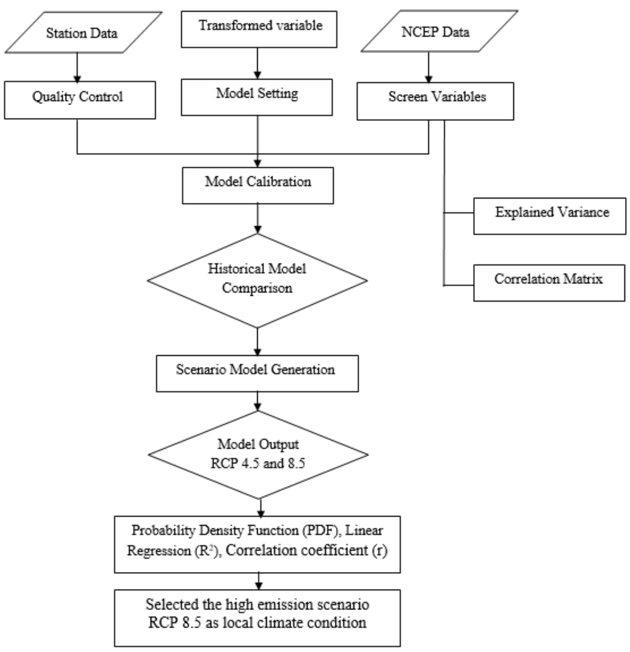
Global NEST Journal, Vol 24, No 2, pp 267-275

Statistical downscaling of regional climate model: A case study in Penang Island, Malaysia

Yee L.Z.[1], Tan K.W.[1]*and Huang Y.F.[2]
[1]Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, Perak, Malaysia
[2]Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Selangor, Malaysia
Received: 05/08/2020, Accepted: 20/12/2021, Available online: 28/03/2022
*to whom all correspondence should be addressed: e-mail: tankokweng@utar.edu.my
https://doi.org/10.30955/gnj.003429

Graphical abstract



Abstract

Flooding events are the most devastating natural disasters that Malaysia had experienced these recent years. It has become a challenge for the government and society to mitigate the flood events that occurs naturally, especially when climate change had worsened the situation. This study focuses on changes of rainfall, maximum and minimum temperature in Penang. With the fast growth rate of resident population and its economic status, obtaining the suitable and viable future climate scenario is crucial and significant to enhancing the flood forecasting capabilities in the Penang Island. The dataset obtained from CanESM2 GCM was used to generate the regionalized rainfall and temperature data based on Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment report (AR5) Representative Concentration Pathways (RCPs) of 4.5 and 8.5 W/m^2 emission scenario. The downscaling period (2005–2100) for maximum temperature (Tmax) and minimum temperature (Tmin) and rainfall time were developed using the statistical downscaling technique. For reliability analysis, the model outputs were analysed with Probability Density Function (PDF), Linear Regression and Pearson Correction Test. Based on the reliability analysis outcomes, the possible causes for acknowledging that the Penang Island could follow the high emission scenario RCP 8.5. This indicates that the average monthly rainfall and temperature of Penang Island may increase consequential of the regional climate change, resulting from an increasing population, the industrial development and ever-escalating transportation needs.

Keywords: Regional climate model, statistical downscaling, flood risk, Penang Island, Malaysia.

1. Introduction

Changes of climate system is unavoidable in this century as stated in the Sixth Assessment Report (AR6) published by Intergovernmental Panel on Climate Change (IPCC, 2021). With the increasing average surface temperatures (approximately +1.5°C), the extreme weather such as prolonged drought and intensive precipitation will likely to occur in most regions. The rapid increment of atmospheric greenhouse gases (GHG) concentration is the root cause invoking the climate change process, which have been reported in Sixth Assessment Report (IPCC (2021). Consequently, global warming which is one of the climate changes effects has caused the glaciers and ice sheets to melt at a fast rate which subsequently leads to rising global sea levels. It also tends to alter the hydrologic cycles and caused extreme weather events (IPCC 2013; 2021). This will eventually amplify the number of extreme weather events happened such as floods and droughts (Horton and McMichael 2008). Flooding events are recognised as the most devastating natural disasters that onslaught Malaysia more frequently these recent years. According to the Malaysian Department Irrigation and Drainage, (2000), a total 85 out of 185 rivers are at the risk of recurrent flooding. It also stated that approximately 29,800 km^2 (Total 9 % of the total Malaysia's land area) is vulnerable to flood disaster. A total of 4.82 million (22% of Malaysia's population) citizens will be affected by floods. Climate change is the foremost main factor to worsen the situation. Neither the government nor society is in a position to minimize or stop the flood event completely, as it is a natural phenomenon (FRMP, 2012).

Penang Island is undergoing rapid economic development and most of human settlements are located at the coastal areas. It has been affected by monsoon interchange season annually, for example the increased frequency and magnitude of flood events has been observed at the low-laying area in Penang Island. Different flood studies for Penang Island have stated that it is vulnerable to the impacts of climate change due to the sea level rise, temperature and rainfall pattern variation. In addition, the flood issue in the island can also be worsened with increase of rainfall intensity and expansion of subsidence surface areas. Gao *et al.,* (2021) analysed the Penang Island land subsidence to obtain the flood inundated areas. The land subsidence would increase 2.0% and 5.9% of the flood inundated area, based on the sea level rise projections. Osman *et al.,* (2021) stated that in this region, flooding and urban waterlogging issue are frequently observed due to climate changes and urbanization development. Othman *et al.,* (2021) reported that due to climate change, rainfall in northern region of Peninsular Malaysia, including Penang, has increased since last decades. Based on historical record, the average maximum rainfall per hour has increased six-fold from an average of 31 mm in the 1990s to 180 mm in the 2010s.

