Projections of Cardiovascular Disease Mortality in Peninsular Malaysia Using Statistical Downscaling Based on Cluster Approach

Aina Izzati Mohd Esa, Syafrina Abdul Halim[*], Norhaslinda Ali

Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400, Selangor, Malaysia

Abstract Projecting the mortality of cardiovascular disease in future is crucial in preparing the mitigation strategies. The purpose of this research is to estimate number of deaths of the cardiovascular disease in Peninsular Malaysia based on future temperature projections using the cluster approach. Ward's method is used to identify the number of clusters of 45 meteorological stations by calculating the shortest distance between the two coordinates of the stations. The output of global climate model (GCM) is incredibly useful for the projection of future temperature, but the large bias in the observational datasets may lead to inaccurate projection. To tackle the bias, a good fitted model for temperature series is important in order to ensure that the mean and variability of the observed series are well captured. It is important to estimate the parameters for each cluster precisely. Furthermore, a good fitted model for temperature series is also crucial in order to ensure that the mean and variability of the observations are well captured. Thus, this study proposed the appropriate statistical distribution for the temperature series to be associated in the bias correction method (BCM) using the quantile mapping (QM) technique to reduce the biases between observations and historical GCM temperature data series. Next, Ward's method is applied to determine the optimal number of clusters for Peninsular Malaysia. The results have shown that the proposed model is able to reduce the temperature series biases between the GCM and the observations. Six clusters throughout Peninsular Malaysia have been selected based on Ward's method. The projection number of deaths of cardiovascular disease under is estimated to increase between 2006 and 2100 in all clusters across Peninsular Malaysia, based on the temperature projections.

Keywords Bias Correction Model, Cardiovascular Disease, Global Climate Model, Ward, Quantile Mapping

1. Introduction

Climate change has been shown in recent research to dramatically increase temperature-related mortality. However, the potential effects of climate change-related health risks will differ greatly between diseases. The population's exposure to local social and environmental stressors, and access to health care, potential health outcomes will differ across regions and countries. Global climate models (GCMs) is the common model to assess future climate. GCMs has a long history of growth and have a rare opportunity to physically model global climate and uncertainty in novel ways [1]. However, the resolution of GCMs is too coarse and is associated with distortions due to the structure of the model, the processes of parameterization, assumption, and calibration [2]. As a

result, GCMs cannot be used to forecast future climates directly. Downscaling is the process of reducing a coarse resolution to the GCM output's finest resolution. Statistical and dynamical downscaling are the two types of downscaling. The process of developing statistical relationships between local climate variables and large atmospheric variables is known as statistical downscaling. Statistical downscaling is used in this study due to several advantages such as it is less computational, good in capturing the bias, and can provide finer resolution outputs than dynamical downscaling [3]. Statistical downscaling encompasses a wide range of techniques. Bias correction method (BCM) is one of the statistical downscaling approaches in which it can correct the bias between observations and historical GCM data that remain valid under future conditions [4]. One of the techniques in BCM namely quantile mapping (QM), is often used to capture changes in the mean and variability of a GCM [5]. Grouping similar stations in a good structure could explain the climate of Peninsular Malaysia accurately [6]. The estimated parameter values without clustering may not represent the whole Peninsular Malaysia quite well as each region has different climatic characteristics as well as different topographical factor. It is important to estimate the parameters for each cluster precisely [7]. Furthermore, a good fitted model for temperature series is also crucial in order to ensure that the mean and variability of the observations are well captured. Thus, this study proposed the appropriate statistical distribution for the temperature series to be associated in the bias correction method (BCM) using the quantile mapping (QM) technique to reduce the biases between observations and historical GCM temperature data series. Next, Ward's method is applied to determine the optimal number of clusters for Peninsular Malaysia. Therefore, the aim of this study are to project the future series of daily mean temperatures (2006-2100) and to calculate the cardiovascular disease mortality rate (2006-2100) in Peninsular Malaysia based on the projections of temperature using the cluster approach.

2. Literature Review

2.1. Temperature in Malaysia

According to the Intergovernmental Panel on Climate Change (IPCC), Southeast Asian countries, especially developing countries like Malaysia, would be the most vulnerable to heatwaves [8]. From 1961 to 2002, surface temperature in most parts of Malaysia showed major warming trends [9]. Warming patterns of between 2.7°C and 4.0°C/100 years were observed at various stations in Peninsular Malaysia and northern Borneo. Kuching and Bintulu stations had lower rates of between 1.0°C and 1.5°C/100 years, while Miri had no noticeable warming or cooling pattern. From 2008 to 2010, the lowest heatwave

index was 27.3°C in Kuching and the highest heatwave index was 35.0°C in Sandakan in East Malaysia [10]. From 2001 to 2010, the highest heatwave index experienced in Kuala Lumpur with an increase of 9.1°C, and the lowest heatwave experienced in Alor Setar with an increase of 0.1°C across Peninsular Malaysia [11]. In addition, the moderate heatwave index was found at 4.2°C in Kuantan, and the longest heatwave, spanning 24 days, occurred in Ipoh, Perak, with amplitudes ranging from 29.4°C to 33.0°C [11]. The heatwave's characteristics were also compared to spatial distribution maps, which revealed that the southeast, northeast, and west parts of Peninsular Malaysia experience the most heatwaves. During dry season, the highest heatwave index occurred between March and July [11]. Tang [12] had investigated daily mean temperature in five different locations: Kota Kinabalu, Kuching, Malacca, Kuantan, and Subang Jaya. The annual moving average of daily mean temperature in Kota Kinabalu has been discovered to be on the rise, with temperatures fluctuating between 26°C and 28.5°C. In Kuantan, the temperature varies from 25°C to 28°C, while in Subang Jaya, temperature ranges from 26°C to 28.7°C. Meanwhile, Kuching has a temperature range of 25.5°C to 27.5°C, while Malacca has a temperature range of 26°C to 28.5°C. Kuching had the smallest annual moving average increment of daily mean temperature, followed by Kota Kinabalu, Malacca, Kuantan, and Subang Jaya. Due to the slower rate of development, Kuching experienced the smallest temperature rise. This confirmed that the heatwave conditions in Malaysia are concerning, necessitating further research and investigation because of the effects of heatwaves that may directly affect the agriculture, economics, and human health [10-11].

2.2. Cardiovascular Disease (CVD)

Cardiovascular diseases (CVDs) are a category of diseases that affect the heart and blood vessels. Coronary heart disease (CHD), one of the CVDs, is the leading cause of death globally. CHD mortality has risen by more than threefold in Malaysia over the last 40 years and it continues to grow. CVDs were ranked third in the 1950s, steadily rising to first place in 1970, and continued to grow in 1989, with 16.5 times increase from 1.8% in 1950 to 29.6% in 1989 [13]. This suggests that CVDs have surpassed respiratory diseases, neoplasms diseases, infectious diseases, metabolic diseases, and blood diseases as the leading cause of death among the six disease groups. The four main diseases under CVD are coronary heart disease, cerebrovascular disease, hypertension, and rheumatic heart disease. Deaths due to CHD increased from 32.7% in 1965 to 38.2% in 1989, while deaths due to cerebrovascular disease, hypertension, and rheumatic heart disease decreased from 33.1%, 16.0%, and 4.5% in 1975 to 30.1%, 1.4%, and 1.7% in 1989 [13]. Other types of CVDs have decreased from 46% in 1965 to 28.6% in 1989 [13].

According to the Department of Statistics Malaysia (DOSM), in 2016, CHD was the leading cause of death with 13.2%, followed by pneumonia (12.5%), cerebrovascular disease (6.7%), transportation incidents (5.4%), and malignant neoplasms of the trachea, bronchi, and lung (2.2%) [14]. In the next two years, the proportion rises by 13.9% and 15.6%, respectively [15-16]. Although the number of deaths of CHD decreased marginally in 2019, it remained as the leading cause of death with 15.0%, followed by pneumonia (12.2%), cerebrovascular disease (8.0%), transportation incidents (3.8%), and malignant neoplasm of the trachea, bronchus, and lung (2.4%) [17].

