Application of statistical downscaling model (SDSM) for long term prediction of rainfall in Sarawak, Malaysia

M. Hussain[1,2], K. W. Yusof[2], M. R. Mustafa[2] & N. R. Afshar[3]
[1]Hydro Department, Sarawak Energy Berhad, Malaysia
[2]Department of Civil Engineering, Universiti Teknologi PETRONAS, Malaysia
[3]Department of Civil Engineering, Universiti Malaysia Sarawak, Malaysia

Abstract

Long-term prediction of rainfall over a catchment is a challenge for hydrologists. It is required for water resources management, hydropower energy forecasting and flood risks assessment in river basins. Several large scale climate phenomena affect the occurrence of rainfall around the world i.e El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) are most famous for their effect on India, North and South America and Australia. This study is motivated to evaluate the performance of Statistical Downscaling Model (SDSM) developed by annual and monthly sub models for rainfall downscaling from Global Climate Models (GCMs) over the two districts in Sarawak. It is noted that the monthly sub-models have better performance over the annual sub-models. However, both monthly and annual sub-models have poor correlation with the recorded rainfall for the calibration and validation period. Results indicate that both stations show increasing trend in the future annual rainfall under H3A2 and H3B2 scenarios of HadCM3. SDSM predict that the annual rainfall at Belaga and Limbang is expected to increase by 37.8% and 22.7% respectively by 2074. Overall the SDSM approximates the average rainfall very well during the calibration and validation period but the correlation between observed and forecasted rainfall was not so good. And there is a need to improve the statistical downscaling modelling to develop better correlation between predictand and predictors to have better model performance over the wet regions like Sarawak.

1 Introduction

Accurate river flow forecasting is one of the most frequently analysed problems in hydrology as it is driven by the rainfall forecasting. Due to this reason, different rainfall prediction models coming from several areas of knowledge under different approaches and with specific goals have been developed resulting various levels of success. Climate change and anthropogenic effects are widely regarded as the two main drivers of stream flow change (Piao *et al.* [1]). Climate change, leading to and changes in intensities and patterns of rainfall, as well as changes in evapotranspiration, has a significant impact on regional hydrological processes (Labat *et al.* [2]). The increasing concentration of greenhouse gases in the atmosphere due to human activities such as land use changes and the dependence upon fossil fuels has resulted in global warming and a global energy imbalance (Wentz *et al.* [3]). The prediction of the forthcoming climate change at regional scales is vital for climate change adaptation issues. The climate impact studies are based on scenarios that reflect different social balances between the world and local growth, the financial and environmental values. Several climate scenarios established by the Special Report on Emissions Scenarios – SRES (IPCC SRES, 2000) are the research basis of international climate-change projects.

   The study of rainfall forecasting is very important to a country such as Malaysia where the rain is abundant. Malaysia receives rainfall from 2000 mm to 5000 mm annually where it is greatly influenced by two monsoon periods in November to March and May to September. Though having abundant rainfall can have its advantages but it also has its disadvantages as too much of rainfall is often associated with flood events that can cause casualties. The state of Sarawak is not an exception, Sarawak is well known for its long and wide rivers. Rivers have been the main transportation and still are in certain remote areas. Numerous activities can always be found in the vicinity of the rivers, from commercial, industrial to residential. These activities have started since decades ago and still continue to grow and spatially expanding through times providing incomes ranging from small farmers to the largest corporations. Unfortunately, these areas are expected to experience frequent flood events due to the combine effects of the increase in the mean sea level which contribute to backwater in rivers near coastal areas and rainfall intensity.

   To date, the main tools to predict the variability and changes in climate variables i.e temperature, rainfall and humidity on global and continental levels, are Global Climate Models that are also called General Circulation Models (GCMs). These advanced and numerical-based coupled models interpret global systems such as sea-ice, the oceans, and atmosphere (Fowler *et al.* [4]). Although these models are very helpful in the investigation and predictions regarding future changes in climate, the outputs of these models are based on a large grid scale (250 to 600 km). Because of their coarse resolution, the outputs cannot be used

successfully to investigate the environmental and hydrological impacts of climate change on a regional scale (Wilby *et al.* [5]).

In practice, the choice of downscaling method not only hinges on the time, data and technical resources available, but also the intended application (Wilby *et al.* [6]). Moreover, the value of a downscaling method can be reflected only after it is put in a practical impact study to inform climate risk assessment and adaptation options appraisal.

