Malaysian Journal of Civil Engineering 27 Special Issue (2):311-324 (2015)

PERFORMANCE ASSESSMENT OF DIFFERENT BIAS CORRECTION METHODS IN STATISTICAL DOWNSCALING OF PRECIPITATION

Kamal Ahmed*, Shamsuddin Shahid, Sobri Harun & Nadeem Nawaz

Department of Hydraulics and Hydrology, Faculty of Civil Engineering, University Technology Malaysia

*Corresponding Author: *kamalbaloch2004@gmail.com*

Abstract: Global circulation models (GCMs) are widely used for the modeling and assessing the impacts of climate change. These models do not always accurately simulate climate variables due to the risk of considerable biases. Several bias correction methods have been proposed and applied so far. The selection and application of appropriate bias correction can improve accuracy and reduce uncertainty in downscaled precipitation in arid regions. In this study, initially multilayer perceptron (MLP) neural network was applied to downscale the mean monthly precipitation. The MLP model was calibrated by using National Center for environmental prediction (NCEP) reanalysis dataset and monthly precipitation observations located in selected hyper-arid, arid and semi-arid regions. Later, the performance of four bias correction methods namely, power transformation, simple multiplicative, variance inflation and quantile mapping were evaluated by comparing the mean and standard deviation of observed and downscaled precipitation. It has been found that the power transformation method is the most reliable and suitable method for downscaling precipitation in the arid region.

Keywords: *Downscaling, precipitation, arid, bias correction, multilayer perceptron.*

1.0 Introduction

The increasing concentration of greenhouses gases in the atmosphere is resulting in climate change. The impacts of climate change are observed in most parts of the world. Modeling the impacts of climate change is necessary to avoid major crises in water, food and energy sector in the future. Global circulation models (GCMs) are currently considered as the most important and appropriate tool for the modeling and assessing the impact of climate change (Chu *et al*., 2010, Goyal *et al*., 2012). Generally, GCM simulations are not directly used because of biased representation. This biasness is due to the assumptions made during the development phase of circulation models due to lack of incomplete knowledge about atmospheric phenomena. Therefore, GCM models may not simulate climate variables accurately, and there is always a difference between observed and simulated climate variables. Thus, it is necessary to use bias correction methods to remove biasness from GCM output (Salvi *et al*., 2011). Application of accurate bias correction method can largely remove errors from the model. Several bias correction methods are developed and successfully applied in different studies such as (Ines and Hansen, 2006; Sharma *et al*., 2007; Elshamy *et al*., 2009; Li *et al*., 2010; Gudmundsson *et al*., 2012; Teutschbein and Seibert, 2012b; Acharya *et al*., 2013; Lafon *et al*., 2013; Müller and Thompson, 2013; Wang and Chen, 2013 and Sachindra *et al*., 2014). These methods range from simple scaling approaches to more advanced approaches employing frequency, probability and empirical distributions (Teutschbein and Seibert, 2012a). The selection and application of appropriate bias correction method are challenging in different arid climate classes i.e. hyper-arid, arid and semi-arid. The rainfall in such regions is very scarce, erratic and infrequent. In this study, initially, we used a nonlinear regression-based model i.e. multilayer perceptron neural network for downscaling precipitation without any bias correction method. Multilayer perceptron is the most popular, flexible and simplest type of artificial neural network. This method is convenient for time series data analysis and is used in downscaling by e.g. Wilby and Wigley (1997), Schoof and Pryor (2001), Harpham and Dawson (2006), Cannon (2008), Huth *et al*. (2008), and Hadipour *et al*. (2014). Later, we used four popular bias correction methods which includes power transformation method (Leander *et al*., 2008), simple multiplicative method (Berg *et al*., 2012), variance inflation method (Hessami *et al*., 2008) and quantile mapping (Panofsky and Brier, 1968). The main objective of this study is to assess the performance of these methods in an arid region. The performances of these methods are assessed by comparing the mean and standard deviation of observed and simulated values. This study is carried out in Nokkundi (hyper-arid), Panjgur (arid) and Barkhan (semi-arid) regions located in province Balochistan, Pakistan.