2.3. Cluster Analysis

Cluster analysis is the method of combining objects from the same cluster with objects from other clusters. The formulation of a query, the selection of a distance metric, the selection of a clustering process, the definition of the number of clusters, the interpretation of profile clusters, and the evaluation of clustering validity are all part of cluster analysis. The aim of cluster analysis is dealing with the underlying structure of data in order to gain insights of previously unknown data, as well as to identify significant features and probable, classification to determine the degree of similarity among data points, and compression to organize and summarize the data into understandable segments [18-19]. Cluster analysis is becoming more common in atmospheric science for redefining climate divisions using specific climate and meteorological data [18-20]. Several studies have used cluster analysis to determine the spatial and temporal trends of monthly precipitation [6, 21-23]. It has been suggested that cluster analysis can be used in variable such as temperature to investigate the regional characteristics of temperature distribution through pattern and spatial analysis temperature variability on an annual and monsoon-seasonal basis for various climatic regions [6]. Several studies have used the k-means clustering approach to classify the spatial and temporal patterns [23]. The k-means method allows users to determine the number of clusters to be used. However, incorrectly specifying the number of clusters can lead to severe errors in the analysis [6]. On the other hand, Ward's method is less sensitive in determining the number of clusters and capable of capturing extreme values [7]. Hence, Ward's method is used for clustering.

2.4. Bias Correction Method Quantile Mapping

BCMs use a transfer function to correct the bias between observations and historical climate variables. The advantages of the transfer function are assumed to be stationary and remain valid for the future period [4]. Several BCMs have been developed to downscale the meteorological variables from the GCM outputs, varying from the simple or linear method to the advanced or nonlinear method [4-5, 24]. Linear scaling, variance scaling, power transformation, QM, daily bias correction, local intensity scaling, daily translation, delta change method, multiple linear regression, multiple linear regression with randomization, analogue method, and nearest neighbour analogue are only a few examples of BCMs [4-5, 24-31]. QM is one of the best BCM methods because of the ability of QM to adjust the mean, the quantile, the variances, and preserving the extremes [4-5, 26-27]. The aim of QM is to match the GCM outputs' probability density distribution (PDF) with the PDF of observed data [4, 32]. This can be accomplished by adjusting the occurrence distributions of meteorological variables using a transfer function [33]. QM can be classified into three types: distribution-derived QM, parametric QM, and nonparametric QM [34]. As it offers the best combination of accuracy and robustness [26-27], QM based on distribution derived is used in this study. Previous research has assumed that temperature series are usually normally distributed with a symmetrical shape of distribution [4-5, 26-27, 35-37]. However, the distribution's shape of temperature series is positively skewed. Thus, Gamma distribution with three parameters is proposed in this study. The addition of a location parameter, θ which can be used to determine the best central value for describing the data. In particular, the temperature series is fitted with a QM with three parameters: shape (α), scale (β), and location (θ) of the Gamma distribution to reduce the biases between observations and GCMs temperature.

3. Methods/Methodology

3.1. Data

The historical daily mean temperature for Peninsular Malaysia was obtained from 45 weather stations between 1976 and 2005 [23]. The Interdisciplinary Climate Research Model (MIROC5) outputs are used, which were modelled as part of the Phase 5 Coupled Model Comparison Project (CMIP5). In this study, the historical GCM of daily mean temperature (1976-2005) and the future GCM of daily mean temperature (2006-2100) were obtained under RCP4.5 and RCP8.5. Representative Concentration Pathways (RCPs) are a collection of estimated initial values for radiative forcing, CO_2 concentration, and temperature anomaly up to the year 2100. RCPs were created using economic activity, energy sources, population growth, and other socioeconomic variables [38]. RCP8.5 represents increasing greenhouse gas emissions over time, while RCP4.5 represents a stabilization scenario in which radiative forcing stabilizes shortly after 2100 while remaining below the long-run radiative forcing target level. Radiative forcing is the result of a process that changes the balance of incoming and

outgoing energy in the Earth-atmosphere system. As shown in Table 2, each RCP has its own emissions trajectory and radiative forcing. The mortality data for unstable angina, STEMI, and NSTEMI being recorded at Malaysia's hospital were obtained from [39-42].

Table 1. The list of stations and the coordinates of 45 stations throughout Peninsular Malaysia

State	Station Number	Longitude (°)	Latitude (°)
Perlis	1	100.225	6.525
	2	100.275	6.525
Kedah	3	100.325	6.125
	4	100.375	6.125
Penang	5	100.375	5.375
Perak	6	101.025	4.575
	7	101.075	4.575
	8	101.325	4.525
	9	101.375	4.525
	10	100.825	4.825
	11	100.875	4.825
	12	100.725	4.225
	13	100.775	4.225
Selangor	14	101.525	3.125
	15	101.575	3.125
	16	101.525	3.025
	17	101.575	3.025
	18	101.625	2.725
	19	101.675	2.725
Melaka	20	102.225	2.225
	21	102.275	2.225
Negeri Sembilan	22	102.225	2.725
	23	102.275	2.725
Johor	24	102.925	1.825
	25	102.975	1.825
	26	103.325	2.025
	27	103.375	2.025
	28	103.625	1.625
	29	103.675	1.625
	45	103.825	2.425
Kelantan	30	102.225	5.525
	31	102.275	5.525
	32	103.025	5.425
	33	103.075	5.425
Terengganu	34	103.125	4.325
	35	103.425	4.525
Pahang	36	102.325	3.975
	37	102.375	3.975
	38	103.225	3.775
	39	103.275	3.775
	40	103.325	3.775
	41	102.425	3.425
	42	102.475	3.425
	43	103.025	3.075
	44	103.075	3.075

Table 2. The emission trajectory and radiative forcing

RCP	Radiative forcing (Wm^2)	CO_2 equivalent ($p.p.m$)	Temperature anomaly (℃)	Pathway
RCP4.5	4.5 post 2100	650	2.4	Overshoot-free stabilization
RCP8.5	8.5 in 2100	1370	4.9	Rising

3.2. Ward's Method

Hierarchical (HCA) and non-hierarchical (NCA) cluster analysis are the two types of cluster analysis (NHCA). HCA is used in this study to compress, organize, and summarize data into manageable clusters. HCA connects the coordinates that are the most similar. Starting with an agglomerative approach. It begins with N data points that are grouped together to form individual clusters. To form a new cluster, the shortest distance between N and $N-1$ neighbours is measured. After that, consider the distance between the remaining $N-2$ coordinates and the newly created cluster. The two coordinates are then connected to the shorter distance, either by adding data coordinates to the two clusters or by constructing a new cluster from two new data points. This process is repeated until is a single cluster of N coordinates is achieved, regardless of the absolute distance between them. This is an incredibly good algorithm for reducing dimensionality. The results of HCA are often represented as a dendrogram, a tree-shaped graph that contains crucial information about the cluster's measured distances and the couplings formed. HCA has advantages over the NCA where the cluster number determination is not sensitive. One of the methods in HCA is Ward's form. Ward's approach has the advantage of reducing overall within-cluster variance. The formula for measuring the distance between two points is

$$D = \sum_{i=1}^{n}\sum_{j=1}^{n} \sqrt{(x_i - x_i)^2 + (y_j - y_j)^2} \qquad (1)$$

where, n is 45 number of stations and (x_i, y_j) is the coordinates of each stations.