There are two main approaches called statistical downscaling (SD) and dynamical downscaling (DD) for downscaling outputs of a GCM. SD methods are much simpler than DD methods to downscale the outputs of a GCM. Using SD methods, global-scale climate variables such as mean sea level pressure, zonal wind, temperature, geo-potential height, etc. are linked with local-scale variables (regional-scale variables) such as observed temperature, precipitation and humidity, and this is done by producing some statistical/empirical relationships (Wetterhall *et al.* [7]). To date, many statistical models have been developed and are available. SDSM is being used widely throughout the world (Huang *et al.* [8]) to downscale the most important climate variables such as temperature, precipitation, and evaporation, etc. for assessing hydrologic responses in climate change scenarios. This SDSM model is developed through a combination of multiple linear regression and the stochastic weather generator (Wilby and Harris [9]; Mahmood and Babel [10]; and Hassan *et al.* [11]). The objectives of this study are to (1) download and screen the HadCM3 GCM for grids of two regions in Sarawak, (2) calibrate and validate the SDSM model with the observed rainfall record, and (3) forecast the rainfall corresponding to the two climate change scenarios available for HadCM3 GCM (H3A2 and H3B2) for the next 60 years as 2015–2044 and 2045–2074.

2   Materials and methods

2.1 Study area and data

The Sarawak state of Malaysia is the biggest state located on the Island of Borneo with total area 124,450 km^2. It experiences a wet and humid tropical climate throughout the year. The annual rainfall is between 2500 mm to 5000 mm. In this study, two rainfall stations have been selected as a case study, one at Belaga Town (Central Sarawak) and other at Limbang Town (Northern Sarawak). The details of these stations is given in Table 1 and shown in Figure 1. Observed daily historical rainfall data of these rain gauge stations were collected from the Department of Irrigation and Drainage, Sarawak (DID).

2.2 General circulation models

In this study, the output from the Hadley GCM 3 model (HadCM3) has been used to predict the future rainfall. It uses a 360 days per year and has a spatial grid with dimensions 2.5° latitude × 3.75° longitude. This GCM contains a complex model

Table 1: Details of rainfall stations.

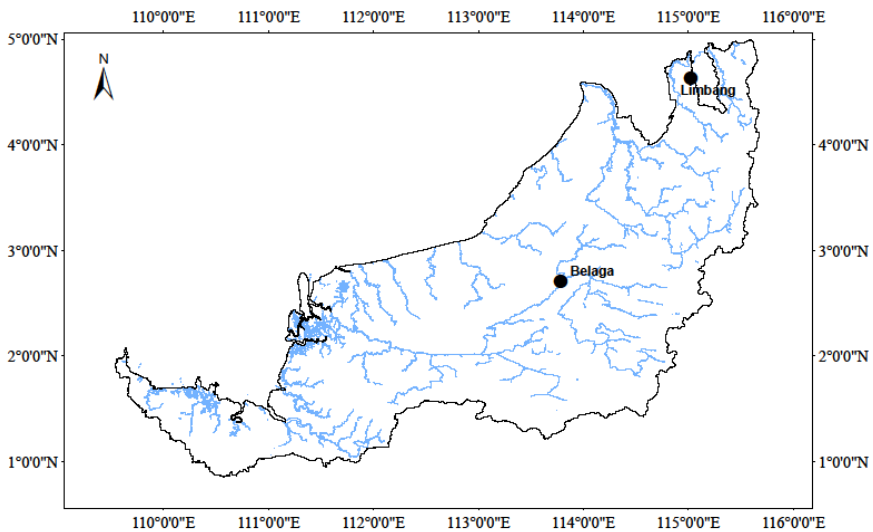| Region	Station ID	Station Name	Location		Height above MSL	Observed (1971–2010) Annual Rainfall (mm)
			Lat (°)	Long (°)		
Belaga	2737103	Belaga	002	113	40	3447
Limbang	4650007	Tegarai	004	115	15	4616

Figure 1: Location of rainfall station.

of land surface processes, which includes: (1) 26 land cover classifications; (2) 4 layers of soil, where temperature, freezing, and melting points are tracked; and (3) a detailed evapotranspiration function that depends on temperature, vapour pressure, vegetation type, and ambient carbon dioxide concentrations. It is considered the most mature and popular of the GCMs (Toews and Allen [12]). The predictors of NCEP and HadCM3 (H3A2 and H3B2) were obtained from a Canadian website: http://www.cics.uvic.ca/scenarios/index.cgi?Scenarios, for the periods of 1961–2001 and 1961–2099, respectively. H3A2 and H3B2 are respectively the IPCC emission scenarios A2 and B2 of HadCM3. These predictors are especially processed for the SDSM model. The NCEP predictors (2.5×2.5°) are first interpolated to HadCM3 grid resolution (2.5×3.75°) to eliminate spatial differences. Subsequently, the NCEP and HadCM3 predictors were normalized by utilizing long-term mean and standard deviations of 1961–

1990. The normalized predictors are only available for HadCM3 and CGCM2 in such a form that can be downloaded according to the coordinates of the study area and used directly for SDSM [10].