2.0 Study Area and Data

Baluchistan is a mountainous, desert and an arid province, located at 30.12°N, 67.01°E of Pakistan. The map of Baluchistan is shown in Figure 1. Physically, it is an extensive plateau of rough terrain divided into basins by ranges of sufficient heights and ruggedness. Geographically, it is divided in to four distinct zones: upper high lands, lower high lands, plains, and deserts. Topography varies from 5 m to 3700 meters above mean sea level, which can strongly influence the climate. The climate of the province lies in the class of hyper-arid, arid and semi-arid. Monsoon season which usually starts at the end of June and lasts till September and western disturbance which originates from Mediterranean Sea brings rainfall during December to March to these regions. In this study, we selected the region of Nokkundi as hyper-arid, Panjgur as arid and Barkhan as semi-arid. The locations of these regions are highlighted in figure 1.

Nokkundi is located at 28°49′N 62°46′E. The mean annual rainfall recorded at this station is 35 mm. The western disturbance and monsoon contributes 87% and 8% of total rainfall. The months of April, May, June, August, September, October and November are very dry having average rainfall of 1.5 mm. Panjgur is located at 26°58'0N 64°5'60E. The mean annual rainfall recorded at Panjgur station is 102 mm and lies in the arid region. The western disturbances shares 55 % and monsoon shares 20 % of total rainfall. Barkhan is located at 29°54'0N 69°31'0E. This station receives 417 mm rainfall annually and lies in the semi-arid climate class. Monsoon season contributes 65 % of its rainfall. The lowest rainfall is recorded during the months of November to January. The rainfall in these regions is scanty and unevenly distributed.

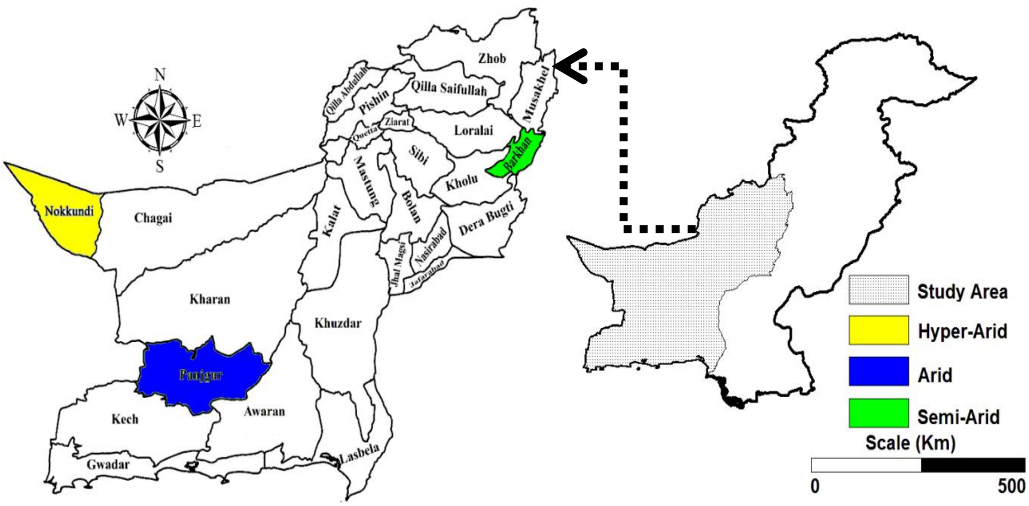


Figure 1: Location of study area in Pakistan Map

The observed data used in this study was recorded at selected stations from 1961 to 2001. The data of NCEP reanalysis global circulation model outputs at selected rainfall recording stations were extracted from Canadian climate change scenarios network (http://www.cccsn.ec.gc.ca). NCEP reanalysis atmospheric data have been used as the model predictors.

