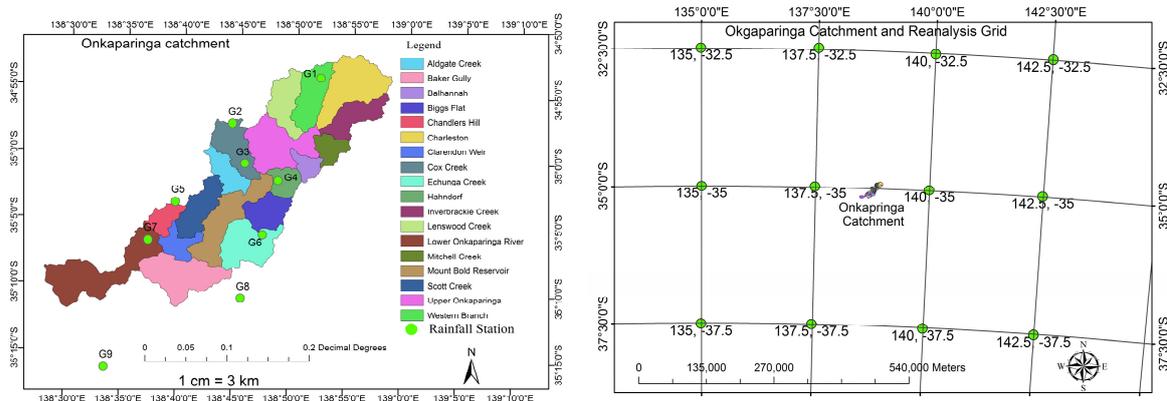
Extreme rainfall has a non-normal distribution with discontinuous events which results in difficulties in simulation. The Generalized Linear Model (GLM) is capable of handling non-normal distributions. Chandler and Wheater (2002) and Yang *et al.* (2005) used a GLM for spatio-temporal stochastic simulation of daily rainfall. Although GLMs have been used for downscaling of rainfall most previous studies have focused on reproduction of the mean rainfall. Very few studies have looked at the downscaling of extreme rainfall events using a GLM technique. Hu *et al.* (2012) and Liu *et al.* (2012) used the GLIMCLIM (Generalized Linear Model for daily CLIMate) software package to fit a GLM to the daily rainfall and to downscale extreme rainfall indices such as dry and wet spell lengths for present and future climates. Yang *et al.* (2005) fitted a generalized extreme value (GEV) distribution to the observed annual maxima rainfall and compared this with median, 90<sup>th</sup> and 99<sup>th</sup> percentiles of annual maxima generated from a GLM simulated daily rainfall time series. They observed that the GLM produced a reasonable simulation of extreme rainfall events.

In statistical downscaling predictors from GCM data are used for projecting future rainfall. The Coupled Model Comparison Project Phase 5 (CMIP5) is a new generation of GCMs that have recently become available to the scientific community. Compared to earlier GCMs, CMIP5 has incorporated additional components that represent the atmosphere's interaction, land use and vegetation, atmospheric chemistry, aerosols and the carbon cycle (Taylor *et al.*, 2011). Four emission scenarios have been developed for future climates in CMIP5 and these are termed Representative Concentration Pathways (RCPs). RCP4.5 represents a medium mitigation scenario whereas RCP8.5 is a high emission scenario. In this study the extreme rainfall has been downscaled using a GLM technique from NCEP reanalysis data. MK3.6 historical data were used to assess whether there is any bias in the downscaled extreme rainfall compared to the simulation by NCEP reanalysis data. Furthermore, future changes in annual maximum daily (AM) rainfall were observed for RCP4.5 and RCP8.5 scenarios of the MK3.6 model. As far as we are aware, this is the first attempt where a GLM technique has been applied for simulating extreme rainfall events from CMIP5 data at least for the Australian climate.