2.3 Statistical downscaling model (SDSM)

SDSM developed by Wilby *et al.* [13] is a hybrid of multiple linear regression (MLR) and the stochastic weather generator (SWG). MLR establishes a statistical/empirical relationship between NCEP, large-scale variables, and local-scale variables, and produces some regression parameters. These calibrated parameters, along with NCEP and GCM predictors, are then used by SWG to simulate up to 100 daily time series in order to create a better correlation with the observed time series.

In SDSM, some suitable predictors from the atmospheric predictors are selected through a multiple linear regression model, utilizing the combination of the correlation matrix, partial correlation, P value, histograms, and scatter plots. Multiple co-linearity must be considered during the selection of predictors. There are two kinds of optimization methods: (1) ordinary least squares (OLS) and (2) dual simplex (DS). The OLS produces comparable results with DS and is also faster than DS [8]. There are three kinds of sub-models – monthly, seasonal, and annual – that comprise the statistical/empirical relationship between the regional-scale variables (temperature and precipitation), and large-scale atmospheric variables. Annual sub-models drive the same kind of regression parameters for 12 months and the monthly sub-model represents 12 regression equations, giving different calibrated parameters for each of the 12 months. There are also two kinds of sub-models, conditional and unconditional; any of them can be used according to the local-scale variables. The unconditional sub-model is used for independent or unconditional variables such as temperature. The conditional sub-model is used for variables such as precipitation and evaporation [13, 14]. Most of the time, precipitation data is not distributed normally, but in the case of temperature, the data is distributed normally. SDSM can transform the data to make it normal before using the data in regression equations [15]. For example, [8, 15] used the fourth root for precipitation to render it normal before using it in a regression equation. Two kinds of daily time series, namely (1) daily historical site data and (2) NCEP daily predictors, are used to develop SDSM. The outputs of this model are daily time series, which can be produced by forcing the NCEP or HadCM3 predictors [8].

2.4 Model performance

A simulation of mean daily and monthly rainfall during the calibration and validation of the SDSM time series were checked by using the coefficient of correlation (R) and root mean square error (RMSE), and it is defined as:

$$R = \frac{\sum_{i=1}^{n}(Obs_i - \overline{Obs}) \cdot (Pred_i - \overline{Pred})}{\sqrt{\sum_{i=1}^{n}(Obs_i - \overline{Obs})^2 \cdot \sum_{i=1}^{n}(Pred_i - \overline{Pred})^2}} \qquad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \qquad (2)$$

In which, *Obs* = observed data value; *Pred* = predicted data value; $\overline{Obs}$ = mean observed data value and $\overline{Pred}$ = predicted mean data. The closer R value to 1 and RMSE value to 0, the predictions are better. To build confidence with the analysis's performance, mean rainfall including wet and dry spell lengths is compared graphically with the observed data. These graphical comparisons are able to identify pattern and variations captured by all models [11].

3   Results and discussions

3.1 Screening predictors for SDSM

A correlation analysis has been applied to choose a sensible combination of predictand and NCEP predictors available for the both grids. Each predictor was selected based on the highest correlation with the predictand (rainfall). For Belaga, r_850 (850 hPa relative humidity) and for Limbang, shum (surface specific humidity) became domain predictor variable. Although some predictor variables (e.g. shumas for Belaga and 5_zas for Limbang) did not show a higher correlation with rainfall, but they were still selected as predictor variables because the combinations of two or more variables would describe the conditional process of rainfall, which depends on many intermediate process. In general, the selection of predictor variable in this study are mostly similar to the ones applied in other studies, such [10, 11, 14].

3.2 Model calibration

The downscaled daily and monthly rainfall during the calibration period simulated by SDSM (using the NCEP variables) is shown in Table 2. Both models are not able to capture the daily data series very well and vice versa for the monthly data series. The difference in average of observed and simulated SDSM over the both regions are between 0.01–0.15 mm/day for daily rainfall and 0.48–4.47 mm/month for monthly rainfall. The coefficient of correlation varies between 0.393–0.504/month for monthly rainfall and 0.096–0.147/day for the daily rainfall. RMSE varies between 16.58–20.94 mm/day for the daily rainfall and 119.31–164.51 mm/month for the monthly rainfall. However monthly sub models have the better correlation with the recorded rainfall as compared to the annual models. And it is decided to use monthly sub-models for the validation period and future rainfall downscaling over the both regions.

