Malaysian Journal of Civil Engineering 24(1):1-12 (2012)

APPLICATION OF STATISTICAL DOWNSCALING MODEL FOR LONG LEAD RAINFALL PREDICTION IN KURAU RIVER CATCHMENT OF MALAYSIA

Zulkarnain Hassan[1]* and Sobri Harun[1]

[1]*Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia*

*Corresponding Author: *zulkarnain23@live.utm.my*

Abstract: The climate impact studies in the hydrology are often relying on the climate change information at a fine spatial resolution. However, Global Climate Models (GCMs) which is regarded as the most advanced models yet for estimating the future climate change scenarios are operated on the coarse spatial resolution and not suitable for climate impact studies. Therefore, in this study, the Statistical Downscaling Model (SDSM) was applied to downscale rainfall from the GCMs. The data from single rainfall station located in the Kurau River were used as input of the SDSM model. The study included the calibration and validation with large-scale National Centers for Environmental Prediction (NCEP) reanalysis data, and the projection of future rainfall corresponding to the GCMs-variables (HadCM3 A2). The study results shows that during the calibration and the validation stage, the SDSM model can be well acceptable in regards to its performance in the downscaling of the daily and annual rainfall. For the future period (2010-2099), the SDSM model estimates that there were increases in the total average annual rainfall and generally, the area of rainfall station become wetter.

Keywords: *statistical downscaling, global climate model, rainfall, Kurau river, SDSM*

1.0 Introduction

In recent years, many studies have focusing the effects of Global Climate Models (GCMs) and climate change on rainfall variability in different parts of the world. Even through GCMs can be applied directly for climate change assessment, these models have the largest scale resolution and need to be downscaled into a site-scale (smaller-scale) (Yimer *et al.*, 2009). Therefore, the statistical methods by using the Regression Models are applied in this study. These models generally involve the establishing linear or nonlinear relationships between sub grid-scale (e.g. single-site) parameters and coarser-resolution (grid-scale) predictor variables (Wilby and Wigley, 1997). The relationship can be named as 'transfer function'.

One of the examples of the Statistical Downscaling tools that implement the Regression Models is the Statistical Downscaling Models (SDSM). The SDSM facilities have undergone rapid development of multiple, low cost, single-site scenarios of daily surface weather variables under the present and future climate forcing. As an addition, the tool performs ancillary tasks of data quality control and transformation, predictor variable pre-screening, automatic model calibration, basic diagnostic testing, statistical analyses and graphing of climate data (Wilby *et al.*, 2007). Therefore, the main objective of this study is to downscale current and future rainfall corresponding to GCMs-variables, by applying the SDSM model.

2.0 Material and Methods

2.1 Study Area

The rainfall station chosen for this study was located in the catchment of Kurau River of Perak State of Malaysia (Figure 1).The coordinate of rainfall station is 5^{o} 12'N 100^{o} 41'E. This catchment is selected as a study area because there is a dam located at the downstream of basin. This dam provides irrigation water for double cropping planting intensity to the Kerian-Sg. Manik irrigation scheme. About 10,000 farmers with some 24,000Ha of paddy land are depending on this rice cultivation industry. The dam also provides fresh water to meet the domestic and industrial demands to Kerian District as well as Larut Matang District. Therefore, assessing the impact of future climate change at this catchment will gives an insight upon which appropriate decisions about the water resource development can be made.

2.2 *Global Climate Models (GCMs)*

The GCMs-predictors of Hadley Center's GCM (Hadley Centre 3rd Generation - HadCM3) A2 future scenario is run (named as SRES A2) for 1961–2099. The re-analysis data predictors of National Center of Environmental Prediction (NCEP) on HadCM3 computational grid for 1961–2000 at daily time steps was used in the study. Both GCMs-predictors were obtained from the Canadian Climate Impacts Scenarios (CCIS) website for a daily time step.