3.0 Methodology

The methodology adopted in this study is discussed in this section. Initially, different regions which fall in hyper-arid, arid and semi-arid climate classes were selected. The NCEP reanalysis data of these regions covering sufficient grid points were extracted from Canadian climate change scenarios network website. The potential predictors were selected by the application of Principal component analysis in Statistical Package for the

Social Sciences (SPSS). Principal Component Analysis is a multivariate statistical technique that has been most widely used in climatological studies for extracting potential predictors (Hannachi *et al*., 2007). The selected predictors were used to downscale climate data by using Multilayer perceptron neural network in SPSS environment.

3.1 Multilayer Perceptron (MLP) Neural Network

Multilayer perceptron's is most popular, flexible and simplest type of artificial neural network. It is convenient for time series data analysis. This method is used to map nonlinear relationship between predictor and predictand. The main function of neural network is to improve the performance function between the predicted and observed values. MLP is composed of an input layer, any number of hidden layers, and an output layer of neurons. This method consists of nodes grouped into input, hidden and output layers. Single hidden layer is considered enough to approximate continuous functions, but there is no widely accepted rule regarding the number of hidden nodes (Harpham and Dawson, 2006; Hsieh, 2009; Gaitan *et al*., 2013).

Usually statistical techniques assume some assumption regarding data distribution while MLP makes no assumption. MLP can be trained to approximate virtually any smooth, measurable function. It can model highly non-linear functions and can be trained to accurately generalize when presented with new, unseen data. These features of the multilayer perceptron make it an attractive alternative to developing numerical models, and also when choosing between statistical approaches (Gardner and Dorling, 1998). MLP used following equations for modeling precipitation:

$$y_k = F\left(\sum_{j=1}^{h} w_j G(s_i) + b_k\right) \tag{1}$$

whereas F represents linear activation function of the output neuron, b_k is threshold; w_j represents the connection, G is the hyperbolic tangent sigmoid used as activation function for the hidden nodes, and can be expressed as follows:

$$G(s_i) = \frac{e^{s_i} - e^{-s_i}}{e^{s_i} + e^{-s_i}} \tag{2}$$

where s_i the weighted sum of all incoming information and is also referred to as the input signal

$$s_i = \sum_{i=1}^{n} w_i x_i \tag{3}$$

where x_i is the inputs to the network, w_i is the connection weights between nodes of the input and hidden layers.

3.2 Bias Correction

In this study, we used four popular bias correction methods i.e. power transformation method, variance inflation method, simple multiplicative method and quantile mapping. These methods are discussed below:

3.2.1 Power Transformation Method

Linear bias correction methods correct mean but do not consider variance. The precipitation is usually varied spatially and highly nonlinear in nature. Power transformation is a nonlinear method which corrects both mean and variance of precipitation (Shabalova *et al*., 2003; Leander *et al*., 2008). Power transformation can be employed as:

$$P^* = a.P^b \tag{4}$$

where P* is corrected precipitation, P is simulated precipitation, a & b are parameters obtained during calibration period.

3.2.2 Simple Multiplicative Method

This method is linear in nature and is used to correct mean and standard deviation. This method considers mean observed and mean modelled precipitation while correcting modelled precipitation (Berg *et al*., 2012). The precipitation can be corrected by following equation:

$$P_{corrected} = P_{\mathrm{mod}} \times \frac{\overline{P_{obs}}}{\overline{P_{\mathrm{mod}}}} \tag{5}$$

where P_{mod} is mean modeled precipitation and P_{obs} is mean observed precipitation.

3.2.3 Variance Inflation and Bias Correction Method

The modeling error calculated from Equation (6), (7), and (8) in deterministic part of precipitation is added in the downscaled precipitation via multilayer perceptron neural network.

$$e_i = \sqrt{\frac{VIF}{12}} Z_i S_e + b \tag{6}$$

where e_i is modeling error, Z_i is a normally distributed random number, S_e is standard error of estimate and VIF is variance inflation factor and b is bias correction.