Prospectively, precipitation projection is crucial to assess the probability of flood and drought occurrence in a local area (Tan and Loh 2017). Based on Sixth Assessment Report (AR6), Intergovernmental Panel on Climate Change (IPCC) applied a different general circulation model (GCMs) to assess the future climate scenario based on different radiative forcing. It has stressed on the high likelihood of extreme weather events occurred in Southeast Asia region. The GCMs are feasible to predict the precipitation pattern in a particular region which are developed based on grid scale (200 to 600 km). Due to the coarse resolution of GCMs, the accuracy will be questionable when it is applied into local scale (50-100km) hydrological study. Hence, the regional climate model (RCM) downscaled from global climate models (GCM) is important tool to predict the variation of precipitation and temperature in local level (Goyal *et al.,* 2012, Trzaska and Schnarr 2014). It contains higher resolution and additional information, which can represent a better local landscape, hydrological profile, and local atmospheric processes (Shivam *et al.,* 2017).

Data scientists synthesize the GCMs data with different kind of downscaling approaches i.e dynamic and statistical methods. It has concluded that statistical downscaling method (SDSM) was faster and reliable approaches to generate the RCMs for the hydrological studies. For instance, Ang *et al.,* (2016) and Fung *et al.,* (2019) applied the SDSM on CanESM2 based on RCP4.5 and 8.5 scenario. The changes of rainfall and daily temperature patterns have been simulated under RCP 8.5 in Malaysia. The model output of projected rainfall downscaled by SDMS has produced the Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI). Tahir *et al.,* (2018) stated that statistical downscaling of precipitation and temperature is important to investigate the impact of climate change on the catchment. Under Coupled Model Intercomparison Project Phase 5 (CMIP5) model of CanESM2, the precipitation is expected to increase from last quarter of this century. Hassan and Harun (2011) apply the SDSM approach using GCM HadCM3 to generate RCM for Perak State, Malaysia. The downscaled daily precipitation and temperature for the period (2010-2099) showed the increase of total average annual rainfall and temperature. Abnormal precipitation pattern is estimated to be occurred in the future. It shows how to use of different downscaled RCMs to assess the rainfall pattern, eventually predicts the potential climate induced disasters in the study area. Unfortunately, lack of flood risk studies in Penang are found based on regional climate modelling approach. Due to high economy activities and high population density in the Penang Island, it is important to enhance the flood risk reduction capabilities by coupling with a reliable regional climate model (RCM). However, the issue of high reliable regional climate model remains as a major stumbling block to develop an effective flood disaster management plan in Penang Island. To address this issue, this study aims to demonstrate the application statistically downscaling work to develop the RCM for Penang Island. Hence, CanESM2 has been applied in this study with objectives to develop the regional climate model based on statistical approach and analyse the trend of climate variation based on selected representative concentration pathway (RCP).

2. Background of study area

Penang Island was chosen as the study area. The main constituent island of the state of Penang, Malaysia has the coordination of 5.59° to 5.12°N latitude and 100.17° to 100.56°E longitude. Penang Island is apart from the mainland of Peninsular Malaysia with the Malacca Strait located in between as shown in Figure 1. Approximately half of the Penang state population living in the island, whilst the others are residing in the Penang part of the mainland. The population number of Penang Island was recorded as 1,776,800 in 2018 with the density of 1,684/km^2 (DOS 2018).

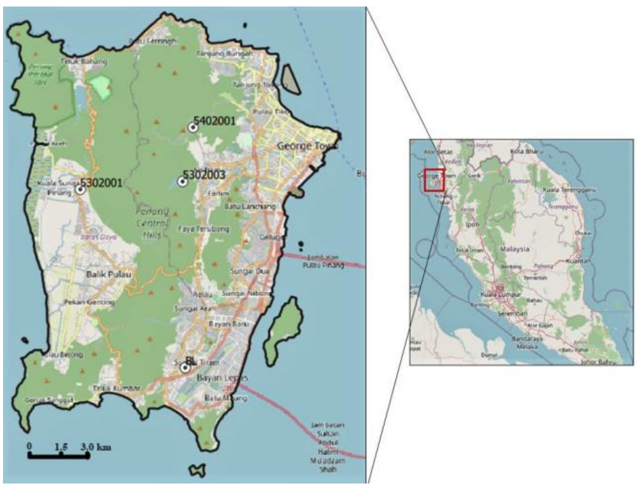


Figure 1. Study area and distribution of meteorological stations.

With the high density of population in Penang Island, the failure of flood disaster management and contingency action could lead to severe loss of lives, disease transmission and economic damages. Penang Island has a tropical monsoon and rainforest type of climate that falls under Köppen climate classification (*Af*). There are four weather monitoring facilities in Penang Island. According to the Malaysian Meteorological Department (MMD), the average annual precipitation of Penang Island was recorded approximately 2,670 mm with an average temperature of 23–32^{o}C.

3. Methodology

3.1. CanESM2 and NCEP/NCAR reanalysis data setup

The GCM model (CanESM2) output selected, and the statistical downscaling software (SDSM ver.5.3) used are both introduced in this section. The CCCma (Canadian Centre for Climate Modeling & Analysis) is a part of the Climate Research Division of Environment Canada. It is currently located at the University of Victoria, Victoria, British Columbia. The aim of CCCma is to form and apply the climate models in order to improve the understanding of climate change. Besides that, it also generates quantitative projections for future climate of Canada and globally. The model had also been used to contribute input to the Coupled Model Inter-comparison Project- phase 5 (CMIP5) (Flato *et al.,*2000; Flato *et al.,* 2013).

The CanESM2 model is into its fourth generation as a coupled global climate model. This model was evolved from the earlier CanESM1 version. It includes the fourth-generation atmospheric component (CanAM4), a spectral model which employs a T63 triangular truncation associated with spatial resolution of 128 x 64 horizontal grids and 35 vertical layers. The CanESM2 consists of 40 levels of ocean component of the upper ocean with approximately 10 m resolutions and 1.41^{o} x 0.94^{o} of the horizontal resolution (Arora and Boer 2014).