HadCM3 is a coupled atmosphere-ocean GCMs developed at the Hadley Centre of the United Kingdom's National Meteorological Service. This GCMs contained a complex model of land surface processes, included 23 land cover classifications; four layers of soil where temperature, freezing, and melting are tracked; and a detailed evapotranspiration function that depends on temperature, vapour pressure, vegetation type, and ambient carbon dioxide concentrations (Mohammed, 2009). HadCM3 was chosen because the model is widely used in many climate-change impact studies. Furthermore, HadCM3 provides daily predictor variables, which can be used for the SDSM model. In addition, HadCM3 has the ability to simulate for a period of thousand years, showing little drift in its surface climate. Its predictions for temperature change are average, and for the precipitation, the predictions' increases are below average

(McCarthy *et al.*, 2001). Among the SRES scenarios, A2 was considered as among the worst case scenarios, projecting high emissions for the future.

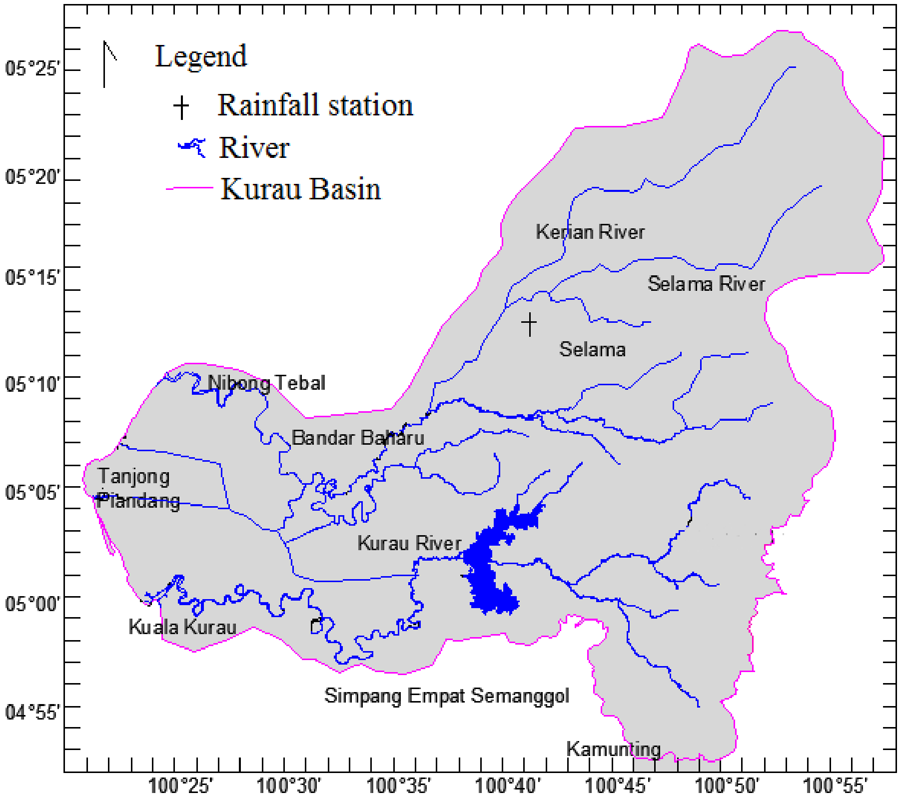

Figure 1: Location of the observed rainfall station

2.3 Statistical Downscaling Model (SDSM)

The SDSM model was introduced by Wilby *et al.* (2002). The method consists of two steps. In the fist step, it determines whether rainfall occurs on each day or not. This is defined as;

$$w_t = \propto_0 + \sum_{j=1}^{n} \propto_j \hat{u}_t^{(j)} + \propto_{t-1} w_{t-1} \tag{1}$$

where, t is time (days), w_t is the conditional possibility of rain occurrence on day t, $\hat{u}_t^{(j)}$ is the normalized predictor, α_j is the regression parameter deduced by an ordinary least square method and w_{t-1} and α_{t-1} are the conditional probabilities of rain occurrence on day *t-1* and lag-1 day regression parameter respectively. These two parameters are optional, depending on the study region and predictand. Uniformly distributed random number r_t ($0 \leq r_t \leq 1$) was used to determine the rain occurrence and supposed the rain would happen if $w_t \leq r_t$.