3.3. Quantile Mapping (QM)

QM defines a transfer function to reduce the biases between observations and historical GCM temperature data series. The transfer function is

$$b = \frac{1}{CDF_x}\left(CDF_y \times y\right) \qquad (2)$$

where, b represents the corrected temperature between observations and historical GCM, y represents the observations of daily mean temperature, CDF_x and CDF_y represents the cumulative distribution function (CDF) of historical GCM and observations of daily mean temperature, respectively. Equation (1) is used to calculate future daily mean temperature projections under RCP4.5 and RCP8.5. The observations (y) were replaced with future GCM temperature under RCP4.5 (v_1) and RCP8.5 (v_2) while the historical GCM temperature (x) were

replaced with the corrected temperature (b). Let p_1 and p_2 represent the projection of future daily mean temperature under RCP4.5 and RCP8.5.

$$p_1 = \frac{1}{CDF_h}\left(CDF_{v_1} \times v_1\right) \qquad (3)$$

$$p_2 = \frac{1}{CDF_h}\left(CDF_{v_2} \times v_2\right) \qquad (4)$$

The PDF of Gamma distribution with three parameters is

$$f(x; \alpha, \beta, \theta) = \frac{(x-\theta)^{\alpha-1}e^{-\frac{(x-\theta)}{\beta}}}{\beta^{\alpha}\gamma(\alpha)} \qquad (5)$$

where, α the shape, β the scale, θ the location, and $\gamma(\alpha)$ gamma function. The parameter values can be estimated using Maximum Likelihood Function (MLE). Hence, the $\ln L$ is

$$\ln L = -n\alpha \ln \beta - n \ln \gamma(\alpha) + (\alpha - 1)\sum_{i=1}^{n}\ln(x_i - \theta) - \frac{\sum_{i=1}^{n}(x_i-\theta)}{\beta} \qquad (6)$$

By taking the partial derivatives with respect to α, β, and θ. Then, let the equation equals to zero,

$$\frac{\delta \ln \alpha}{\delta \alpha} = \sum_{i=1}^{n}\ln(x_i - \theta) - n \ln \beta - n\psi(\alpha) = 0 \quad (7)$$

$$\frac{\delta \ln \beta}{\delta \beta} = \frac{\sum_{i=1}^{n}(x_i-\theta)}{\beta^2} - \frac{n\beta}{\alpha} = 0 \qquad (8)$$

$$\frac{\delta \ln \theta}{\delta \theta} = (\alpha - 1)\sum_{i=1}^{n}\frac{1}{x_i-\theta} + \frac{n}{\beta} = 0 \qquad (9)$$

The value of α, β, and θ cannot be estimated directly because of $\sum_{i=1}^{n}\frac{1}{x_i-\theta}$. The value of α, β, and θ can be calculated in a closed form without having to solve the nonlinear equations simultaneously [43]. The formula to estimate θ is

$$\theta = \sum_{i=1}^{n}\theta_i x_i, i = 1,2,3, \dots, n \qquad (10)$$

where,

$$\theta_1 = 1 + \left(1 - \frac{1}{n}\right)^n$$

$$\theta_i = \left(1 - \frac{1}{n}\right)^n - \left(1 - \frac{(i-1)}{n}\right)^n, i = 2,3,4, \dots, n$$

x represents the daily mean temperature value's first order statistics. The formula to estimate α and β are

$$\alpha = \frac{(\bar{x}-\theta)^2}{s^2} \qquad (11)$$

$$\beta = \frac{\bar{x}-\theta}{\alpha} \qquad (12)$$

where, $\bar{x}$ represents the average of daily mean temperature and s^2 represents the variance of daily mean temperature.

3.4. Validation Framework

The results of corrected temperature with QM for all clusters are compared to the observations using k-fold cross validation to reduce the effects of the training period selection and to ensure the assessment's robustness [44]. To calibrate the parameters, first use the period from 1976 to 1996 as the training period. Then, as a single experiment, correct the temperature biases for the remaining 10 years. The training period is then advanced by one year per time, and the BCM and validation are completed for the remaining years. This operation is performed triple times. Therefore, 20 continuous 30-year periods are chosen as training periods and 10 remaining years as validation periods. The root-mean-square error (RMSE) and mean absolute error (MAE) are used to determine the efficiency of corrected temperature with QM for all clusters (MAE). Furthermore, using the baseline period of 1976 to 1996 as a reference, the future temperatures for all clusters are projected from 2006 to 2100 under RCP4.5 and RCP8.5. RMSE and MAE are used to test the efficiency of all clusters for future temperature projections under RCP4.5 and RCP8.5.

3.5. Attributable Annual Deaths

The attributable annual deaths formula was used to project the impact of temperatures on CVD mortality rates in the future (2006-2100) [45-47],

$$AAD = y_0 \times ERC \times POP \tag{13}$$

where, AAD represents attributable annual deaths, y_0 represent the baseline annual mortality, ERC represents attributable change in mortality due to cardiovascular disease at each temperature, and POP represents the total population. The constant values of y_0 and POP were obtained from the National Cardiovascular Disease Database (NCVD) and DOSM in 2015, respectively. A distributed lag non-linear model (DLNM) was used to model the relationship between temperature and cardiovascular disease mortality [45-46]. DLNM has the advantage of capturing the lagged dependencies of exposure-response relationships by using two-function modelling (lag-response relationships and exposure-response, respectively) and complex non-linear. The exposure-response and lag-response relationships are then incorporated into the cross-basis function. In the exposure-response relationship, a quadratic B-spline is used, while in the lag response relationship, a natural cubic B-spline is used [46].

4. Results and Discussion

Table 3. The minimum value of Kelly-Gardner-Sutcliffe

Number of clusters	Value
4	22.2270
6	19.4703

Table 4. The list of 45 stations of 6 clusters throughout Peninsular Malaysia

Cluster	State	Station
1	Perlis	1,2,3,4,5
	Kedah	
	Penang	
2	Perak	6,7,8,9,10,11,12,13
3	Selangor	14,15,16,17,18,19,20,21,22,23
	Melaka	
	Negeri Sembilan	
4	Johor	24,25,26,27,28,29,45
5	Kelantan	30,31,32,33,34
	Terengganu (Station 34)	
6	Terengganu (Station 35)	35,36,37,38,39,40,41,42,43,44
	Pahang	

Figure 1 shows the cluster dendrogram for 45 stations throughout Peninsular Malaysia. Ward's method has advantages when the determination of cluster number is less sensitive compared to the k-means method [6]. However, a different number of clusters could give different results. Hence, the optimal number of clusters could give more reliable results using the Ward's method [18]. Based on Table 3, six clusters have a minimum value of Kelly-Gardner-Sutcliffe which is 19.4703 compared to four clusters which is 22.2270. Thus, six clusters are the optimum number of clusters that will be selected in this study. Table 4 lists the number of stations for each cluster throughout Peninsular Malaysia.
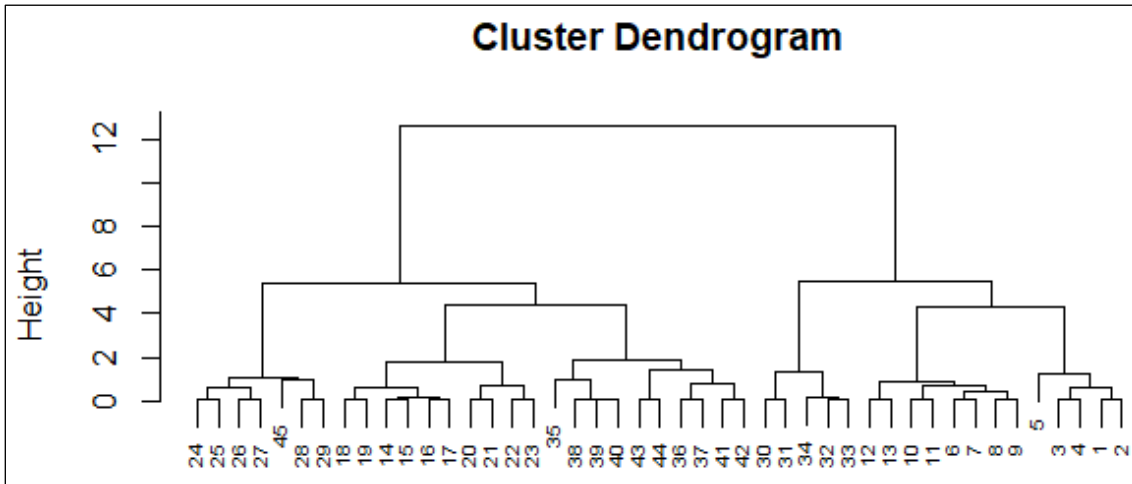
Figure 1. The cluster dendrogram for 45 stations throughout Peninsular Malaysia using Ward's method

Table 5. The descriptive statistics for observations daily mean temperature of 6 clusters throughout Peninsular Malaysia

Cluster	Descriptive	Value
1	Mean	25.7665
	Median	25.7299
	Skewness	0.1436
	Kurtosis	0.2906
2	Mean	26.9084
	Median	26.8880
	Skewness	0.1075
	Kurtosis	-0.0570
3	Mean	24.7683
	Median	24.7625
	Skewness	0.0019
	Kurtosis	-0.0114
4	Mean	26.9084
	Median	26.8880
	Skewness	0.0843
	Kurtosis	0.1173
5	Mean	26.1734
	Median	26.1602
	Skewness	0.1082
	Kurtosis	0.0757
6	Mean	26.2549
	Median	26.2449
	Skewness	0.1219
	Kurtosis	-0.0168

Table 5 shows the descriptive statistics for the observed daily mean temperature of 6 clusters throughout Peninsular Malaysia. The result shows that the value of mean is greater than the value of median which indicates that observed daily mean temperature for all clusters is positively skewed. It is also supported by the positive value of skewness. Thus, Gamma distribution with three parameters is used to fit the observed daily mean temperature for all clusters throughout Peninsular Malaysia. The value of kurtosis for cluster 1, 4, and 5 are positive while in contrast, the value of kurtosis for cluster 2, 3, and 6 show negative values. The positive value of kurtosis represents indicate that a distribution is peaked and possess thick tails while the negative value of kurtosis indicates flatter distribution.