The NCEP/NCAR Reanalysis data set is a continually updated (since 1948) globally gridded data set that represents the state of the Earth's atmosphere. It is a combination product of the National Centres for Environmental Prediction (NCEP) and the National Centre for Atmospheric Research (NCAR). It has a resolution of roughly 210 km horizontally and 28 levels vertically. It has more than 80 variables, for example, geo-potential height, temperature, relative humidity, wind components and so on (NCAR/UCAR Research Data Archive, 2016). In this study, all the available data covered the Penang Island were collected and examined. The time period of the data collected was 1976 to 2005 as corresponding to availability of historical data set. This data set was downloaded from the Canadian Climate Data and Scenarios website (climate-scenarios.canada.ca/
?page=pred-canesm2) (Canadian Climate Data and Scenarios, 2019).

3.2. Model predictors

The independent variables used in the model are normally referred to as the model predictors. They are also the controlling variables of the model. The model predictands are the responding/ dependent variables for the modelling. Simply saying, we can obtain the relationship between predictors and predictands by input both of them into the model calibration. Table 1 shows the description of the predictor variables of the NCEP/NCAR reanalysis data set.

3.3. Station data (Predictands) validation

In this study, the station data were obtained from the 4 stations i.e stations no.5302001, 5302003, 5402001, and the Bayan Lepas (BL). At least 20 years of data for all the rainfall and temperature stations were used for the station data validation and the range of this period as shown in Table 2.

3.4. Downscaling experimental setup

The Quality Control function is important for the identification of gross data errors, specification of missing data codes and outliers prior to model calibration and downscaling process. Handling and processing of missing and imperfect data is essential for practical usage of data. In many instances it may be appropriate to transform the predictors and/or the predictand before the model calibration. This ensures a better fitting between predictors and predictand since their relationship are always changing and not linear. In this study, all 26 variables were transformed with lag transformation ranging -9 to 9 (total 20 lag transformation of each variable) to capture the best predictor in the following section.

The variable (Predictors) screening process is to assist the user in selecting the appropriate downscaling predictor by examining which of the large-scale variables that have a stronger relationship with the localized variables. This stage is usually the most challenging step in downscaling process as the result largely depends on the selection of the predictors. The variables with the highest partial correlation and the lowest or zero p-value associated with other variables are chosen. This step can also filter some of the variables with high correlation with predictand but are also highly associated with other variables as well. Such variables did not add much value to the decision-making process. In this study, the significant level was set to be 10 % in defining the dependency of variable with others. Hence, the selection of final predictors was based on the partial correlation of the variables and their *p*-value. The variables that are beyond the significant level were omitted from the selection and the predictors.

The partial correlation indicates the correlation between predictors and predictand without the influence of other variables, while the *p*-value measures how extreme the correlation between predictors and predictand is, in terms of probability. Smaller *p*-value indicates that this association is less likely to occur by chance.

Table 1. Description of Predictor Variables

Predictor variables	Description of predictor variables	Predictor variables	Description of predictor variables
mslpgl	Mean seal level pressure	**p5zhgl**	500hPa Divergence of true wind
p1_fgl	1000hPa Wind speed	**p850gl**	850hPa Geopotential
p1_ugl	1000hPa Zonal wind component	**p8_fgl**	850hPa Wind speed
p1_vgl	1000hPa Meridional wind component	**p8_ugl**	850hPa Zonal wind component
p1_zgl	1000hPa Relative vorticity of wind	**p8_vgl**	850hPa Meridional wind component
p1thgl	1000hPaWind direction	**p8_zgl**	850hPa Relative vorticity of wind
p1zhgl	1000hPa Divergence of true wind	**p8thgl**	850hPa Wind direction
p500gl	500hPa Geopotential	**p8zhgl**	850hPa Divergence of true wind
p5_fgl	500hPa Wind speed	**prcpgl**	Total precipitation
p5_ugl	500hPa Zonal wind component	**s500gl**	500hPa Specific humidity
p5_vgl	500hPa Meridional wind component	**s850gl**	850hPa Specific humidity
p5_zgl	500hPa Relative vorticity of wind	**shumgl**	1000hPa Specific humidity
p5thgl	500hPa Wind direction	**tempgl**	Air temperature at 2m

Table 2. Time period of data validation for all rainfall and temperature stations

Station name	Location (Lat, Long)	Baseline period
5302001	5.3917, 100.2125	1983 - 2005
5302003	5.3958, 100.2653	1983 - 2005
5402001	5.4236 , 100.2708	1983 - 2005
BL-Rainfall	5.3000 100.2667	1984 - 2005
BL-tempmin	5.3000 100.2667	1984 - 2005
BL-tempmax	5.3000 100.2667	1984 - 2005

Table 3. Selected predictors with respective partial correlation coefficient and p-value