## 2. STUDY AREA AND DATA

The Onkaparinga catchment in South Australia is the study area for this research. This is the main study catchment for the Goyder Institute's Climate Projections project. It has a total catchment area of 560 km<sup>2</sup>. Annual rainfall varies with a strong gradient from approximately 400 mm near the coast to 1170 mm in upstream areas (Teoh, 2003). Daily rainfall data from nine rainfall stations (Figure 1) were collected from the Bureau of Meteorology (BOM), Australia for the period 1980 to 2005. The NCEP/NCAR reanalysis datasets (termed as NCEP hereafter) (Kalnay *et al.*, 1996) were used as predictors for the calibration and validation of the model. Moreover, predictor data from the CSIRO\_MK3.6 GCM (denoted as MK3.6 hereafter) were used

to compare the historical rainfall against downscaled rainfall based on GCM predictor values (historical) for the period 1981 to 2005. RCP4.5 and RCP8.5 scenarios data of MK3.6 were used as predictors for the projection of rainfall for the period 2041 to 2060.



**Figure 1.** Onkaparinga catchment with selected rainfall stations (left) and the reanalysis grid (2.5° X 2.5°) around the study area (right)

### 3. METHODOLOGY

A GLM based downscaling technique was applied using the GLIMCLIM software package (Chandler, 2002) in the Onkaparinga catchment to downscale the multi-site daily rainfall. Details of this downscaling approach is available at Beecham *et al.* (2013); Hu *et al.* (2012) and Liu *et al.* (2012). In addition to seasonality, spatial variation and temporal autocorrelation, large scale atmospheric and circulation variables (air temperature at 700 hPa, relative humidity at 500, 700 and 850 hPa, zonal wind components at 700 and 850 hPa, meridional wind component at 500 and 700 hPa) were obtained from NCEP reanalysis and MK3.6 were used as predictors in the downscaling model. These exogenous daily data were extracted at twelve NCEP reanalysis grid (2.5° X 2.5°) points and averaged, as shown in Figure 1. The exogenous data were bias corrected by subtracting the mean and divided by the standard deviation of the daily NCEP values over the period 1980 to 2005.

The model was run by for NCEP reanalysis and MK3.6 historical data to simulate the present climate over the calibration (1991 to 2005) and validation (1981 to 1990) periods. Moreover, future daily rainfall series were simulated for RCP4.5 and RCP8.5 scenarios of MK3.6. For each simulation, 1000 set of synthetic daily rainfall series were stochastically generated. Extreme rainfall indices such as consecutive dry days (CDD), rainfall events more than 10mm/day (R10) (10 mm/day is the more or less 90<sup>th</sup> percentile of daily rainfall at all the rainfall stations) and annual maximum daily (AM) rainfall were estimated from the simulated daily rainfall series. A wet day was defined as a day with rainfall greater than 0.5mm. Model efficiency statistics such as coefficient of determination ( $R^2$ ) and mean bias (MB) were used to assess the agreement between the observed and median simulated indices. The 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of simulated values were used for defining a 95% confidence bound. AM series generated from different simulations were fitted to a Log Pearson Type III distribution to estimate the magnitude of AM rainfall for different annual recurrence intervals (ARIs).