In the second step, it determines the estimated value of rainfall on each rainy day. It can be defined by a z-score as;

$$Z_t = \beta_0 + \sum_{j=1}^{n} \beta_j \hat{u}_t^{(j)} + \beta_{t-1} + \varepsilon \quad (2)$$

in which Z_t is the z-score on day t, β_j is the calculated regression parameter, and β_{t-1} and Z_{t-1} are the regression parameter and the z-score on day *t-1* respectively. Furthermore, rainfall y_t on day t can be written as;

$$y_t = F^{-1}[\emptyset\,(Z_t)] \quad (3)$$

in which ϕ is the normal cumulative distribution function and F is the empirical function of y_t.

During downscaling with the SDSM, a multiple linear regression between a few selected GCM predictors and rainfall were uutilised. The operation and structure of the SDSM model during calibration can be summarised as shown in Figure 2.

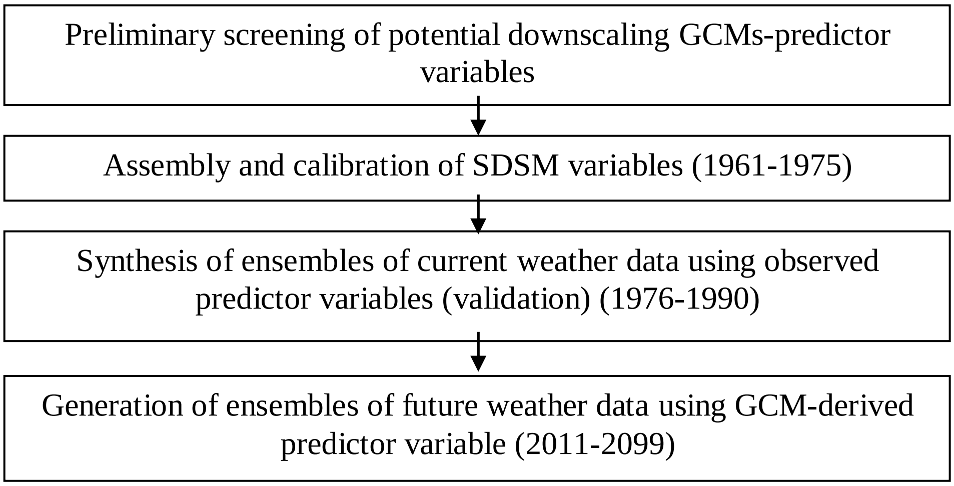

Figure 2: The operation and structure of SDSM

2.3.1 Preliminary Screening of Potential Downscaling Predictor Variables

Several procedures have been suggested to select suitable GCMs-predictor variables, such as partial correlation analysis, step-wise regression, or information criterion. Each statistical analysis can be used to choose a sensible combination of predictors from the available data. However, in SDSM, the task of screening is simple and can be achieved by using linear correlation analysis and scatter plots.

The observed daily data of GCMs-predictor variables representing the current climate condition (1960–2000) derived from the NCEP reanalysis data set was used to investigate the percentage of variance explained by each predictand-predictor pairs. The steps to identify predictor variables that were used in this study being recommended by several researchers (Wilby *et al.*, 2007; Mohammed, 2009; Chu *et al.*, 2009) are; (1) Choose all predictors and run the explained variance on a group of eight or ten of predictors at a time; (2) Of each group, pick a high explained variance of predictor(s); (3) Then, partial correlation analysis is done for selected predictors based on correction of each predictor. There could be a predictor with a high explained variance, but it might be very highly correlated with another predictor. This means that it is difficult to tell that this predictor will add information to the process, and therefore, it will be dropped from the list; (4) Finally, the scatter-plot is used to show the relationship between potential predictor and predictand.

2.3.2 Calibration and Validation

The calibration model process constructs downscaling models based on multiple regression equations, given a daily weather data (the predictand) and a regional-scale atmospheric (predictor) variables (Wilby *et al.*, 2007). It was carried out based on the selected predictor variables that were derived from the NCEP data set. The temporal resolution of the downscaling model for precipitation and temperature downscaling are specified as monthly for each station. Some of the SDSM setup parameters for bias correction and variance inflation were adjusted during calibration to obtain a good statistical agreement between the observed and simulated climate variables. For event threshold, a value of 0.5 was used.