Stations	Predictors	Descriptions	Partial r	p-value
5302001 (rainfall)	p1_ugl (6)	1000 hpa wind speed	0.054	0.000
	p8_ugl (1)	850 hpa wind speed	0.083	0.000
	p8thgl (-5)	850 hpa wind direction	0.067	0.000
	s500gl (-1)	500 hpa specific humidity	0.046	0.000
	tempgl (-9)	Air temperature at 2m	0.057	0.000
5302003 (rainfall)	p1_ugl (-2)	1000 hpa wind speed	0.064	0.000
	p8_ugl (2)	850 hpa wind speed	0.076	0.000
	p8thgl (5)	850 hpa wind direction	0.057	0.000
	prcpgl (-4)	Total precipitation	-0.033	0.000
	shumgl (-5)	1000 hpa specific humidity	0.055	0.000
5402001 (rainfall)	p1_ugl (4)	1000 hpa wind speed	0.055	0.000
	p8_ugl (4)	850 hpa wind speed	0.082	0.000
	p8_vgl (-8)	850hpa meridional wind	0.051	0.000
	p8thgl (-5)	850 hpa wind direction	0.071	0.000
	tempgl (-9)	Air temperature at 2m	0.055	0.000
BL-Rainfall (rainfall)	p1_ugl (-4)	1000 hpa wind speed	0.059	0.000
	p8_ugl (2)	850 hpa wind speed	0.087	0.000
	p8thgl (-8)	850 hpa wind direction	0.072	0.000
BL-tempmin (minimum temperature)	p1_ugl (9)	1000 hpa wind speed	-0.108	0.000
	p500gl (-9)	500 hpa geopotential	0.180	0.000
	shumgl (-2)	1000 hpa specific humidity	0.192	0.000
	tempgl (6)	Air temperature at 2m	0.136	0.000
BL-tempmax (maximum temperature)	mslpgl (6)	Mean sea level pressure	-0.120	0.000
	p1_ugl (-8)	1000hpa wind speed	-0.217	0.000
	p500gl (1)	500 hpa geopotential	0.170	0.000
	s500gl	500 hpa specific humidity	-0.099	0.000
	tempgl (9)	air temperature at 2m	0.101	0.000

Bracket “(x)” indicated the predictor undergone x-lag transformation.

The calibrated model takes the User-defined predictand along with a set of predictors to compute the parameters of multiple regression equations through an optimization algorithm. In this study, the ordinary least square method was chosen for the rainfall optimization while the dual simplex was for temperature optimization. The selected model type was that for the monthly basis for both rainfall and temperature downscaling. There were 12 regression functions developed by SDSM for an entire year. This selection was due to the high variance of climate throughout the year in Malaysia, especially its rainfall.

The type of process was set to be "unconditional" for temperature and "conditional "for rainfall downscaling. For the unconditional model, a direct link is assumed between predictors and predictands (for example, local wind speed) while for the conditional model, it is assumed there was an intermediate process in between (for example, local precipitation).

The parameter such as variance inflation and bias correction were initially set to be 12 and 1.0 respectively. Variance inflation controls the variance magnitude in downscaled data while bias correction was used to compensate the model tendency for over-or-under-estimating the mean of downscaled data. It was suggested that two of these parameters to be adjusted and the values corresponding to best validation result should be chosen. The length of 30 years available observed data sets was divided into three groups. The first group consisted of 20 years data from year 1976 to 1995 was used for the calibration process whereas the second group of year 1976 to 1986 data was used for the station data validation and the third group of 1996 to 2006 was used for determination of RCPs.

The CanESM2 output of the three different scenarios (historical, RCP 4.5 and RCP 8.5) and 'Year Length' parameter of 365 days is used. The historical CanESM2 was used to simulate the historical model (1983-2005) and for the comparison with actual historical data. This will prove that the predictors selected during screening process are suitable. 'Scenario Generator' operation generated 20 ensembles of synthetic daily weather series from futuristic large-scale predictors (GCM output) rather than NCEP/NCAR Reanalysis data.

For the validation part, it involved comparing the mean of generated climate data by 'Scenario Generator' and observed station data of each month over a 10 year's period. This comparison included graphical way (plotting graphs) and statistical way (computing coefficient of correlation, linear regression, standard normal distribution and probability density function between two sets of data).

Correlation coefficient (r) is a quantitative statistic which computes the correlation and dependency between two random variables. The result ranges from -1 to 1. When the result is close to 1 or -1, it represents a high relationship between two sets of variables in positive or negative trend respectively. In opposing, no relationship between variables when r=0. The computation of correlation coefficient is shown as below;

$$r = \frac{\sum_{i=1} n\left(x_i - \bar{x}\right)(y_i - y\text{Ǥ})}{\sqrt{\sum_{i=1} n\left(x_i - \bar{x}\right)^2 \sum_{i=1} n(y_i - y\text{Ǥ})^2}}.$$

Where x denotes first set of variables and y denotes the second set of variables. In this research, correlation coefficient was also used in analysis of the relationship between two chosen indices in later section.

Linear regression is defined as a linear approach to modeling the relationship between a dependent variable and one or more independent variables (Freedman, 2007). In this study, our goal is forecasting and hence the linear regression was used for the comparison of the simulated data and also observed station data. A High R-square value detonates that the simulated data is correlated with the observed station data.

In probability theory, the normal distribution is a type of continuous probability distribution for the real-valued random variable. For the normal distribution case, the simplest one is known as the standard normal distribution. The computation of standard normal distribution is shown as below:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

If the mean equals zero and the standard deviation equals one, it will be known as the probability density function (PDF). The PDF is a function whose value at any given sample point in the sample space can be interpreted as providing a relative likelihood that the value of random variable would equal the sample. In this study, the area under the graph of standard normal distribution and probability density function were computed or the validation of data and selection of RCP for the future trend as well. Finally, the selection of RCP climate scenario for Penang station will be based on all the four statistical tests mentioned above (Casella and Berger 2001).

4. Results and discussion

A total of 26 NCEP/NCAR variables (Predictors) which performed lag-transformation from lag -9 to lag 9 were analysed by partial correction test to identify the sensitivity of predictors to predictands (station data). The variables with the least or zero value of significant level (p-value) and highest value of partial correlation (partial r) were selected as the predictors. The list of selected predictors for all the stations with partial correlation coefficient and significant level respectively is shown in Table 3. The screening outcome shows that the 100 hpa wind speed (p1_ugl) is sensitive to all predictands for rainfall and temperature parameter. In addition, the 850 hpa wind speed (p8_ugl) and the 850 hpa wind direction (p8thgl) are also sensitive to the predictand for rainfall.