Table 6. The correlation coefficient between and within clusters across Peninsular Malaysia

Correlation coefficient between clusters	Correlation coefficient within clusters
0.0259	0.9182

Table 6 shows the correlation coefficient between and within clusters across Peninsular Malaysia. The correlation coefficient within clusters is higher than the correlation coefficient between clusters. This indicates that there is no strong relationship between clusters while there is a strong relationship within clusters.

Table 7 shows the statistical analysis of observations, historical GCM, and corrected daily mean temperature of all clusters throughout Peninsular Malaysia after QM. The results show that the mean and variance values for all clusters are almost identical to the observations. Closer values of RMSE and MAE to zero indicate the smallest difference between observation and corrected daily mean temperature, indicating better simulation accuracy. Cluster 5, cluster 6, cluster 3, cluster 1, and cluster 4 have lower RMSE and MAE values than cluster 2, cluster 6, cluster 3, cluster 1, and cluster 4. In comparison to the other clusters, this shows that QM provides reliable results for cluster 5 [48].

Table 7. Statistical analysis of observations, historical GCM, and corrected daily mean temperature of all clusters throughout Peninsular Malaysia after QM

Cluster	Data	Frequency statistics (Daily)		Time series statistics (Daily)	
		Mean	Variance	RMSE	MAE
	Historical GCM	25.7665	1.2697	-	-
1	Observed Temperature	27.3949	1.0008	0.5171	0.2674
	Corrected Temperature	27.9117	1.0390		
2	Observed Temperature	24.2317	0.8453	0.2049	0.0420
	Corrected Temperature	24.4364	0.8596		
3	Observed Temperature	26.9084	0.9992	0.4422	0.1956
	Corrected Temperature	27.3504	1.0323		
4	Observed Temperature	26.8266	0.9972	0.6626	0.4390
	Corrected Temperature	27.4887	1.0471		
5	Observed Temperature	26.7451	1.0649	0.1924	0.0370
	Corrected Temperature	26.5528	1.0497		
6	Observed Temperature	26.5305	1.0627	0.4372	0.1911
	Corrected Temperature	26.9673	1.0980		

Table 8. The RMSE and MAE of yearly corrected temperature with QM for all clusters throughout Peninsular Malaysia compared to observations from 1976 to 2005.

Cluster	Data	K-fold cross validation	
		RMSE	MAE
1	Corrected Temperature	0.00004	0.00004
2	Corrected Temperature	0.00005	0.00004
3	Corrected Temperature	0.00007	0.00006
4	Corrected Temperature	0.00005	0.00004
5	Corrected Temperature	0.00005	0.00004
6	Corrected Temperature	0.00007	0.00006

Table 8 shows the RMSE and MAE of yearly corrected temperature with QM for all clusters throughout Peninsular Malaysia compared to observations from 1976 to 2005. The value of RMSE and MAE for all clusters are less zero. This indicates that the average of the 20 experiments shows that the QM can reduce biases in the temperature of GCM significantly.

Table 9. The RMSE and MAE of projection of future temperature of all clusters throughout Peninsular Malaysia under RCP4.5 and RCP8.5 scenarios

Cluster	Data	K-fold cross validation	
		RMSE	MAE
1	RCP4.5	0.0000340	0.0000288
	RCP8.5	0.0000309	0.0000254
2	RCP4.5	0.0000338	0.0000286
	RCP8.5	0.0000298	0.0000255
3	RCP4.5	0.0000327	0.0000277
	RCP8.5	0.0000347	0.0000303
4	RCP4.5	0.0000327	0.0000284
	RCP8.5	0.0000316	0.0000267
5	RCP4.5	0.0000311	0.0000269
	RCP8.5	0.0000323	0.0000278
6	RCP4.5	0.0000345	0.0000306
	RCP8.5	0.0000311	0.0000266

Table 9 shows the RMSE and MAE of projection of future temperature of all clusters throughout Peninsular Malaysia under RCP4.5 and RCP8.5 scenarios. For both clusters, the RMSE and MAE values are less than zero. This means that the projection of future temperature of all clusters under the RCP4.5 and RCP8.5 scenarios across Peninsular Malaysia are reliable.

Table 10 shows the estimated parameter values of the shape $\hat{\alpha}$, scale $\hat{\beta}$ and location $\hat{\theta}$ for corrected temperature for all clusters in Peninsular Malaysia and future GCM under RCP4.5 and RCP8.5. All the values of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$ are highly significant based on 95% confidence interval.

Table 10. The estimated parameter values of the shape $\hat{\alpha}$, scale $\hat{\beta}$, and location $\hat{\theta}$ for corrected temperature of 6 clusters in Peninsular Malaysia and future GCM under RCP4.5 and RCP8.5.