Table 2: The performance of models during the calibration period.

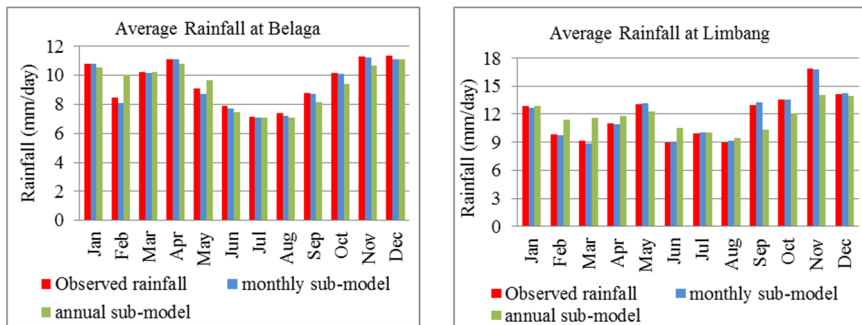|	Belaga			Limbang		
	X	RMSE	R	X	RMSE	R
Daily Rainfall (mm)						
Obs	9.49			11.79		
SDSM_A	9.36	16.58	0.096	11.70	21.11	0.103
SDSM_M	9.34	16.46	0.120	11.80	20.94	0.147
Monthly Rainfall (mm)						
Obs	288.92			358.87		
SDSM_A	285.05	124.93	0.393	356.32	164.51	0.413
SDSM_M	284.45	119.31	0.476	359.35	156.00	0.504

Figure 2: Average rainfall during model calibration period.

3.3 Validation of models

The downscaled daily and monthly rainfall simulated by SDSM using the NCEP, H3A2 and H3B2 predictor variables during the validation period (1991–2001) period is shown in Table 3. In general, all of the predictors able to estimate average daily rainfall very well and the difference in the daily rainfall with reference to observed rainfall vary from 0.25–0.93 mm/day for Belaga and 1.13–2.33 mm/day for Limbang. For the monthly rainfall, NCEP variables have the better correlation with the predictand as compared to H3A2 and H3B2 predictors. The estimated monthly rainfall varies by 3.3–28.31 mm/month for Belaga and 39.95–73.52 mm/month for Limbang. The performance of the model is better at Belaga than Limbang as Limbang has higher monthly rainfall.

Table 3:   The performance of models during the validation period.

	Belaga			Limbang		
	X	RMSE	R	X	RMSE	R
Daily Rainfall (mm)						
Obs	9.56			14.15		
SDSM	8.63	15.63	0.147	11.82	22.45	0.141
SDSM A2	9.81			13.02		
SDSM B2	10.29			12.76		
Monthly Rainfall (mm)						
Obs	291.00			430.72		
SDSM	262.69	101.43	0.501	357.20	234.88	0.237
SDSM A2	294.30	118.85	0.181	390.77	237.71	0.129
SDSMB2	308.56	117.33	0.217	382.90	244.33	0.124

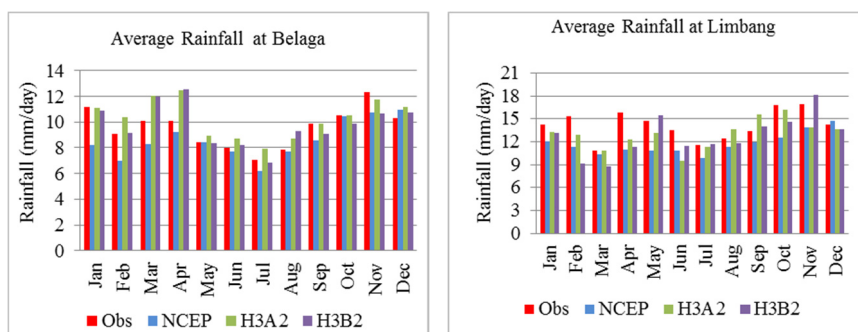Figure 3:   Average daily rainfall at Limbage during validation period.

3.4 Downscaling future rainfall under climate change scenarios

The future rainfall at Belaga and Limbang have been downscaled for the period of 2015–2044 and 2045–2074 using the HadCM3 output for H3A2 and H3B2 scenarios. The forecasted annual rainfall have been compared with the baseline period of 1971–2010 for the possible changes in future rainfall trend. It is noted that for Belaga region, the annual rainfall would be increased by 17.8% and 18.0% during 2015–2044 and 37.8% and 34.2% during 2045–2074 under H3A2 and H3B2 scenarios respectively. For Limbang region, the annual rainfall would be increased by 7.7% and 6.4% during 2015–2044 and 22.7.8% to 19.8% during 2045–2074 under H3A2 and H3B2 scenarios respectively.