### 2. RESULT AND DISCUSSION

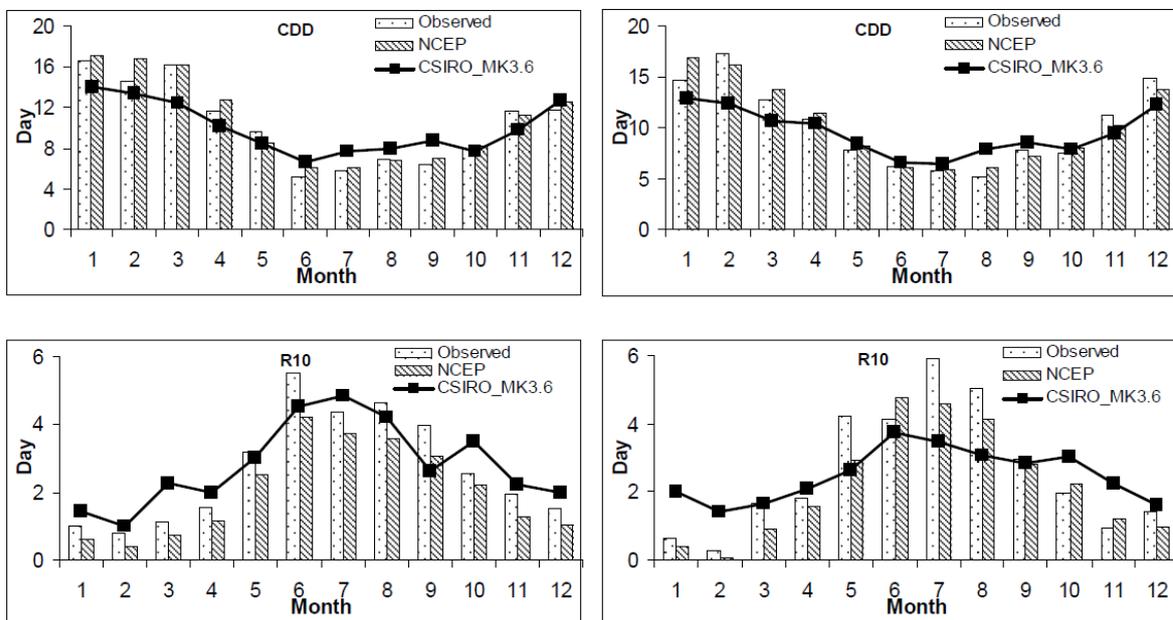
The performance statistics of extreme rainfall indices such as CDD and R10 driven by NCEP reanalysis and MK3.6 historical datasets for the calibration and validation periods are listed in Table 1. Performance statistics were estimated for each rainfall station separately. Table 1 shows the range (low to high) of the performance statistics over the study area. Significant variation was observed in the model performance among the rainfall stations. For example, in the calibration period,  $R^2$  varies from 0.45 to 0.59 and MB varies from -0.24 to 0.96 for CDD using NCEP data. The performance of the model driven by NCEP data was consistent over the calibration and validation periods in term of  $R^2$  and MB for both indices CDD and R10. It should be noted that the simulation by MK3.6 data showed less performance in reproducing the extreme rainfall indices as compared to the simulation driven by NCEP reanalysis data. This indicates the existence of biases in the NCEP and MK3.6 data.

Figure 2 shows the mean observed and median simulated CDD and R10 values for different months of the year. The index statistics were estimated for all nine stations separately and were then averaged. The model was able to reasonably simulate the CDD statistics for the calibration and validation periods when NCEP reanalysis datasets were used. In contrast CSIRO\_MK3.6 showed less satisfactory results. CDD was underestimated for January to April and overestimated for July to September. The study reveals that MK3.6 has a tendency to underestimate CDD in the dry season and overestimates CDD in the wet season both in the calibration and validation periods. In particular it appears that the model produces too many rainy days in the dry season.

**Table 1.** Model efficiency statistics for simulating extreme rainfall indices driven by NCEP reanalysis and CSIRO\_MK3.6 historical datasets over the calibration and validation periods.