| Cluster	Data	Parameter	Estimated value (SE)	95% CI
1	Corrected values	α	2602.6467 (24.8719)	(2602.1808, 2603.1126)
		β	0.0203 (0.0002)	(0.02029, 0.0203)
		θ	-24.9117 (-0.2228)	(-24.9075, -24.9159)
2	Corrected values	α	2393.5693 (22.8738)	(2393.1409, 2393.9977)
		β	0.0190 (0.0002)	(0.01899, 0.0190)
		θ	-20.9238 (-0.2000)	(-20.9201, -20.9275)
3	Corrected values	α	2486.3362 (23.7603)	(2485.8912, 2486.7812)
		β	0.0204 (0.0002)	(0.02039, 0.0204)
		θ	-23.3117 (-0.2228)	(-23.3075, -23.3159)
4	Corrected values	α	2460.3256 (23.5118)	(2459.8852, 2460.7660)
		β	0.0206 (0.0002)	(0.02059, 0.0206)
		θ	-23.2668 (-0.2223)	(-23.2626, -23.2710)
5	Corrected values	α	2292.6727 (21.9096)	(2292.2623, 2293.0831)
		β	0.0214 (0.0002)	(0.02139, 0.0214)
		θ	-22.5036 (-0.2151)	(-22.4996, -22.5076)
6	Corrected values	α	2293.9304 (21.9216)	(2293.5198, 2294.3410)
		β	0.0219 (0.0002)	(0.02189, 0.0219)
		θ	-23.2204 (-0.2219)	(-23.2162, -23.2246)
-	Future GCM (RCP4.5)	α	1744.1813 (16.6680)	(1743.8691, 1744.4935)
		β	0.02767 (0.0003)	(0.02767, 0.02768)
		θ	-21.4962 (-0.2054)	(-21.4923, -21.5000)
-	Future GCM (RCP8.5)	α	1355.9014 (12.9575)	(1355.6587, 1356.1441)
		β	0.03637 (0.0003)	(0.03637, 0.03638)
		θ	-22.0165 (-0.2104)	(-22.0126, -22.0204)
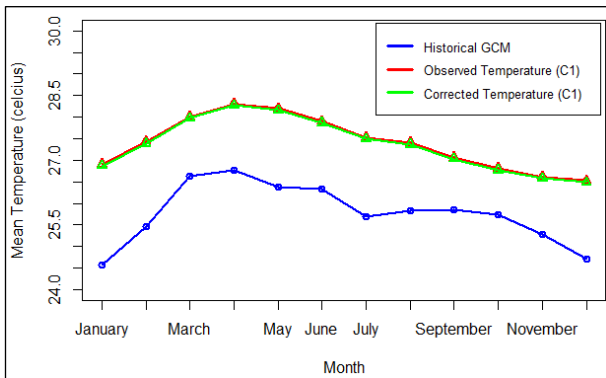
(a)
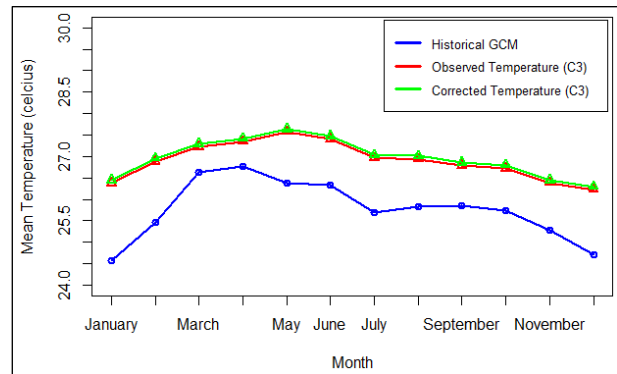
(c)
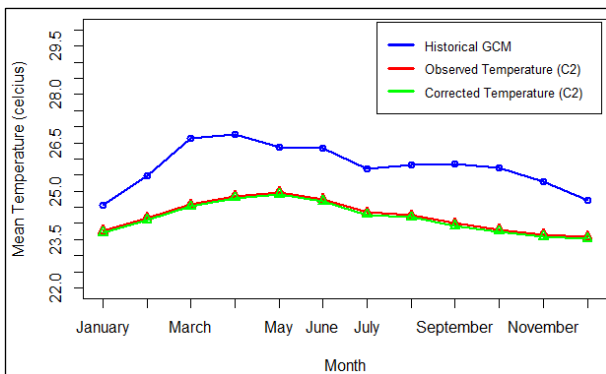
(b)
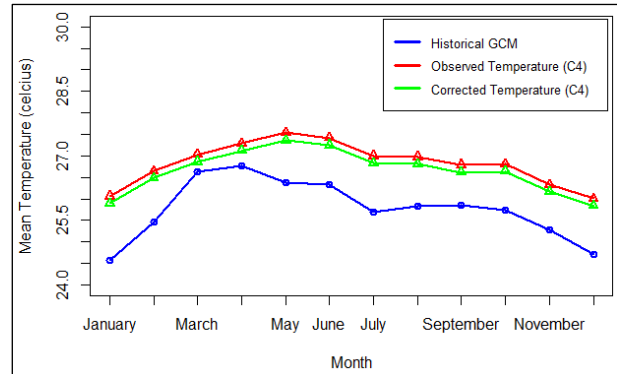
(d)
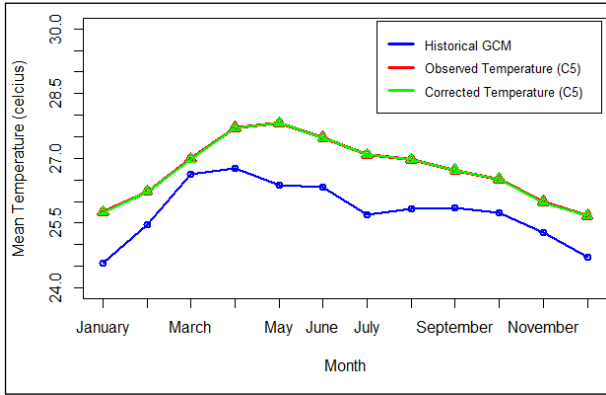
(e)
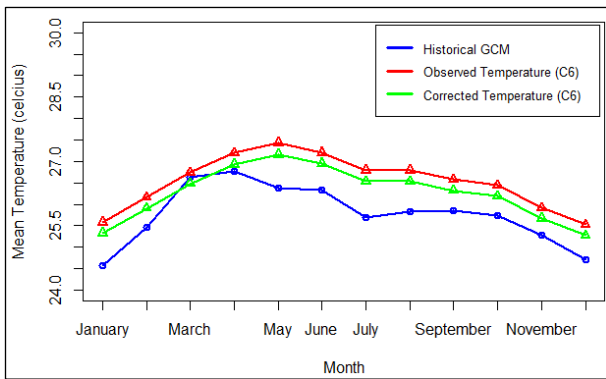
(f)

Figure 2. The monthly mean temperature of historical GCM, observed and corrected temperature for (a) 1st cluster, (b) 2nd cluster, (c), 3rd cluster, (d) 4th cluster, (e) 5th cluster, and (f) 6th cluster throughout Peninsular Malaysia.


(a)

(b)

(c)

(d)