The partial correlation of predictors for the rainfall predictand was in the range of |0.033 - 0.087|; and the temperature predictand was in the range of |0.108 - 0.217|. This indicates that the daily rainfall tends to be less predictable as it has higher fluctuation (Ang, 2016). As shown in Table 3, station 5302003 is loaded with lowest partial correlation (-0.033) whereas BL-Rainfall station is determined as the highest partial correlation value (0.087) for rainfall predictand.

For the temperature predictand, the BL-tempmin station is loaded with the highest value of partial correlation and the BL-tempmax station held the lowest, which are 0.108 and -0.217 respectively. The negative sign of the partial correlation or otherwise known as the inverse partial correlation is a relationship between two variables whereby, they move in opposite direction e.g if variable X increase, then variable Y will decrease (Baba *et al.,* 2004).

The predictor 1000hpa wind speed tends to affect both the rainfall and temperature parameters. The altitude of 1000hpa and 850hpa are approximately at respectively 114m and 1456m above sea level. Penang Island is located at 3m above sea level with Penang Hill at height of 833m above sea level. The 1000hpa altitude acts as a wind boundary and includes the surface of Penang Island. Back and Bretherton (2005) had stated that surface wind speed contributes to the small amount of daily rainfall. Based on Figure 2, the surface level wind speed promotes the evaporation rate and formed shallow cumulus population. These cumuli eventually increase the relative humidity of atmosphere and hence also increase the deep convection and precipitation rate. Besides that, the surface wind (resulting from the surrounding island sea current) also acts as an agent that regulates the Penang Island regional temperature (Bigg *et al.,* 2003). This has factored in the 1000hpa wind speed as the main predictor that affects maximum and minimum temperature of the stations.

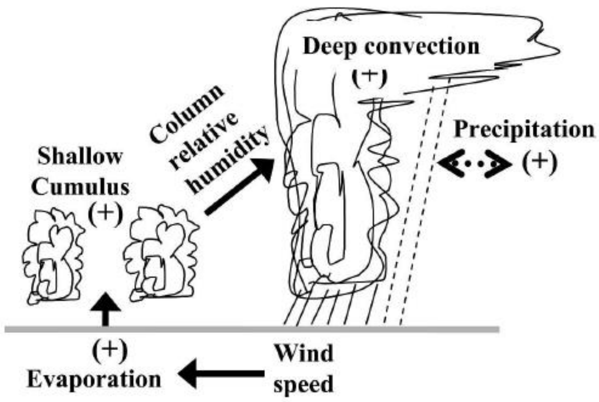


Figure 2. Schematic of convective initiation mechanism (Back and Bretherton, 2005).

There are two common predictors (variables) loaded in the model for rainfall stations (5302001, 5302003, 5402001 and BL-Rainfall) and these are the 850hpa wind speed and the 850hpa wind direction. These refer to the wind speed and wind direction at altitude of approximately 1456m above sea level. At higher level of altitude, wind speed and wind direction are larger compared to those of the lower altitude. According to Turgut and Usanmaz (2016), the average wind directions increased from 169° at lower altitude to 260° at higher altitude based on the data retrieved from a station located nearby the sea. In near the surface, there are plenty of structures, vegetation and mountains to limit the speed of the wind but at altitude of 850hpa there are none. This result in the greater wind speed at high altitude compared to lower altitude. A combination of higher wind speed and wind direction results in greater fluxes that increased the deep convection and hence the rainfall rate is higher (Back and Bretherton, 2005).

For the temperature station (BL-tempmax and BL-tempmin), the common variables loaded in the model are 500hpa geopotential and air temperature at 2.0 m. The air temperature is routinely measured at 2 meters height above ground at the meteorological station (BL-Rainfall). This enables us to identify the temperature profile, as it varies with height. The variable 500hpa geopotential refers to the altitude whereby the pressure is 500hpa (approximately 5.5km above sea level) and it forms a layer of atmosphere from the ground to the 500hpa height. This layer is also accordingly known as “thickness” from the meteorological perspective.

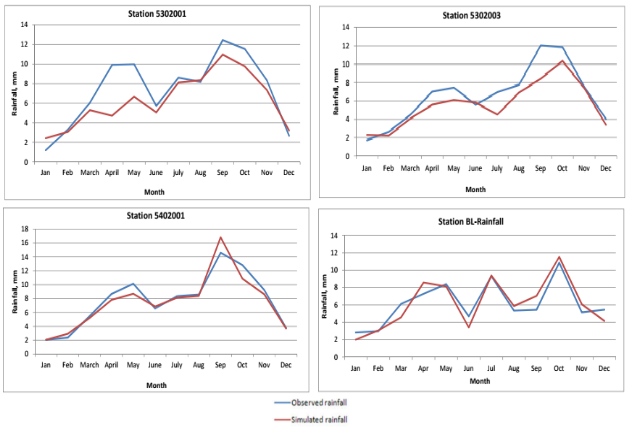


Figure 3. Comparison of observed and simulated average monthly rainfall.

The temperature data were obtained by the meteorological stations (BL-tempmin BL-tempmax) located at 2.0 m above sea level. The maximum and minimum temperature of Penang Island are possibly affected due to the coordinate of the island on Earth longitude and latitude, surrounding sea water, wind movement, precipitation, local activities and geographical terrain within the 500hpa geopotential boundary.