Extreme rainfall indices	Periods	Model	Efficiency statistics	
			R <sup>2</sup>	Mean bias (MB)
CDD	Calibration	NCEP	0.45 - 0.59	(-0.24) - 0.96
		MK3.6	0.27 - 0.37	(-1.36) - 0.08
	Validation	NCEP	0.48 - 0.56	(-0.68) - 0.96
		MK3.6	0.30 - 0.42	(-1.77) - 0.20
R10	Calibration	NCEP	0.44 - 0.67	(-1.19) - 0.28
		MK3.6	0.23 - 0.34	(-0.71) - 0.20
	Validation	NCEP	0.48 - 0.67	(-0.63) - 0.02
		MK3.6	0.40 - 0.56	(-0.03) - 0.60

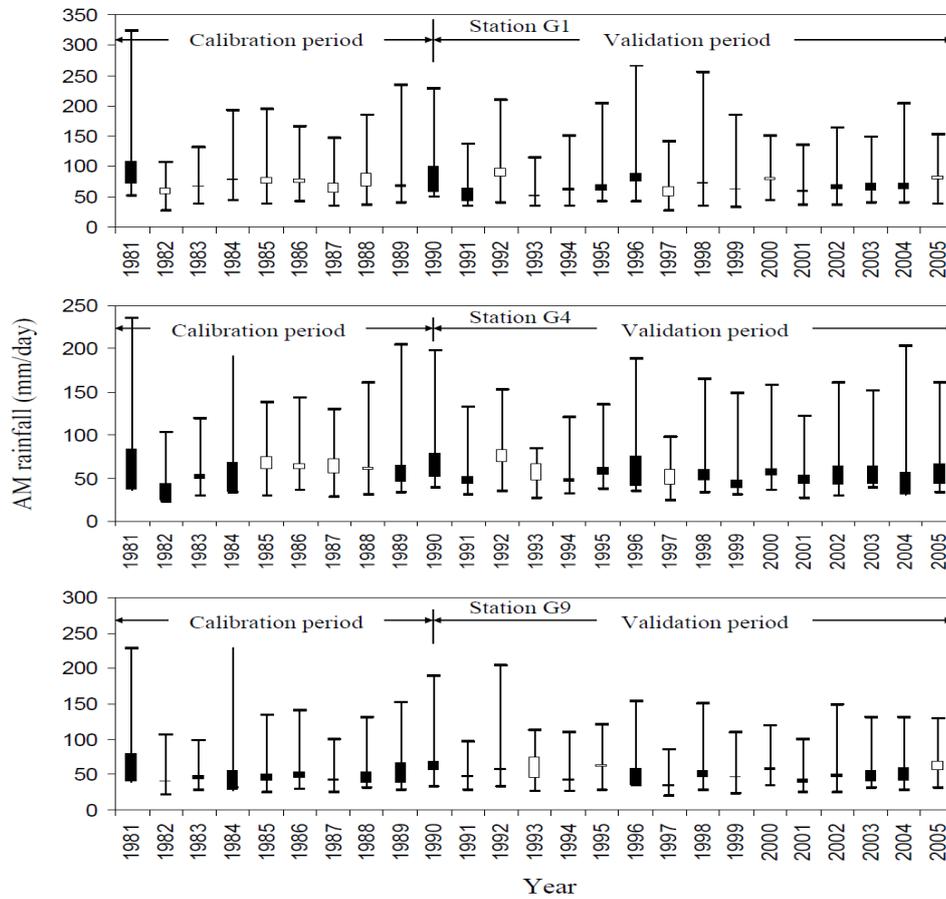
On the other hand the model performance was not satisfactory for simulation of R10 compared to CDD. The model continuously underestimated the R10 statistics for each month in the calibration period for the NCEP reanalysis data. Although the R10 was underestimated, the trend and variability were reasonably well simulated. In the simulation using MK3.6 over the validation period, R10 was overestimated for January, February, October and November whereas it was underestimated for June, July and August. Simulation differences between NCEP reanalysis and MK3.6 can be attributed to the bias in the large scale atmospheric and circulation variables. These biases need to be considered carefully before applying these data for future projections. Otherwise, the projections could be misleading. Due to the short calibration and validation periods, the model might not accurately capture some unusual and extreme storms which may also result in some discrepancies in the simulation result.



**Figure 2.** Comparison of observed and median simulated extreme rainfall indices driven by NCEP reanalysis and CSIRO MK3.6 (CMIP5) datasets for the calibration (left) and validation (right) periods.