Figure 2 shows the monthly mean temperature for historical GCM, observed and corrected temperature of all clusters throughout Peninsular Malaysia. The monthly mean temperature recorded the lowest value of monthly mean temperature in January and December. This is because November and December associate with the northeast monsoon wet seasons when the weather is cooler than in other seasons. Meanwhile, the highest value of monthly mean temperature was recorded in May. Peninsular Malaysia experiences southwest monsoon rainfall between March and October where this area receives minimal rainfall during this period and is considered as a dry period [49]. Different ranges of mean temperature were observed in different clusters due to the climate characteristic differences. The delineated temperature zones are closely linked to geography and topographic characteristics [6, 23].
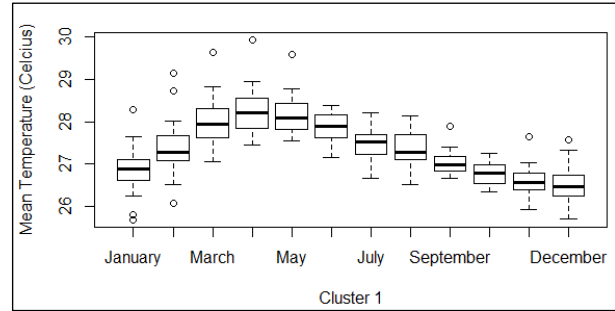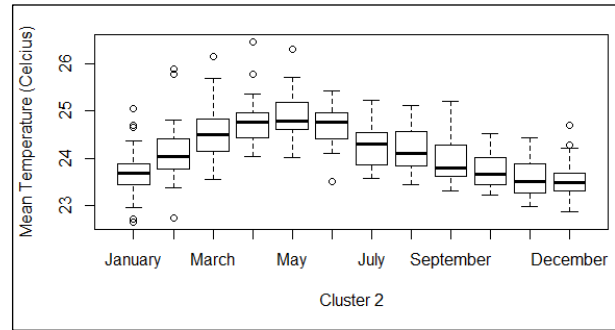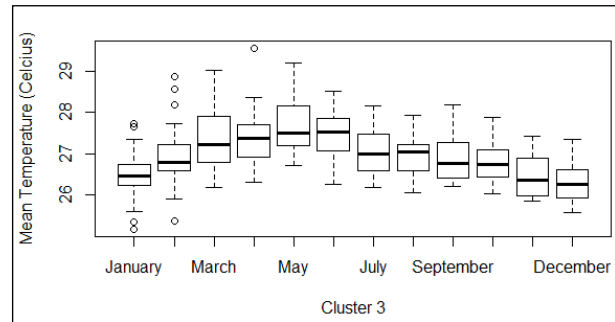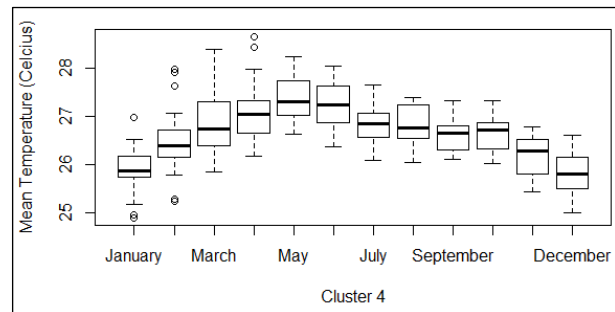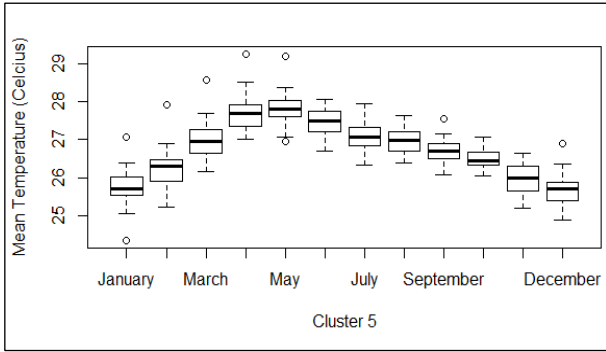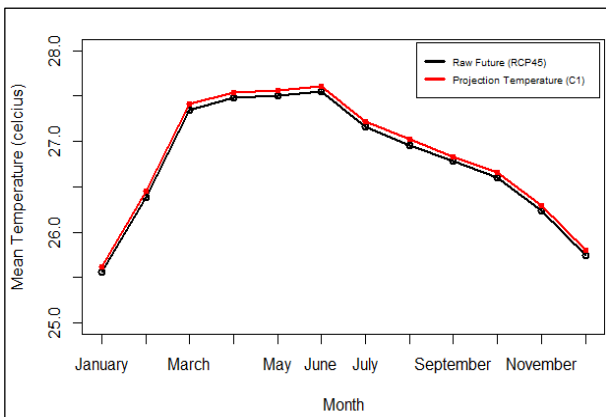
(e)
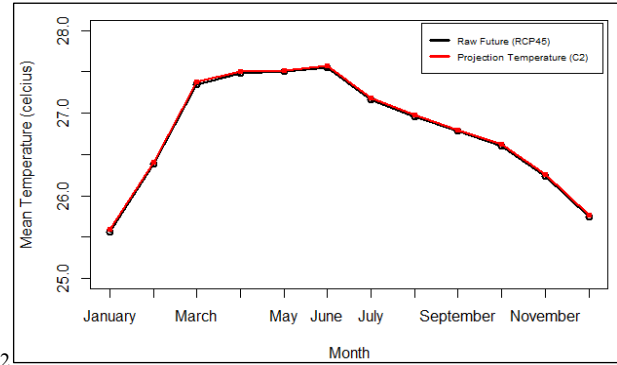
(b)
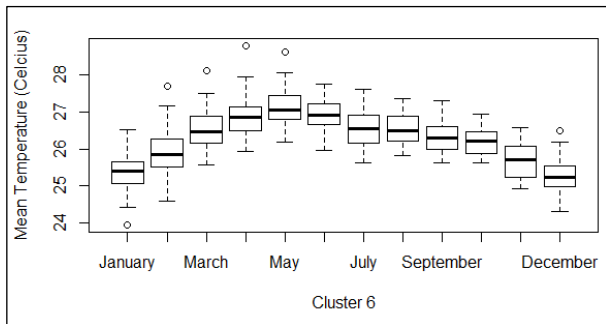
(f)
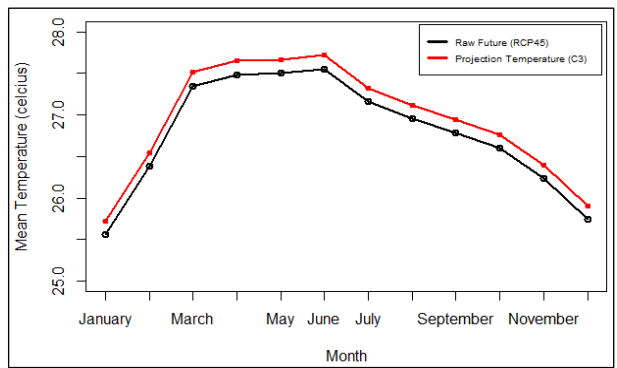
(c)

Figure 3. The boxplot of mean temperature from January until December for (a) 1st cluster, (b) 2nd cluster, (c), 3rd cluster, (d) 4th cluster, (e) 5th cluster, and (f) 6th cluster throughout Peninsular Malaysia.

Figure 3 shows the boxplot of mean temperature from January until December for all clusters. In contrast to the second halves of the year, the outliers are more prominent in the first half of the year. In general, the shape of the mean temperature distribution for all clusters is positively skewed between January and June, indicating that the weather is slightly warmer during this period. However, the weather becomes somewhat cooler towards the end of the year.
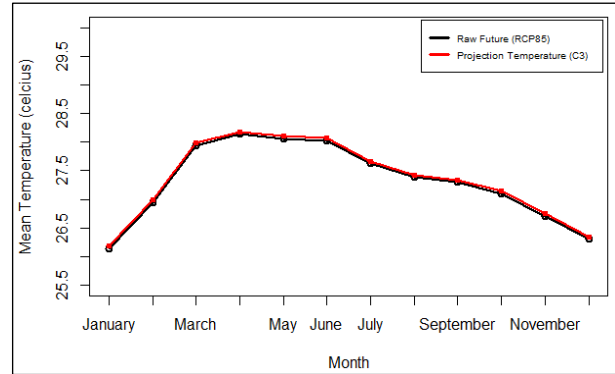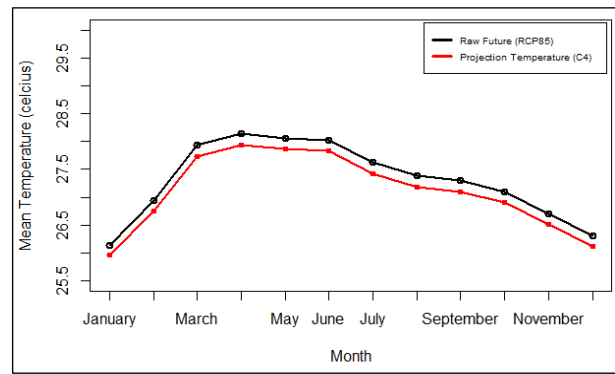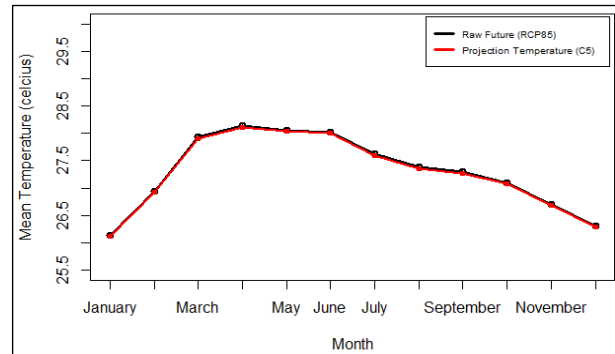

(d)

(a)

(e)