The bias correction and variance inflation factor can be calculated by following equations:

$$b = M_{obs} - M_d \tag{7}$$

$$VIF = \frac{12(V_{obs} - V_d)}{S_e^2} \tag{8}$$

where M_{obs} is observed mean, M_d is deterministic mean, V_{obs} is observed variance and V_d is deterministic variance.

3.2.4 Quantile Mapping

Quantile map (QM) was originally developed by Panofsky and Brier (1968). This method relies on cumulative distributive function. The principal of quantile mapping is to adjust the distribution of model output with observed data set. Gudmundsson *et al*. (2012) formulated QM in following relation:

$$x = h(x) \tag{9}$$

where is h (x) transfer function which is used to establish relationship between observed and modelled output. The distribution of data set and parameters were found through EASY FIT program and then the following expression is used:

$$x = F^{-1} F_m(x) \tag{10}$$

where F^{-1} is the quantile function corresponding to observed precipitation, F_m is the cumulative distribution function of modeled precipitation.

3.3 Performance Evaluation

The performances of these methods are evaluated by statistical measures that include monthly mean and standard deviation of observed and modeled precipitation. The evaluation was done during validation period of the downscaled model.

4.0 Results and Discussion

4.1 Downscaling

Multilayer perceptron is applied to downscale NCEP reanalysis data. The potential predictors of each station were extracted from principal component analysis (PCA). The data of each station was trained from 1961 to 1990 and tested during 1991 to 2001. Initially, the performance of the model was not promising in reproducing similar patterns of observed.

The comparisons of mean monthly observed and mean monthly simulated values have large differences not only in term of mean precipitation but also have large variations in standard deviations. The distribution pattern was also found different from the observed pattern. These differences occur due to high variations in mean monthly observed precipitation. Since, the study area was arid, and rainfall is highly erratic and infrequent behavior in arid regions. In order to overcome this problem; we decide to downscale each month separately. For this purpose, we developed some macros in MS Excel, which were used to separate each month from the year for calibration and validation period. The separated data were transferred and run into the model. Comparisons of the results showed that this procedure is effective for downscaling precipitation in the arid region. It is observed that the model has improved the results of mean rainfall and standard deviation. The downscaled results of hyper-arid, arid and semi-arid regions are shown in Figure 2. It can be seen that the mean monthly precipitation are highly correlated with observed values. In only few months, the model has slightly over and underestimated. Overall, the downscaling results were found satisfactory in term of mean rainfall but not in standard deviation. For example, in the case of Barkhan region, the month of June has observed standard deviation of 55.98 while MLP produced 21.19, similar types of differences were also found in other months and stations. However, the distribution pattern was found same as observed. In such cases, the model was not compatible to cover high variation in the mean standard deviation. Therefore, we applied four popular bias correction methods to adjust variations in mean and standard deviation.