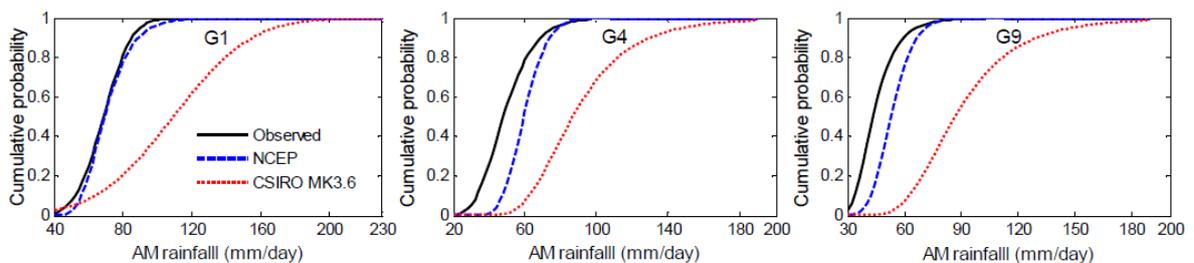
AM rainfall was estimated from the simulated daily rainfall driven by NCEP reanalysis and MK3.5 data for all rainfall stations. Figure 3 shows the comparison of observed and downscaled AM rainfall over the period 1981 to 2005 at rainfall stations G1, G4 and G9. Due to the space limitation only three rainfall stations are presented here which represent the upper, middle and downstream (near the coast) regions of the

Onkaparinga catchment (Figure 1). Figure 3 represents the 2.5<sup>th</sup>, 97.5<sup>th</sup> percentile and median simulated AM rainfall for the NCEP reanalysis data. Simulation of AM rainfall by the model driven by NCEP reanalysis data considerably varies from one station to another. For example, the model performed better at rainfall stations G1 and G9 compared to G4.



**Figure 3.** Comparison of AM rainfall between observed and downscaled by NCEP data for 1981 to 2005. Top and bottom whisker represent the 97.5<sup>th</sup> and 2.5<sup>th</sup> percentile of simulated AM rainfall by NCEP data. Box represent the difference between the observed and median simulated AM rainfall by NCEP data. Black and white box indicate the overestimation and underestimation of AM rainfall by the model, respectively

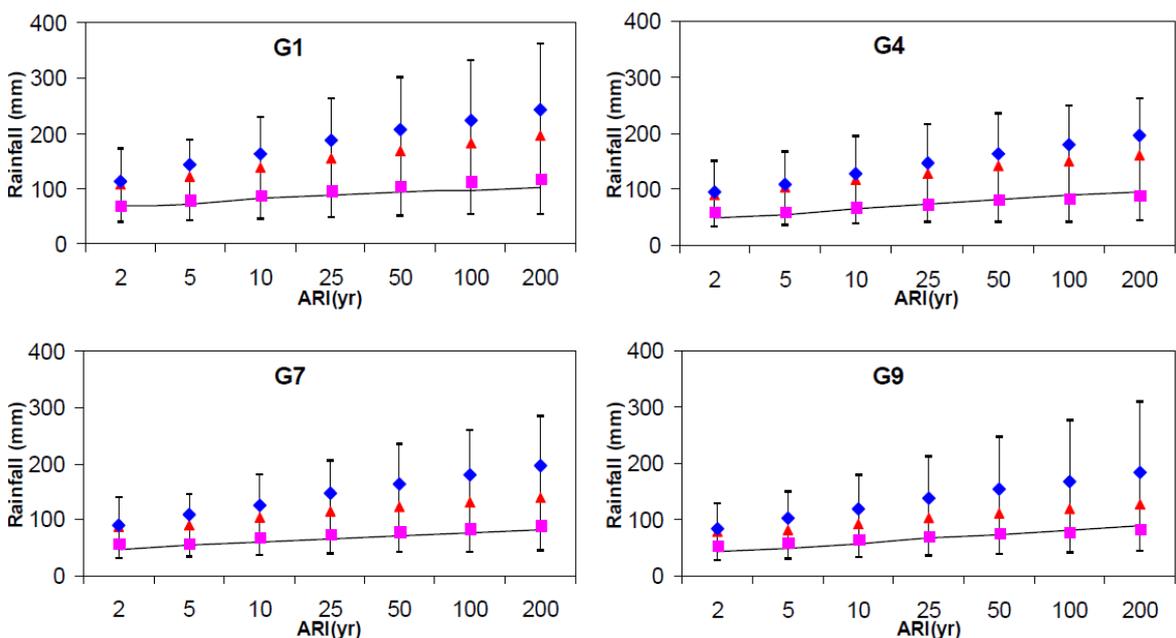
AM rainfall modeled by MK3.6 data were overestimated for most of the year at all stations over the period 1981 to 2005 as shown in Figure 4. In order to evaluate the performance of the simulation by NCEP and MK3.6 data, the average difference of AM rainfall from the observed AM values were estimated for each station for the period 1981 to 2005 (not shown here). Both NCEP and MK3.6 model overestimated AM rainfall for all stations. However, the NCEP model performed better compared to the MK3.6 model. The latter model gave poorer simulations for rainfall station G2 with an average difference of 64.21 mm.