(f)
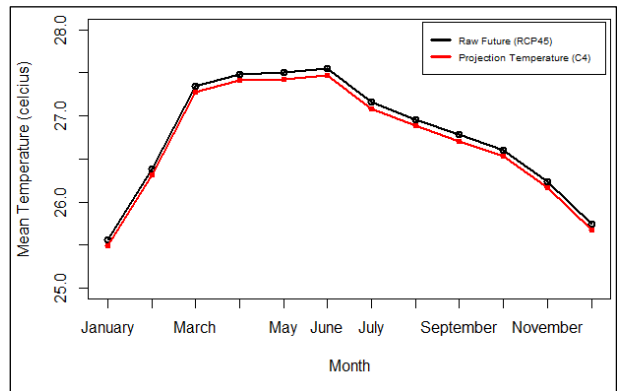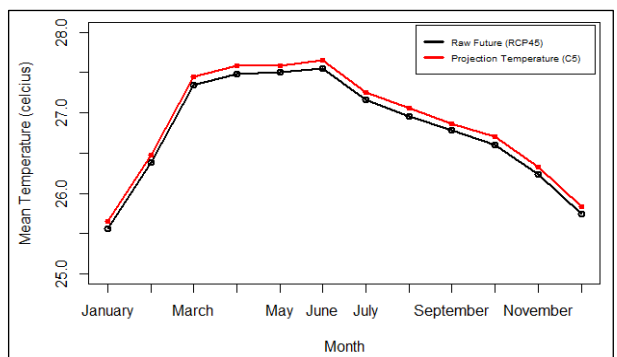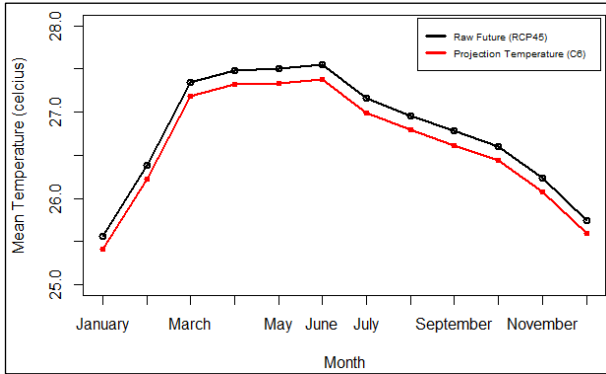
Figure 4. The yearly mean temperature of raw future GCM and corrected temperature (a) 1ˢᵗ cluster, (b) 2ⁿᵈ cluster, (c), 3ʳᵈ cluster, (d) 4ᵗʰ cluster, (e) 5ᵗʰ cluster, and (f) 6ᵗʰ cluster throughout Peninsular Malaysia under RCP4.5.
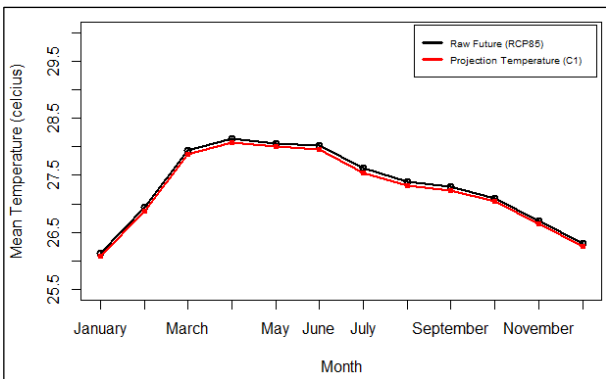
Figure 4 shows the monthly mean temperature of raw future GCM and corrected temperature of all clusters under RCP4.5. The projection of temperature for cluster 1, 3, and 5 is higher than the raw future temperature whereas the projection of temperature for cluster 4 and 6 is lower than the raw future temperature. Nevertheless, the projection of temperature for cluster 2 is similar to raw future temperature. All clusters recorded the highest value of monthly mean temperature in June while the lowest value of monthly mean temperature between December and January.
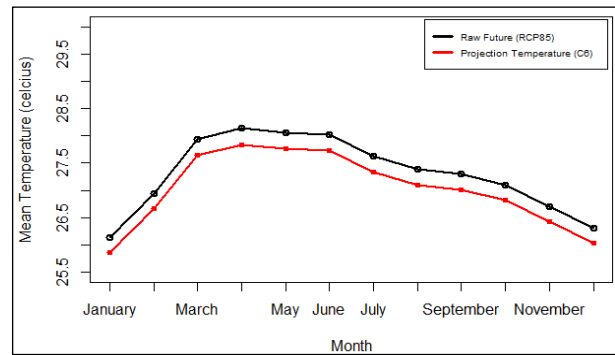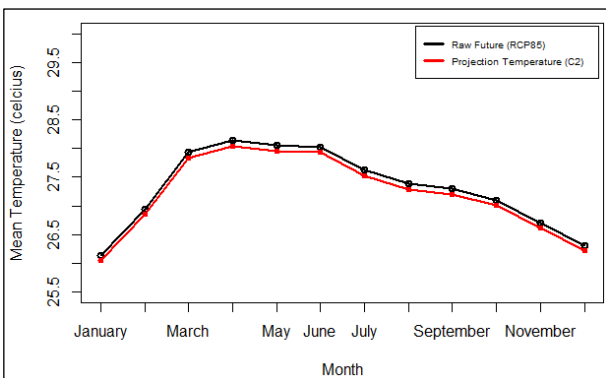

(a)
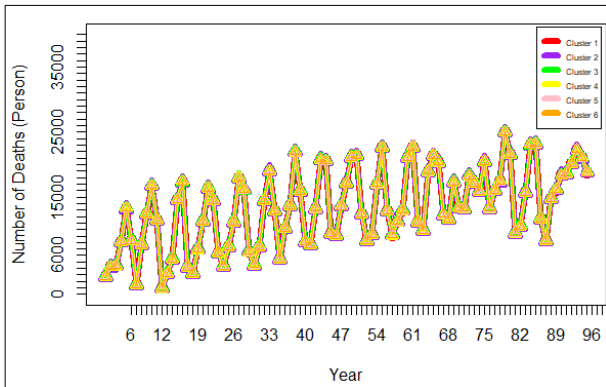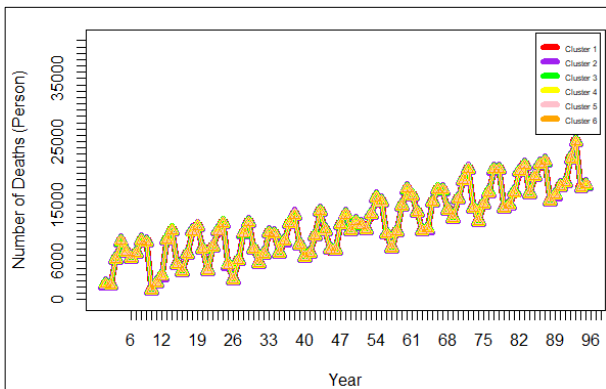

(b)

(c)

(d)

(e)

(f)

Figure 5. The yearly mean temperature of raw future GCM and corrected temperature for (a) 1ˢᵗ cluster, (b) 2ⁿᵈ cluster, (c), 3ʳᵈ cluster, (d) 4ᵗʰ cluster, (e) 5ᵗʰ cluster, and (f) 6ᵗʰ cluster throughout Peninsular Malaysia under RCP8.5.

Figure 5 shows the monthly mean temperature of raw future GCM and corrected temperature of all clusters throughout Peninsular Malaysia under RCP8.5. The projection of temperature for cluster 3 is slightly higher than the raw future temperature whereas the projection of temperature for cluster 1, 2, 4, and 6 is lower than raw future temperature. The temperature projection for clusters 5 is close to the raw future temperature. The highest monthly mean temperature was recorded in June in all clusters, while the lowest monthly mean temperature was recorded between December and January in all clusters.

Table 11. The increment projection of temperature under RCP4.5 and RCP8.5 for 6 clusters throughout Peninsular Malaysia

Cluster	Increment	RCP4.5	Increment	RCP8.5
1		0.1269		0.2604
2		0.1270		0.2607
3	0.05℃	0.1265	0.1℃	0.2592
4		0.1276		0.2616
5		0.1268		0.2599
6		0.1278		0.2624

Table 11 shows the increment projection of temperature under RCP4.5 and RCP8.5 for 6 clusters throughout Peninsular Malaysia. The increment projection of temperature under RCP4.5 by 0.05 ℃ of all clusters whereas the increment projection of temperature under RCP8.5 by 0.1℃. The projection of temperature under RCP4.5 for cluster 5 has higher ratio while the projection of temperature under RCP4.5 for cluster 1 has lower ratio. Meanwhile, the projection of temperature under RCP8.5 for all clusters have similar ratio except for cluster 1.

(a)

(b)

Figure 6. The projection of number of deaths for all clusters throughout Peninsular Malaysia under (a) RCP4.5 and (b) RCP8.5 between 2006 and 2100

Figure 6 shows the projection of the number of deaths for all clusters throughout Peninsular Malaysia under RCP4.5 and RCP8.5 between 2006 and 2100. For both clusters, the projected number of deaths is usually the same. Between 2006 and 2100, the number of deaths projected under RCP4.5 and RCP8.5 are fluctuating. From 2006 to 2100, the number of deaths under RCP4.5 and RCP8.5 had increased. However, as compared to the number of deaths projected under RCP4.5, the number of deaths projected under RCP8.5 is higher.