After the model calibration, validation process is needed. Validation process enables to produce synthetic current daily weather data based on inputs of the observed time series' data, and the multiple linear regression parameters produced using independent observe data, which neglected during calibration procedure. In this study, daily rainfall from 1961-1975 was used for calibration and 1967-1990 was then used for validation. The choices of 1961-1990 as the calibration and validation periods was made based on the availability of the rainfall data.

The performances of SDSM on calibrating and validating data were evaluated based on the coefficient of correlation (R), which being defined as;

$$R = \frac{\sum(obs-\overline{obs})(pred-\overline{pred})}{\sqrt{\sum(obs-\overline{obs})^2 \sum(pred-\overline{pred})^2}} \tag{4}$$

where, *obs* = observed stream flow value; *pred* = predicted stream flow value; $\overline{obs}$ = mean streamflow observed value; and $\overline{pred}$ = mean streamflow predicted. R is measure of how well the predicted values from a forecast model fit with the real-life data with a perfect fit gives a coefficient of 1.0.

2.3.3 Downscaling for Future Emission

The regression weighted produced during the calibration process is applied to generate a future daily weather data. The study assumes that relationship between predictor and predictand under the observed conditions (during calibration) remains valid under the future climate conditions. Hundred ensembles of synthetic daily rainfall time series were produced for HadCM3 A2 for a period of 139 years (1961 to 2099). The outcome was averaged and divided by three (3) period of time, which are 2020s (2010-2039), 2050s (2040-2069) and 2080s (2070-2099).

3.0 Results

3.1 Selection of Predictors

Correlation coefficients between selected predictors and daily rainfall are presented in Table 1. This combination has been selected for the SDSM model due to their maximum correlation daily rainfall. This table also reports the partial correlation and *P* value between the predictors and rainfall that help identify the amount of explanatory power for each predictor

3.2 Results of the Calibration and Validation Models

The calibration model process constructs downscaling models based on multiple regression equations, given the observed daily rainfall and NCEP-reanalysis. The model structures of calibration have been categorised as the condition for rainfall and un-condition for temperature. Figure 3 exhibits the calibration result of the SDSM model downscaling (1961-1975) of daily rainfall. It can be seen that the SDSM model shows a good agreement between the observed and simulated mean daily rainfall and variance. However, there was under-estimation for the graph of average dry and wet-spell length for several months.

Table 1: Partial correlation between NCEP-reanalysis with observed rainfall

	Rainfall	
	Partial r	P value
ncepp__fas.dat	-0.059	0.0001
ncepp5_fas.dat	0.05	0.001
ncepshumas.dat	0.04	0.0011
nceptempas.dat	0.045	0.002