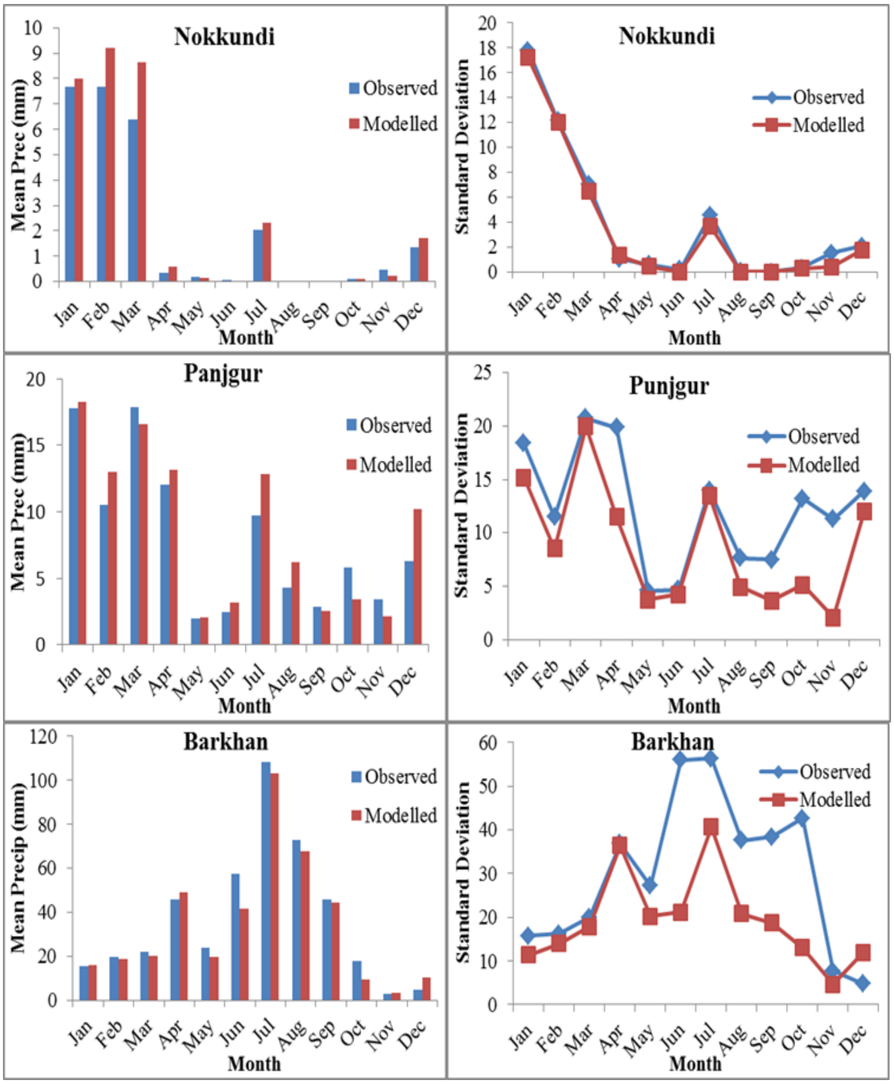


Figure 2: Downscaling results (mean precipitation and standard deviation) of Nokkundi (Hyper-arid), Panjgur (Arid), and Barkhan (Semi-arid) during Validation Period

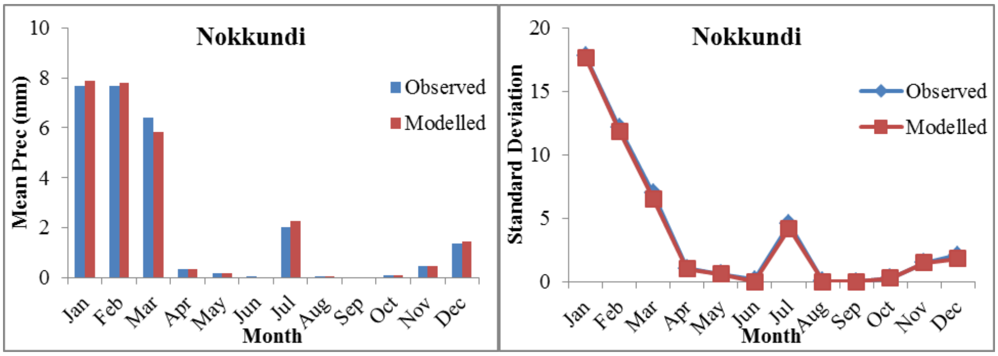
4.2 Application of Bias Correction Methods in Hyper-Arid region

Nokkundi is selected as a hyper–arid region in the present study. Rainfall in this region is scarce, erratic and infrequent in nature. The result obtained from bias correction method i.e. Power transformation method, variance inflation method, simple multiplicative method and quantile mapping method are shown in Table 1. It can be seen from the table that the mean observed and corrected precipitations are very close with each other. The observed mean of time series precipitation is 2.19 mm and the closer mean is achieved by simple multiplicative method which is 2.19 mm.