Table 4. Validation of historical model (SDSM) vs. station data (Observed) for Rainfall and Temperature Downscaling using coefficient of correlation

	Station Name	Correlation coefficient (r)
Rainfall	5302001	0.9665
	5302003	0.9467
	5402001	0.9652
	BL- Rainfall	0.9344
Minimum temperature	BL- tempmin	0.9724
Maximum temperature	BL- tempmax	0.9727

Table 5. Results of tests for rainfall and temperature stations

	Correlation coefficient (r)		Linear regression (R^2)		Probability density function (coverage %)	
Station name	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
5302001	(0.948)	0.944	0.606	(0.655)	73.22	(75.81)
5302003	0.868	(0.908)	0.665	(0.717)	97.00	(98.00)
5402001	(0.839)	0.816	0.614	(0.623)	99.50	(100.0)
BL-Rainfall	0.895	(0.950)	0.632	(0.682)	94.61	(99.50)
BL-tempmin	0.901	(0.919)	0.605	(0.643)	90.79	(99.65)
BL-tempmax	0.850	(0.890)	0.622	(0.631)	93.08	(98.78)

*Brackets indicates that the data has bigger value.

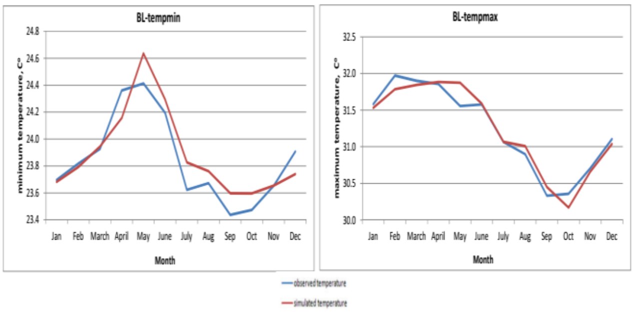


Figure 4. Comparison of observed and simulated maximum temperature and minimum temperature.

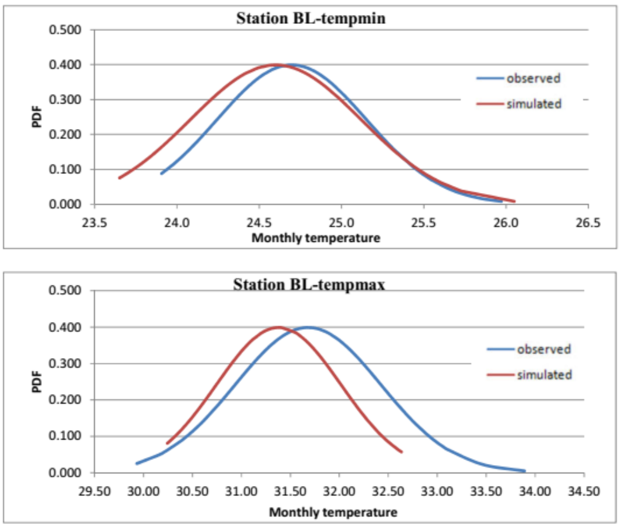


Figure 5. Non-cumulative probability density function (PDF) of monthly temperature for RCP 4.5 climate scenario and observation.

4.1. Comparison of historical model and Observation data (1983-2005)

In order to evaluate the performance of the SDSM downscaled models, the comparison of historical model and Observation data for 1983-2005 period was conducted in this study. The only difference between the observed and historical model average monthly rainfall was determined for station 5302001 as shown in Figure 4. There are differences of 5.1mm and 3.31mm respectively for April and May. Other than that, the SDSM performed well for all rainfall and temperature stations. For the BL-tempmax and BL-tempmax stations, there are slightly differences between the observed data and historical model (<0.3°C). This however is acceptable as it is less than 1 % (<0.01) in the historical model.

Statistical and graphical methods were used to evaluate the performance of SDSM software. The average monthly rainfall, maximum temperature and minimum temperature of each month throughout validation period were computed for historical model and observed station data. The graphical comparisons of these two sets of results are shown in Figures 3 and 4. These graphs represent the monthly average rainfall and monthly temperature for maximum and minimum. The generated average maximum and minimum temperature data are similar to the observed data. It is supported by the correlation coefficient r, which range from 0.9344 - 0.9665 for four rainfall stations, 0.9724 - 0.9727 for temperature stations (See Table 4).

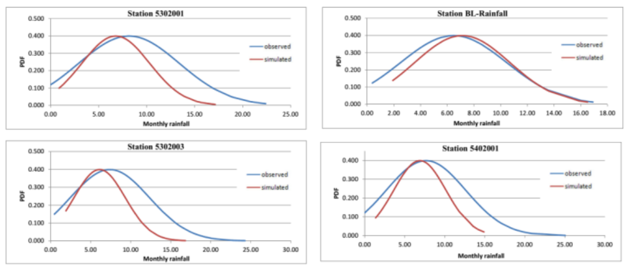


Figure 6. Non-cumulative probability density function (PDF) of monthly rainfall for RCP 4.5 climate scenario and observation.

4.2. Representative concentration pathways (RCPs)

The observed rainfalls and temperatures of the baseline period had been used to generate the projected time series up till year 2100 under RCP 4.5 and RCP 8.5 climate scenario. Similar to the data validation, the Correlation coefficient linear regression test and probability density function test were used to identify the local climate model scenario for Penang Island. As shown in Table 5, the value indicated the result meets the minimum criteria (r>0.9; R^2> 0.6) for correlation coefficient and linear regression and higher overlapped curve area (>90%) for probability density function (PDF).

The result shows that climate model based on RCP 8.5 has meet most of minimum criteria. For the PDF test, the overlapped curve area ranged from 73.22 - 99.50 % for the RCP 4.5 scenario and 75.81- 100 % for the RCP 8.5 scenario. In order to determine the similarity of the simulation and observation, Figures 5–8 tabulate the PDF curves and the mean values. The closer the mean value of the simulated data (RCPs) to the observed data will have the more identical graphs between simulated and observed data. The mean value of observed data ranges from 6.476 to 8.159 mm for the rainfall stations and from 24.70 to 31.7°C for the temperature data. For the RCP 4.5 scenario, the rainfall means values range from 6.119-7.122 mm and 24.6-31.4°C for temperature. For the RCP 8.5scenario, the range of mean value is 6.100- 7.164 mm and 24.7 -31.5°C for temperature.