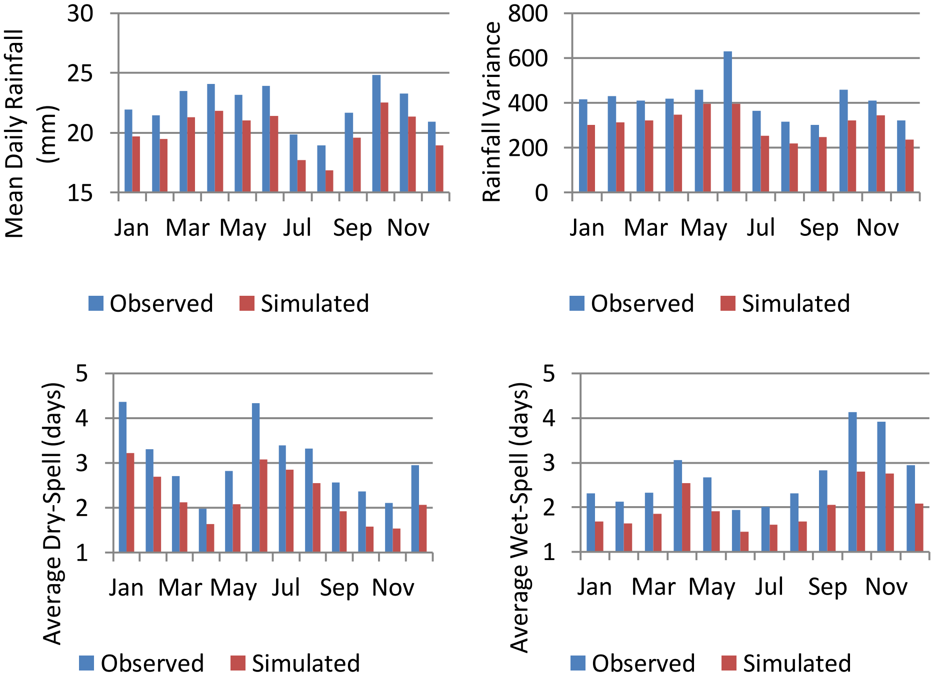

Figure 3: Calibration result of SDSM model downscaling (1961-1975) for daily rainfall

After model calibration, a validation process is required. The validation process produces synthetic current daily weather data based on inputs of the observed time series' data, and the multiple linear regression parameters produced using independent observe data, which were not used during calibration procedure. Figure 4 shows a validation result of the SDSM model downscaling of daily rainfall, and gives a satisfactory agreement between the observed and simulated mean daily rainfall, average dry-spell length and average wet-spell length for all months of the year. However, in the graph of rainfall variance, there was an underestimate simulation for several months. It is very difficult to improve this result since any attempt to improve quality of variance from the model output will automatically affects the other variables which are originally well calibrated. Therefore, the result of rainfall variance was accepted in this case.

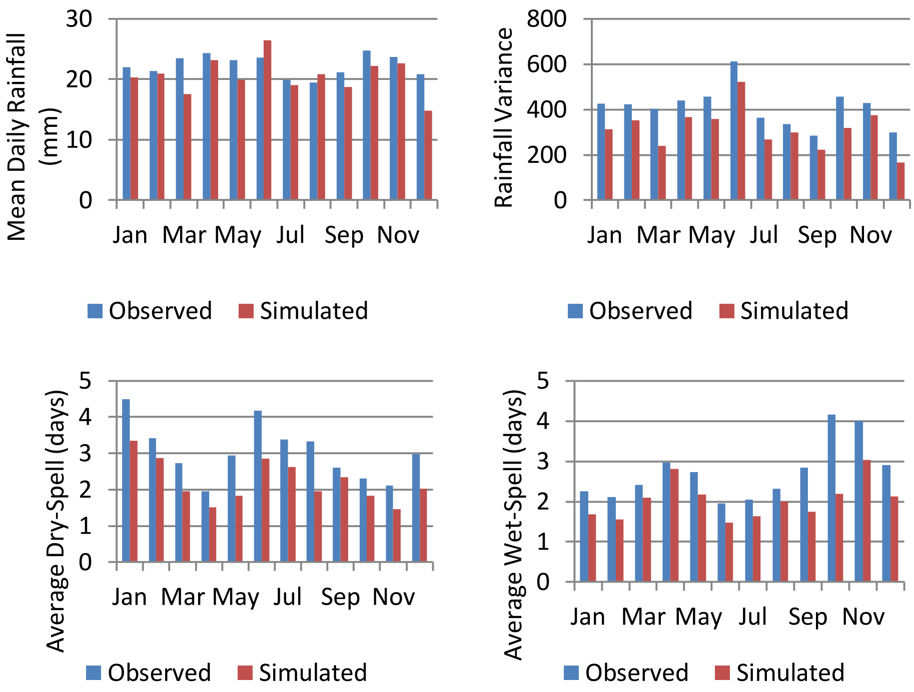

Figure 4: Validation result of SDSM model downscaling (1976-1990) of daily rainfall

The result for calibration and validation are shown in Table 2. The results show the acceptable response of the SDSM model between the observed and downscaling predictands such as rainfall corresponding on the NCEP predictors. For a downscaled of daily rainfall, the *R* do not show a close relationship with observed value. It can be seen that the SDSM model unables to predict well for daily rainfall when $R<0.3$ during calibration and validation. However, the downscaled of monthly and annual rainfall give a close relationship with observed rainfall values ($R>0.4$). Therefore, the SDSM performance indicates the ability of selected NCEP predictors in generating the predictands for the HadCM3 A2 scenarios of the studied site.