Table 1: Bias Correction of Nokkundi (Hyper-arid region)

Parameters	Observed	Modelled	Power Transformation	Variance Inflation	Simple Multiplicative	Quantile Mapping
Mean (mm)	2.19	2.58	2.20	2.29	2.18	2.13
Standard Deviation	3.94	3.62	3.79	3.80	3.18	3.58

This accuracy in mean is achieved, since; this method considers both observed and simulated mean precipitation. The simple multiplicative method gives accurate value in term of mean monthly precipitation but cannot adjust time series values. The performance of this method was also found unsatisfactory in adjusting standard deviation. The performances of variance inflation and quantile mapping are also found satisfactory. The performance of power transformation method was found best in all cases. This method was able to adjust mean and standard deviation properly. The best result obtained from power transformation method is shown in figure 3.



(a) (b)

Figure 3: Bias corrected mean (a) & standard deviation (b) from Power Transformation method for Nokkundi (Hyper-arid)

4.3 Application of Bias Correction Methods in Arid region

Panjgur is selected as an arid region in the current study. The results obtained from different bias correction are showed in Table 2. The results shown in table are monthly mean standard deviation of precipitation during validation period (1991-2001).

Table 2: Bias Correction of Panjgur (Arid region)

Parameters	Observed	Modelled	Power Transformation	Variance Inflation	Simple Multiplicative	Quantile Mapping
Mean (mm)	7.93	8.64	7.53	8.18	7.93	7.92
Standard Deviation	12.28	8.71	11.75	11.55	8.45	9.17

It is clear from the table that all methods properly adjusted mean except variance inflation which gives slightly higher mean. The results of power transformation method and variance are also satisfactory in adjusting standard deviations. The simple multiplicative and quantile mapping were not able to adjust standard deviation properly. The overall performance of power transformation was found best among all method in this region. The result obtained from power transformation is showed in Figure 4.

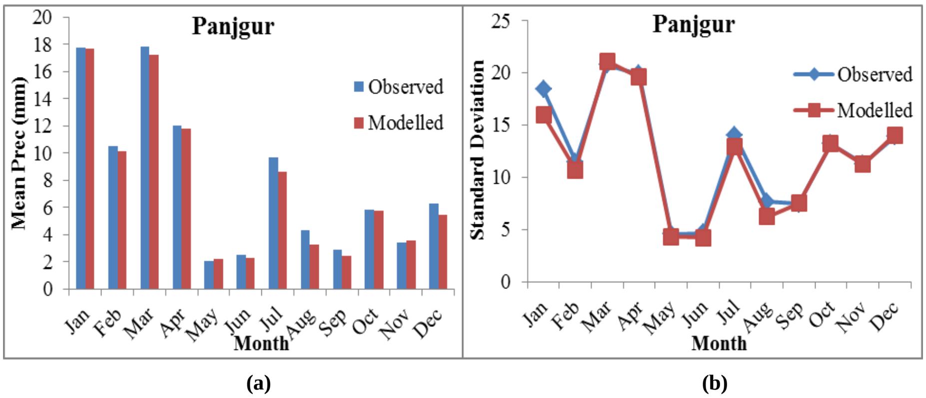


Figure 4: Bias corrected mean (a) & standard deviation (b) from Power Transformation method for Panjgur (Arid)

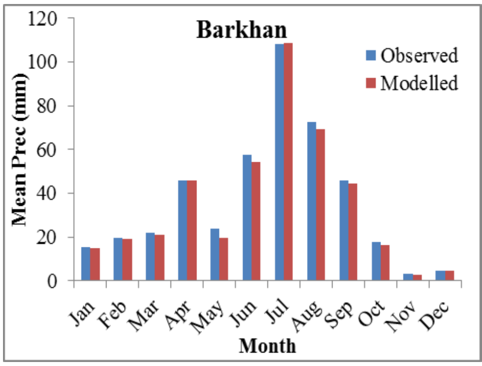
4.4 Application of Bias Correction methods in Semi-arid region

Barkhan is selected as semi-arid region in the present study. The result obtained from bias correction methods are presented in Table 3. The results shown in table are mean and standard deviation of time series precipitation during the validation period. It can be seen from the table that all the methods give more or less same mean as observed. Power transformation and variance inflation method also gives more intimate values to observed standard deviation. Simple multiplicative and quantile mapping method are found unsatisfactory in adjusting the standard deviation. The analysis of mean monthly standard deviation showed an enormous variation during the month of May to September in the observed period. For example, the observed standard deviation in August is found as 37.57 while these methods give variation of 22.41 and 22.27.