**Figure 4.** Comparison of cumulative distribution of annual maximum (AM) rainfall over the period 1981-2005

A Log Pearson Type III distribution was fitted to four sets of AM series (observed, NCEP-simulated, MK3.6 (RCP4.5)-simulated and MK3.6 (RCP8.5)-simulated) and rainfall magnitudes corresponding to seven ARIs were estimated and compared for the period 1981 to 2005, as shown in Figure 5. It was observed in the study that the AM rainfall for different ARIs was reasonably well simulated by the model driven by NCEP data for all rainfall stations. Rainfall frequency analysis based on the downscaled data obtained from the RCP4.5 and RCP8.5 simulations for 2041 to 2060 are presented by the triangle and diamond symbols, respectively, in Figure 5. Future projections of AM rainfall magnitude for the RCP4.5 and RCP8.5 models show significant increases compared to the observed AM values over the period 1981 to 2005 for all ARIs. The difference between the projected magnitudes of rainfall by the two models was lower at the lower ARIs and larger at the larger ARIs.

Although the future projections of AM rainfall show a significant increase during the period 2041 to 2060 compared to the base period (1981 to 2005) for the simulation driven by both RCP4.5 and RCP8.5 scenarios of MK3.6, the result is not certain due to the biases in the simulated AM series by MK3.6 during the calibration and validation periods. Moreover, whether the relationship between the local rainfall and large scale variables over the present climate will remain unchanged also introduces uncertainty in the projection confidence. However, the calibration of the model over a relatively large observed period would capture the long term variability in the relationship and increase the confidence in the future projections, at least to some extent. It is noted that the bias removal technique (subtraction of mean followed by division by standard deviation) used in this study showed poor performance for the model run by MK3.6 compared to the model run by NCEP reanalysis. Moreover, both the reanalysis and GCM data have their own errors and biases. So, the accurate projection of climate variables is still a major challenge in climate research. Adequate bias correction before projecting the future changes in climate variables is essential for climate change impact studies.



**Figure 5** Comparison of magnitude of daily rainfall for different ARIs using four AM series (observed, NCEP-simulated, MK3.5 (RCP4.5)-simulated and MK3.6 (RCP8.5)-simulated represented by horizontal line, square, triangle and diamond symbol respectively). The vertically bar extents show the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the NCEP-simulated rainfall magnitude.