Table 2: The SDSM performance using coefficient of correlation (*R*) for downscaling rainfall using NCEP predictor and observed predictands

	Calibration (1961-1975)	Validation (1976-1990)
Rainfall		
Daily	0.26	0.17
Monthly	0.67	0.50
Annual	0.59	0.43

3.3 Downscaling for Future Emission

The result of downscaling for future emission of SDSM for daily rainfall is shown in Figure 5. This figure exhibits that the mean daily rainfall does not show a constant increase or decrease in rainfall trend. From January to February, an increase of daily rainfall is noticeable. Meanwhile, in March, a decrease of rainfall and in April, a marginal increment of rainfall intensity are clearly exhibited. There is an increment in June until August, with the highest increment of rainfall recorded in June, which gives reading of 48mm in 2080s. From September to October, the rainfall decreases constantly, before return to increament patten in November.

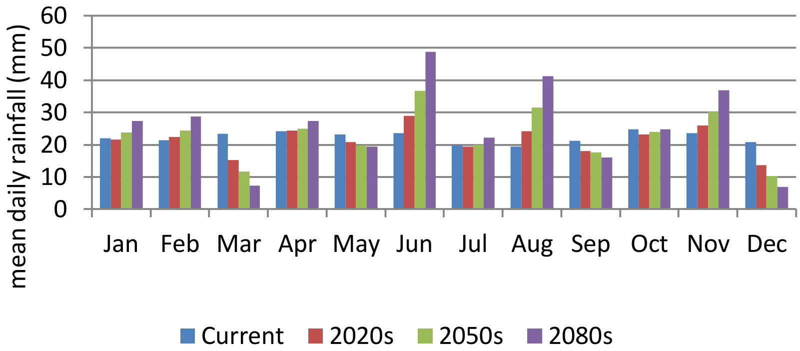

Figure 5: General trend of mean daily precipitation and temperature corresponding to a climate change scenario downscaled with the SDSM

For annual rainfall corresponding to future emission (Table 3), the result shows that there is an increase in trend of annual rainfall. In the 2020s, the simulated annual rainfall is 3828.03mm, and increase by 233.07mm than the present annual rainfall. In the 2050s, there is increases of annual rainfall of 4138.38mm and the increases is measured at 543.42mm than the present annual rainfall. The increase trend is also found in the 2080s with 4892.80mm and the increases given by 1297.84mm than the present annual rainfall.

Table 3: Annual average values for present and simulated rainfall, maximum temperature and minimum temperature corresponding toA2 scenario

	Annual Average Values
Rainfall (mm)	
Present	3594.96
A2 Scenario	
2020s	3828.03
2050s	4138.38
2080s	4892.80

4.0 Discussion

The main fundamental of the SDSM model is the relationships between GCMs-predictor (HadCM3 A2) and predictand (rainfall). Therefore, the selections of GCMs variables are most important parts in the climate change study, and will affect the results of climate assessment. Many studies such as Wilby and Wigly (2000) and Hashmi *et al.* (2009) have applied more than five (5) GCMs-variables in their SDSM analysis. These numbers of GCMs-variables have been chosen in order to show a real condition of climate change in the future. However, the selections of GCMs-variables are difficult and tricky. In addition, the selections of GCMs-variables are still in uncertainty on their methods and there is no standard rule for it. Therefore, in this study, only four (4) of GCMs-variables have been applied. The selections have been tested based on the higher correlation between GCMs-variables and rainfall. Furthermore, the selections of GCMs variables have been looked through by the good performance of the SDSM model during the calibration and validation periods. These results have a similar pattern with the patterns from Karamouz *et al.* (2009), which used only three (3) of GCMs-variables.