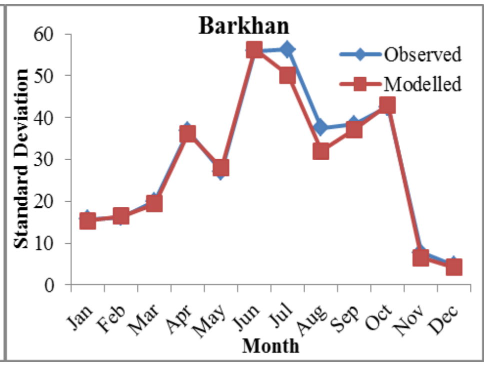
Table 3: Bias Corrections of Barkhan (Semi-arid region)

Parameters	Observed	Modelled	Power Transformation	Variance Inflation	Simple Multiplicative	Quantile Mapping
Mean (mm)	36.44	33.61	35.11	36.46	36.44	36.23
Standard Deviation	29.94	19.22	28.73	30.01	19.55	20.10

It was observed that these methods are incompatible to cover such high variations. The distribution pattern of simulated values is found satisfactory and following the similar pattern of observed values. The overall performance of power transformation and variance inflation method was found satisfactory in all aspects. However, the variance inflation method cannot be considered as a reliable method because this method considers normally distributed random numbers in correcting model errors. Any change in random number can completely change the rainfall trend. Therefore, power transformation method is considered as a reliable method for semi-arid region. The result obtained from power transformation is showed in Figure 5. It can be seen that the model is properly adjusted the mean monthly precipitation by power transformation method, while the month of July, August and September are slightly unadjusted in terms of standard deviation.



(a)



(b)

Figure 5: Bias corrected mean (a) & standard deviation (b) from Power Transformation method for Barkhan (Semi-arid)

5.0 Conclusions

The present study is set out to evaluate the performance of four bias correction methods namely power transformation, simple multiplicative, variance inflation and quantile mapping applied in an arid region. Initially, multilayer perceptron is applied to downscale GCM data without bias correction methods and later four bias correction methods were applied. The performances of these methods were assessed by the comparison of monthly mean precipitation and standard deviation. The result obtained from the comparison showed that the power transformation method is the most reliable and suitable method for removing biasness from GCM model in an arid region and can be used for adjusting the projected precipitation.

6.0 Acknowledgements

The authors would like to acknowledge Ministry of Education Malaysia and Universiti Teknologi Malaysia (UTM) to support this research using ERGS grant number R.J130000.7822.4L084.

References

Acharya, N., Chattopadhyay, S., Mohanty, U. C., Dash, S. K. & Sahoo, L. N. (2013) On the bias correction of general circulation model output for Indian summer monsoon. *Meteorological Applications* 20(3):349-356.

Berg, P., Feldmann, H. & Panitz, H. J. (2012) Bias correction of high resolution regional climate model data. *Journal of Hydrology* 448–449(0):80-92.

Cannon, A. J. (2008) Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli–Gamma Density Network. *Journal of Hydrometeorology* 9(6):1284-1300.

Chu, J. T., Xia, J., Xu, C. Y. & Singh, V. P. (2010) Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theoretical and Applied Climatology* 99(1-2):149-161.

Elshamy, M. E., Seierstad, I. A. & Sorteberg, A. (2009) Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios. *Hydrol. Earth Syst. Sci.* 13(5):551-565.

Gaitan, C., Hsieh, W., Cannon, A. & Gachon, P. (2013) Evaluation of Linear and Non-Linear Downscaling Methods in Terms of Daily Variability and Climate Indices: Surface Temperature in Southern Ontario and Quebec, Canada. *Atmosphere-Ocean*:1-11.