Table 4:   Future changes in annual rainfall (%) with respect to the baseline (1971–2010) under HadCM3 A2 and B2 scenarios.

Scenario	Belaga		Limbang	
	2015–2044	2045–2074	2015–2044	2045–2074
SDSM A2	17.8	37.8	7.7	22.7
SDSM B2	18.0	34.2	6.4	19.8

4   Conclusions

In general, the monthly sub-models developed in SDSM have better correlation of observed climate data series corresponding to the present climate series as compared to the annual sub-models during the calibration period. During the validation period, daily and monthly rainfall downscaled at Limbang has the higher variation as compared to downscaled at Belaga due to the reason Limbang is a wetter region as compared to Belaga and the model under-estimated the rainfall overall at Limbang region.

Both HadCM3 scenarios expect future rainfall at Belaga and Limbang would be increasing significantly during the next 30 to 60 year period, which would cause high floods in both regions. On the other hand, it would improve the annual energy from existing and planned hydropower projects in both regions.

The results also highlights the gaps in downscaling modeling as SDSM under estimate the rainfall when applied to the wettest regions and there is a need to develop a site specific model for the wet region like Belaga and Limbang in Sarawak, which could have better correlation between the predictand and predictors to ensure the good performance of downscaling model.

Acknowledgements

References

[1]   S. Piao, P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudre, D. Labat, and S. Zaehle, "Changes in climate and land use have a larger direct impact than rising CO2 on global river runoff trends", *Proc Natl Acad Sci U S A,* vol. 104, pp. 15242-15247, Sep 25 2007.
[2]   D. Labat, Y. Goddéris, J. L. Probst, and J. L. Guyot, "Evidence for global runoff increase related to climate warming", *Advances in Water Resources,* vol. 27, pp. 631-642, 2004.

[3]   F. J. Wentz, L. Ricciardulli, K. Hilburn, and C. Mears, "How much more rain will global warming bring?" *Science,* vol. 317, pp. 233-235, Jul 13 2007.

[4]   H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling", *International Journal of Climatology,* vol. 27, pp. 1547-1578, 2007.

[5]   R. L. Wilby, L. E. Hay, W. J. Gutowski, R. W. Arritt, E. S. Takle, Z. Pan*, et al.*, "Hydrological responses to dynamically and statistically downscaled climate model output", *Geophysical Research Letters,* vol. 27, pp. 1199-1202, 2000.

[6]   R. L. Wilby, J. Troni, Y. Biot, L. Tedd, B. C. Hewitson, D. M. Smith*, et al.*, "A review of climate risk information for adaptation and development planning", *International Journal of Climatology,* vol. 29, pp. 1193-1215, 2009.

[7]   F. Wetterhall, A. Bárdossy, D. Chen, S. Halldin, and C.-Y. Xu, "Daily precipitation-downscaling techniques in three Chinese regions", *Water Resources Research,* vol. 42, 2006.

[8]   C. Huang, A. G. Barnett, X. Wang, P. Vaneckova, G. FitzGerald, and S. Tong, "Projecting future heat-related mortality under climate change scenarios: a systematic review", *Environ Health Perspect,* vol. 119, pp. 1681-1690, Dec 2011.

[9]   R. L. Wilby and I. Harris, "A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK", *Water Resources Research,* vol. 42, 2006.

[10] R. Mahmood and M. S. Babel, "Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India", *Theoretical and Applied Climatology,* vol. 113, pp. 27-44, 2012.

[11] Z. Hassan, S. Shamsudin, and S. Harun, "Application of SDSM and LARS-WG for simulating and downscaling of rainfall and temperature", *Theoretical and Applied Climatology,* vol. 116, pp. 243-257, 2014.

[12] M. W. Toews and D. M. Allen, "Evaluating different GCMs for predicting spatial recharge in an irrigated arid region", *Journal of Hydrology,* vol. 374, pp. 265-281, 2009.

[13] R. L. Wilby, C. W. Dawson, and E. M. Barrow, "SDSM - A decision support tool for the assessment of regional climate change impacts", *Environmental Modelling and Software,* vol. 17, pp. 147-159, 2002.

[14] J. T. Chu, J. Xia, C. Y. Xu, and V. P. Singh, "Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China", *Theoretical and Applied Climatology,* vol. 99, pp. 149-161, 2010.

[15] M. S. Khan, P. Coulibaly, and Y. Dibike, "Uncertainty analysis of statistical downscaling methods", *Journal of Hydrology,* vol. 319, pp. 357-382, 2006.