### 3. CONCLUSION

This study has compared the performance of the GLM to downscale extreme rainfall in the Onkapinga catchment for simulations using NCEP reanalysis and MK3.6 GCM data. Moreover, it also investigated the future projected changes in the extreme rainfall. Extreme rainfall indices such as CDD, R10 and AM rainfall were downscaled for the present climate (1981 to 2005) from NCEP reanalysis and MK3.6 historical datasets. Future changes of AM rainfall for different ARIs fitted by a Log Pearson Type III distribution were

estimated for the RCP4.5 and RCP8.5 scenarios of MK3.6. The model performed reasonably well for the calibration and validation periods when run with NCEP reanalysis data. In contrast, the performance of the simulation driven by MK3.6 historical data was not satisfactory. This might be due to bias in the large scale atmospheric and circulation variables of MK3.6 relative to the NCEP reanalysis data. Moreover, the influence of non-stationary relationships between local rainfall and large scale variables was also observed in the simulation performance driven by MK3.6 data. Although significant future changes were observed in the AM rainfall for different ARIs for both RCP4.5 and RCP8.5 scenarios, its certainty was questionable due to the bias in the MK3.6 data. Bias correction using centering followed by division by standard deviation technique was not sufficient to correct the bias. Adequate bias correction in GCM data is needed before projecting changes in future rainfall.

## REFERENCES

- Beecham, S. and Chowdhury, R. (2012). Effects of Changing Rainfall Pattern on WSUD in Australia. *Journal of Water Management*, 165(5), 285-298
- Beecham, S., Rashid, M. M. and Chowdhury, R. (2013). Statistical downscaling of multi-site daily rainfall in a South Australian catchment using a generalized linear model. *International Journal of Climatology*, Article under review.
- Chandler, R. E. 2002, GLIMCLIM: Generalised Linear Modelling for Daily Climate Time Series (software and user guide). Research Report No. 227, Department of Statistical Science, University College London.
- Chandler, R. E. and Wheeler, H. S. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resource Research*, 38(10), 1-11.
- Hennesy, K. J., Suppiah, A., Forland, E. and Zhai, P. (1999). Australian rainfall changes, 1910-1995. *Australian Meteorology Magazine*, 48, 1-13.
- Hu, Y., Maskey, S. and Uhlenbrook, S. (2012). Downscaling daily precipitation over the Yellow River source region in China: a comparison of three statistical downscaling methods. *Theoretical and applied climatology*, 112(3-4), 447-460.
- IPCC. (2007). Climate Change 2007: The Physical Sciences Basis. Summary for Policymakers, contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *IPCC Secretariat, Geneva, Switzerland*.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437-471.
- Liu, W., Fu, G., Liu, C. and Charles, S. P. (2012). A comparison of three multi-site statistical downscaling models for daily rainfall in the North China Plain. *Theoretical and applied climatology*, 111, 585 - 600.
- Plummer, N., Salinger, M. J., Nicholls, N., Suppiah, R., Hennesy, K. J., Leighton, R. M., Trewin, B. and Lough, J. M. (1999). Twentieth century trends in climate extremes over the Australian region and New Zealand. *Climate Change*, 42, 183-202.
- Suppiah, R. and Hennesy, K. J. (1998). Trend in total rainfall, heavy-rain events and number of dry days in Australia, 1910-1990. *International Journal of Climatology*, 18(10), 1141-1164.
- Taylor, K., Stouffer, R. and Meehl, G. (2011). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485-498. 10.1175/BAMS-D-11-00094.1
- Teoh, K. S. 2003, Estimating the impact of current farm dams development on the surface water resources of the Onkaparinga River Catchment. DWLBC Report 2002/22, Department of Water, Land and Biodiversity Conservation, for the Government of South Australia,
- Yang, C., Chandler, R. E., Isham, V. S. and Wheeler, H. S. (2005). Spatial-temporal rainfall simulation using generalized linear models. *Water Resource Research*, 41(11), 1-13. 10.1029/2004wr003739