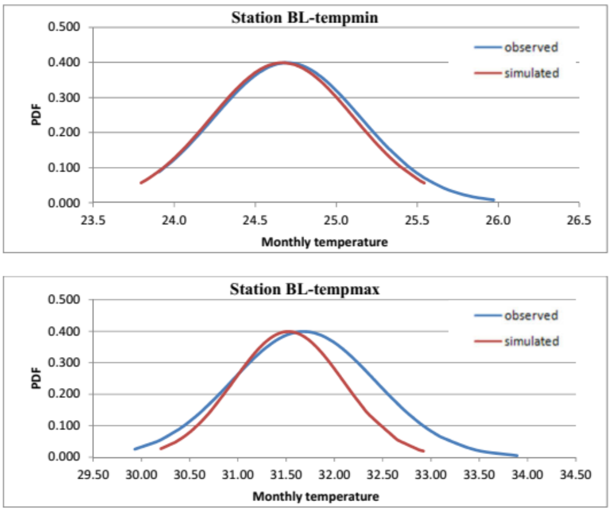


Figure 7. Non-cumulative probability density function (PDF) of monthly temperature for RCP 8.5 climate scenario and observation.

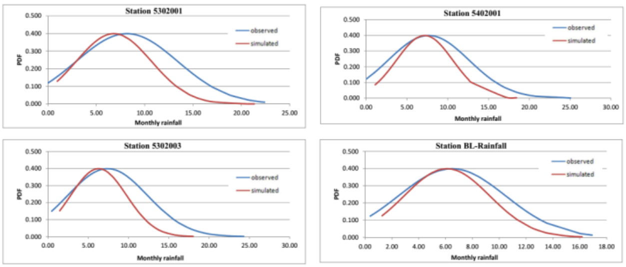


Figure 8. Non-cumulative probability density function (PDF) of monthly rainfall for RCP 8.5 climate scenario and observation.

Based on the RCP 8.5 climate scenario, the monthly rainfall for station 5302001 would have increased by a maximum 10 mm from 1982 to 2100. Such an increment would also contribute to the probability of flood event occurrence especially in the southwest and northeast monsoon season interchange period (September–October) (see Figure 9). Besides, the average minimum temperature of may be gained from 23.9°C to 26.2°C from the period of 1984 to 2100 (see Figure 10). The highest average minimum temperature is recorded as 27.6°C on June. A similar trend was observed for the highest average maximum temperature. It is estimated to increase from approximately 31.9°C to 34°C. There is also a clear trend of the monthly maximum temperature throughout the year. The temperature would be increased from January and hit the peak in March every year. The maximum temperature then reduces to its first crest in May and hit the lowest value in August. After that, the temperature is raised again until the following year of March and the cycle repeats.

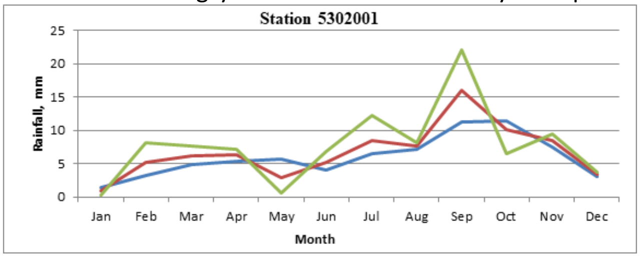


Figure 9. Downscaled climate model (rainfall) for station 5302001 based on RCP 8.5 climate scenarios in different time periods.

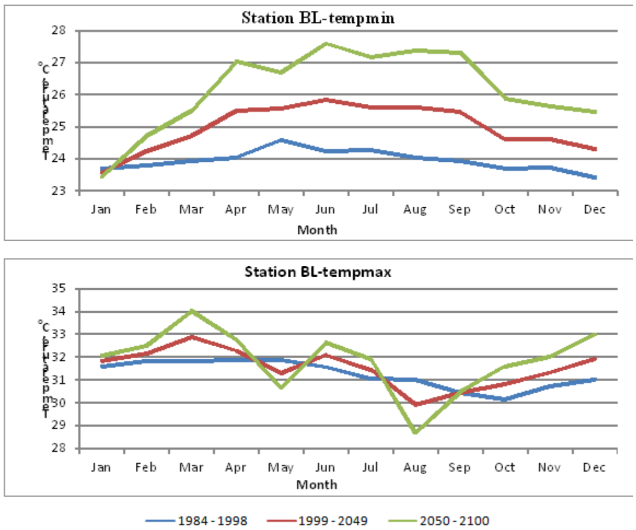


Figure 10. Downscaled climate model (minimum and maximum temperature) for BL station based on RCP 8.5 climate scenarios in different time periods.

5. Conclusion

This study has successfully developed a regional climate model for Penang Island based on representative concentration pathway. The outcome shows that selection of suitable predictors is important to establish a creditable downscaled model. The dominant predictors for rainfall model are found to be the 1000 hpa wind speed (p1_ugl), 850 hpa wind speed (p8_ugl) and 850 hpa wind direction (p8thgl) while for the temperature model, they are loaded with 1000 hpa wind speed (p1_ugl), 500 hpa geopotential (p500gl) and air temperature at 2m (tempgl). Based on the reliability analysis outcomes, the possible causes for acknowledging that the Penang Island could follow the high emission scenario RCP 8.5. It may be due to urbanization and rapid growth of human population in the study area. The rising population subsequently resulted in the increase of industrial production (especially in manufacturing sector), transportation needs and storage services as well. Furthermore, the tourism of Penang Island is bound to increase too as it is one of economy sectors of the State. All of these shall contribute to the higher significant greenhouse gases (GHGs) emission and deforestation, thus causing the inevitable regional climatic change in the area.