In general, the study showed that the SDSM model was able to capture most of the monthly and annual rainfall with a good agreement between observed and simulated rainfall. However, the study found that the model was poor in predicting on a daily rainfall. The results were similar with other studies such as Dibike and Coulibaly (2005), Harun *et al.* (2008), Karamouz *et al.* (2009), and Chu *et al.* (2009). Hence, results from this study is considered fully justified according to early research works. Furthermore, a daily rainfall is the most difficult variables for prediction and it is a condition process which involves an inter-connected with many factors/variables.

The future (2011-2099) annual rainfall results showed that there were an increase in the rainfall’s intensity. For daily rainfall, the results showed that a larger increase of daily rainfall in Jun and August, and a larger decrease in March and December.

5.0 Conclusion

The Statistical Downscaling tools are a tool to downscale the GCMs into a fine scale hydrological region. The SDSM model has been used for this purpose and being widely applied across the world, but not much usage in Malaysia. This study proves that the SDSM model has the ability to perform well during calibration and validation. The study results show that the Kurau River basin will become wetter in future. The results of downscaling of daily rainfall for future emission show an increment trend for each month, except for March, September and December. The increment of annual rainfall has also been observed in this study.

Acknowledgments

The authors would like to thank Department of Irrigation and Drainage Malaysia (DID) and Malaysian Meteorological Department (MMD) for providing data and technical support. The author would also like to thanks to Dawson C.W (developer of SDSM 4.2) for his valuable support and quick feedback through e-mail contacts. The project has been financial support by EScience, under vote 79385 and UTM.

References

Chu, J., Xia, J., Xu, C. Y., and Singh, V. (2009) Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theoretical and Applied Climatology*, 99(1): 149-161.

Dibike, Y. B. and Coulibaly, P. (2005) Hydrologic impact of climate change in the Saguenay watershed: Comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, 307(1-4): 145-163.

Hashmi, M. Z., Shamseldin, A. Y., and Melville, B. W. (2009). Statistical downscaling of precipitation: State of the art and application of Bayesian Multi-model approach for uncertainty assessment. *Hydrology and Earth System Sciences*, 6: 6535–6579.

Harun, S., Hanapi, M. N., Shamsuddin, S., Mohd Amin, M. Z., and Ismail, N. A. (2008) *Regional Climate Scenarios Using a Statistical Downscaling Approach*. Technical Report, Universiti Teknologi Malaysia, Skudai. 1-91pp.

Karamouz, M., Fallahi, M., Nazif, S., and Rahimi Farahani, M. (2009) Long lead rainfall prediction using Statistical Downscaling and Artificial Neural Network modeling. *Scientia Iranica*, 16(2): 165-172.

McCarthy, J., Canziani, O., Leary, N., Dokken, D., and White, K. (2001) *Climate Change 2001: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, New York. 105-110pp.

Mohammed, Y. (2009). *Climate Change Impact Assessment on Soil Water Availablity and Crop Yield in Anjeni Watershed Blue Nile Basin*. Master Thesis, Arba Minch University, Ethiopia. 1-123pp.

Ismail, N. A. (2011) *Statistical Downscaling Downscaling for Climate Change Impact Assesment on Hydrology*. PhD. Thesis, Universiti Teknologi Malaysia, Skudai. 1-350pp.

Wilby, R. L. and Wigley, T. M. L. (1997) Downscaling General Circulation Model output: A review of methods and limitations. *Progress in Physical Geography*, 21: 530–548.

Wilby, R. L., Dawson, C. W., and Barrow, E. M. (2002) SDSM — A decision support tool for the assessment of regional climate change impacts. *Environmental Modelling and Software*, 17: 147–159.

Wilby, R. L., Dawson, C. W., and Barrow, E. M. (2007) SDSM 4.2 - A decision support tool for the assessment of regional climate change impacts. User Manual. *Environment (2007), 17 (2)*: 1-94.

Yimer, G., Jonoski, A. and Griensven, A. V. (2009). Hydrological Response of a Catchment to Climate Change in the Upper Beles River Basin, Upper Blue Nile, Ethiopia. *Nile Basin Water Engineering Scientific Magazine: Special Issue on "Climate and Water*, 2: 49-59.