Gardner, M. W. & Dorling, S. R. (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment* 32(14–15):2627-2636.

Goyal, M., Burn, D. & Ojha, C. S. P. (2012) Evaluation of machine learning tools as a statistical downscaling tool: temperatures projections for multi-stations for Thames River Basin, Canada. *Theoretical and Applied Climatology* 108(3-4):519-534.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. & Engen Skaugen, T. (2012) Technical Note: Downscaling RCM precipitation to the station scale using quantile mapping – a comparison of methods. Hydrol. *Earth Syst. Sci. Discuss*. 9(5):6185-6201.

Hadipour, S., Harun, S. & Shahid, S. (2014) Genetic Programming for the Downscaling of Extreme Rainfall Events on the East Coast of Peninsular Malaysia. *Atmosphere* 5(4):914-936.

Hannachi, A., Jolliffe, I. T. & Stephenson, D. B. (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology* 27(9):1119-1152.

Harpham, C. & Dawson, C. W. (2006) The effect of different basis functions on a radial basis function network for time series prediction: A comparative study. *Neurocomputing* 69(16–18):2161-2170.

Hessami, M., Gachon, P., Ouarda, T. B. M. J., Andr,& St-Hilaire (2008) Automated regression-based statistical downscaling tool. *Environ. Model. Softw*. 23(6):813-834.

Hsieh, W. W. (2009) Machine learning methods in the environmental sciences: *Neural networks and kernels*. Cambridge university press.

Huth, R., Kliegrova, S. & Metelka, L. (2008) Non‐ linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? *International Journal of Climatology* 28(4):465-477.

Ines, A. V. & Hansen, J. W. (2006) Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and forest meteorology* 138(1):44-53.

Lafon, T., Dadson, S., Buys, G. & Prudhomme, C. (2013) Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology* 33(6):1367-1381.

Leander, R., Buishand, T. A., Van Den Hurk, B. J. J. M. & De Wit, M. J. M. (2008) Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output. *Journal of Hydrology* 351(3–4):331-343.

Li, H., Sheffield, J. & Wood, E. F. (2010) Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres* 115(D10):D10101.

Müller, M. F. & Thompson, S. E. (2013) Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal. *Advances in Water Resources* 60:121-134.

Panofsky, H. A. & Brier, G. W. (1968) Some Applications of Statistics to Meteorology. Earth and Mineral Sciences Continuing Education, College of Earth and Mineral Sciences.

Sachindra, D. A., Huang, F., Barton, A. & Perera, B. J. C. (2014) Statistical downscaling of general circulation model outputs to precipitation—part 2: bias-correction and future projections. International Journal of Climatology:

Salvi, K., Kannan, S. & Ghosh, S. (2011) Statistical Downscaling and Bias Correction for Projections of Indian Rainfall and Temperature in Climate Change Studies. *In Proceedings of 2011 4th International Conference on Environmental and Computer Science (ICECS 2011).*

Schoof, J. T. & Pryor, S. (2001) Downscaling temperature and precipitation: A comparison of regression‐ based methods and artificial neural networks. *International Journal of Climatology* 21(7):773-790.

Shabalova, M., Van Deursen, W. & Buishand, T. (2003) Assessing future discharge of the river Rhine using regional climate model integrations and a hydrological model. *Climate Research* 23(3):233-246.

Sharma, D., Das Gupta, A. & Babel, M. S. (2007) Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. *Hydrol. Earth Syst. Sci*. 11(4):1373-1390.

Teutschbein, C. & Seibert, J. (2012a) Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology* 456–457(0):12-29.

Teutschbein, C. & Seibert, J. (2012b) Is bias correction of Regional Climate Model (RCM) simulations possible for non-stationary conditions? *Hydrology and Earth System Sciences Discussions* 9(11):12765-12795.

Wang, L. & Chen, W. (2013) A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. *International Journal of Climatology*.

Wilby, R. L. & Wigley, T. M. L. (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography* 21(4):530-548.