Acknowledgements

We would like to thank the Malaysia Ministry of Higher Education for providing research funding – FRGS/1/2019/TK01/UTAR/02/1. We thank to CCCma, Malaysian Meteorological Department (MMD) and Department of Irrigation and Drainage (DID) for providing data and information to complete the climate modelling downscaling work.

References

Ang J.T. (2016). Drought forecasting using SPI and RDI for Langat River Basin under AR5 Climate change Scenarios. *BEng Thesis*. University Tunku Abdul Rahman.

Arora V. and Boer G. (2014). Terrestrial ecosystems response to future changes in climate and atmospheric CO2 concentration. *Biogeosciences*, **11**(15), 4157–4171.

Baba K., Ritei S., and Masaaki S. (2004). Partial correlation and conditional correlation as measures of conditional independence. Australian and New Zealand. *Journal of Statistics*, **46**(4), 657–664.

Back L.E. and Bretherton C.S., (2005). *The relationship between wind speed and precipitation in the Pacific ITCZ*. Washington: Department of Atmospheric Science, University of Washington.

Bigg G.R., Liss P.S., and Osborn T.J. (2003). Review the role of the oceans in climate. *International Journal of Climatology,* **23**, 1127–1159.

Canadian Climate Data and Scenarios. (2019). CanESM2 predictors: CMIP5 experiments. [online] Available at: <http://climate-scenarios.canada.ca/?page=pred-canesm2> [Accessed 7 August 2019].

Casella G. and Berger R.L. (2001). *Statistical Inference* (2nd ed.). United States: Duxbury Thomson Learning.

DID (2000). *Urban storm water management manual for Malaysia*. Department of Irrigation and Drainage (DID), Kuala Lumpur Malaysia.

Department of Statistics Malaysia, 2018. *Current Population Estimates Malaysia 2018*. Kuala Lumpur Malaysia.

Flato G.M., Boer G.J., Lee W.G., McFarlane N.A., Ramsden D., Reader, M.C., and Weaver A.J. (2000). The Canadian centre for climate modeling and analysis global coupled model and its climate. *Climate Dynamics,* **16**(6), 451–467.

Flato G., Marotzke J., Abiodun B., Braconnot P., Chou S., Collins W., Cox P., Driouech F., Emori S., Eyring V., Forest C., Gleck-ler P., Guilyard E., Jakob C., Kattsov V., and Reason C., Rum-mukainen M. (2013). *Evaluation of Climate Models. Climate Change 2013: The physical science basis. Contri-bution of orking Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: IPCC.

Freedman D. (2007). *Statistical Models: Theory and Practice*. Cambridge University Press, UK.

FRMP. (2012). *Flood Risk Management Planning in Scotland: Arrangement for 2012 – 2016*. Scottish Government.

Fung K.F., Huang Y.F., Koo C.H., and Mirzaei M. (2019). Standardized Precipitation Index (SPI) And Standardized Precipitation Evapotranspiration Index (SPEI) Drought Characteristic and Trend Analysis Using The Second Generation Canadian Earth System Model (CanESM2) Outputs Under Representative Concentration Pathway (RCP) 8.5. *Carpathian Journal of Earth and Environmental Sciences,* **14**(2), 399–408

Gao G., San L.H., and Zhu Y. (2021). Flood Inundation Analysis in Penang Island (Malaysia) Based on In SAR Maps of Land Subsidence and Local Sea Level Scenarios. *Water*, **13**, 1518.

Goyal M.K., Ojha C.S.P., and Burn D.H. (2012). Nonparametric statistical downscaling of temperature, precipitation, and evaporation in a semiarid region in India. *Journal of Hydrologic Engineering*, **17**(5), 615–627.

Hassan Z. and Harun S.B. (2011). *Statistical Downscaling for Climate Change. Cenarios of Rainfall and Temperature. In: United Kingdom-Malaysia-Ireland Engineering Science Conference 2011 (UMIES 2011)*. Kuala Lumpur, 12–14 July 2011. Universiti Teknologi, Johor Bharu, Malaysia.

Horton G. and McMichael T. (2008). *Climate Change Health Check 2020*. Doctors for the Environment, Australia.

IPCC. (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

IPCC. (2021). Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.

NCAR/UCAR Research Data Archive (2016). CISL RDA: NCEP/NCAR Global Reanalysis Products, 1948-2021. [online] Available at: <http://rda.ucar.edu/datasets/ds090.0/> [Accessed 17 February 2021].

Osman S., Chen L.F., Mohammad A.F., Xing L., and Chen Y. (2021). Flood Modeling of Sungai Pinang Watershed under the Impact of Urbanization. *Tropical Cyclone Research and Review*, **10**(2), 96–105.

Othman M., Ahrasan N., and Tan T. (2021). *Report on Climate Change Impacts in Penang*. Penang Green Council, Geogretown Malaysia.

Shivam, Goyal M.K., and Sarma A.K. (2017). Analysis of the change in temperature trends in Subansiri River basin for RCP scenarios using CMIP5 datasets. *Theoretical and Applied Climatology*, **129**(3), 1175–1187

Tahir T., Hashim A.M., and Yusof K.W. (2018). Statistical downscaling of rainfall under transitional climate in Limbang River Basin by using SDSM. *IOP Conference Series: Earth and Environmental Science,* **140**, 1–8.

Tan K.W. and Loh P.N. (2017). Climate change assessment on rainfall and temperature in cameron highlands, malaysia using regional climate downscaling method. *Carpathian Journal of Earth and Environmental Sciences*, **12**(2), 413–421.

Trzaska S. and Schnarr E. (2014). *A Review of Downscaling Methods for climate change projections*. United States Agency for International Development.

Turgut E.T. and Usanmaz O. (2016). An analysis of vertical profiles of wind and humidity based on long-term radiosonde data in turkey. *Applied Science and Engineering*, **17**(5), 830–844.