5. Conclusion

In conclusion, Ward's method could obtain the optimal number of clusters in Peninsular Malaysia based on the Kelly-Gardner-Sutcliffe. QM was able to fit the temperature series as well as eliminate biases in the mean and variability, resulting in an accurate prediction of heat-related CVD. RCP8.5 has a higher CVD mortality estimate than RCP4.5. This is because RCP8.5 expect global temperature to rise of about 4.3°C whereas RCP 4.5 delivers global temperature increase around 2°C to 3°C. However, there are a few caveats to this study that must be addressed. To begin with, this study only corrects temperature in the present study, ignoring any other biases from different sources such as sample size selection and other potential sources of bias. Second, the population is projected to increase in the future. Finally, other causes of CVD and different categories of patients, such as gender, age, and race, must be examined in a future study in order to achieve more meaningful outcomes. Many of these issues are essential for further study.

Acknowledgements

REFERENCES

[1] Watanabe M., Suzuki T., O'ishi R., Komuro Y., Watanabe S., Emori S., Takemura T., Chikira M., Ogura T., Sekiguchi M., Takata K., "Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity," Journal of Climate, vol. 23, no. 23, pp. 6312-6335, 2010.

[2] Khan N., Shahid S., Ahmed K., Ismail T., Nawaz N., Son M., "Performance assessment of general circulation model in simulating daily precipitation and temperature using multiple gridded datasets," Water, vol. 10, no. 12, pp. 1793, 2018.

[3] Luo M., Liu T., Meng F., Duan Y., Frankl A., Bao A., De Maeyer P., "Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: A case study from the Kaidu River Basin in Western China," Water, vol. 10, no. 8, pp. 1046, 2018.

[4] Teutschbein C., Seibert J., "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," Journal of Hydrology, vol. 456, pp. 12-29, 2012.

[5] Fang G., Yang J., Chen Y. N., Zammit C., "Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China," Hydrology and Earth System Sciences, vol. 16, no. 6, pp. 2547-2559, 2015.

[6] Singh S. K., Lo E. Y. M., Qin X., "Cluster Analysis of monthly precipitation over the western maritime continent under climate change," Climate, vol. 5, no. 4, pp. 84, 2017.

[7] Supriyadi S., Hidayati R., Hidayat R. Sopaheluwakan A., "Mapping extreme rain conditions in Sumatra by influence global conditions," IOP Conference Series: Earth and Environmental Science, March., 2017, vol. 58, no. 1, pp. 012041.

[8] Stocker T., "Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change," in Current Situation and Development of Contemporary Education, Cambridge university press, 2014, pp. 1029-1137.

[9] Tangang F.T., Juneng L., Ahmad S., "Trend and interannual variability of temperature in Malaysia: 1961–2002," Theoretical and Applied Climatology, vol. 89, no. 3, pp. 127-141, 2007.

[10] Suparta W., Yatim A.N.M., "An analysis of heat wave trends using heat index in East Malaysia," Journal of Physics: Conference Series, May., 2017, vol. 852, no. 1, pp. 012005.

[11] Suparta W., Yatim A.N.M., "Characterization of Heat Waves: A Case Study for Peninsular Malaysia," Geographia Technica, vol. 14, no. 1, 2019.

[12] Tang K.H.D., "Climate change in Malaysia: Trends, contributors, impacts, mitigation and adaptations," Science of the Total Environment, vol. 650, pp. 1858-1871, 2019.

[13] Khoo K.L., Tan H., Khoo T.H., "Cardiovascular mortality in peninsular Malaysia: 1950-1989," Med J Malaysia, vol. 46, no. 1, pp. 7-20, 1991.

[14] Malaysia, D. S., "Statistics Causes of Deaths in 2017," Department of Statistics, Malaysia, 2017.

[15] Malaysia, D. S., "Statistics Causes of Deaths in 2018," Department of Statistics, Malaysia, 2018.

[16] Malaysia, D. S., "Statistics Causes of Deaths in 2019," Department of Statistics, Malaysia, 2019.

[17] Malaysia, D. S., "Statistics Causes of Deaths in 2020," Department of Statistics, Malaysia, 2020.

[18] Jolliffe I.T., Philipp A., "Some recent developments in cluster analysis," Physics and Chemistry of the Earth, Parts A/B/C, vol. 35, no. 9-12, pp. 309-315, 2010.

[19] Baltaci H., Alemdar C.S.O., Akkoyunlu B.O., "Background atmospheric conditions of high PM10 concentrations in Istanbul, Turkey," Atmospheric Pollution Research, vol. 11, no. 9, pp. 1524-1534, 2020.

[20] Unal Y., Kindap T., Karaca M., "Redefining the climate zones of Turkey using cluster analysis," International Journal of Climatology: A Journal of the Royal Meteorological Society, vol. 23, no. 9, pp. 1045-1055, 2003.

[21] Gómez-Navarro J.J., Raible C.C., Bozhinova D., Martius O., García Valero J.A., Montávez J.P., "A new region-aware bias correction method for simulated precipitation in the Alpine region," Geoscientific Model Development, vol. 11, no. 6, pp. 2231-2247, 2018.

[22] Ringard J., Seyler F., Linguet L., "A quantile mapping bias correction method based on hydroclimatic classification of the Guiana shield," Sensors, vol. 16, no. 6, pp. 1413, 2017.

[23] Wong C. L., Liew J., Yusop Z., Ismail T., Venneker R., Uhlenbrook S., "Rainfall characteristics and regionalization in Peninsular Malaysia based on a high-resolution gridded data set," Waters, vol. 8, no. 11, pp. 500, 2016.

[24] Jakob Themeßl M., Gobiet A., Leuprecht A., "Empirical‐statistical downscaling and error correction of daily precipitation from regional climate models," International Journal of Climatology, vol. 31, no. 10, pp. 1530-1544, 2011.

[25] Thrasher B., Maurer E. P., McKellar C., Duffy P. B., "Bias correcting climate model simulated daily temperature extremes with quantile mapping," Hydrology and Earth System Sciences, vol. 16, no. 9, pp. 3309-3314, 2012.

[26] Chen J., Brissette F. P., Chaumont D., Braun M., "Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America," Water Resources Research, vol. 49, no. 7, pp. 4187-4205, 2013.

[27] Lafon T., Dadson S., Buys G., Prudhomme C., "Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods," International Journal of Climatology, vol. 33, no. 6, pp. 1367-1381, 2013.

[28] Olsson T., Jakkila J., Veijalainen N., Backman L., Kaurola J., Vehviläinen B., "Impacts of climate change on temperature, precipitation and hydrology in Finland–studies using bias corrected Regional Climate Model data," Hydrology and Earth System Sciences, vol. 19, no. 7, pp. 3217-3238, 2015.

[29] Ajaaj A. A., Mishra A. K., Khan A. A., "Comparison of BIAS correction techniques for GPCC rainfall data in semi-arid climate," Stochastic environmental research and risk assessment, vol. 30, no. 6, pp. 1659-1675, 2016.

[30] Luo M., Liu T., Meng F., Duan Y., Frankl A., Bao A., De

Maeyer P., "Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: A case study from the Kaidu River Basin in Western China," Water, vol. 10, no. 8, pp. 1046, 2018.

[31] Tadese M. T., Kumar L., Koech R., "Climate change projections in the Awash River basin of Ethiopia using global and regional climate models," International Journal of Climatology, vol. 40, no. 8, pp. 3649-3666, 2020.

[32] Piani C., Haerter J. O., Coppola E., "Statistical bias correction for daily precipitation in regional climate models over Europe," Theoretical and Applied Climatology, vol. 99, no. 1, pp. 187-192, 2010.

[33] Sennikovs J., Bethers U., "Statistical downscaling method of regional climate model results for hydrological modelling," 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, July., 2009, pp. 3962-3968.

[34] Gudmundsson L., Bremnes J. B., Haugen J. E., Engen-Skaugen T., "Downscaling RCM precipitation to the station scale using statistical transformations–a comparison of methods," Hydrology and Earth System Sciences, vol. 16, no. 9, pp. 3383-3390, 2012.

[35] Wilks D. S., "Statistical methods in the atmospheric sciences (Vol. 100)," 3rd ed, Academic press, 2011, pp. 85-110.

[36] Iizumi T., Takikawa H., Hirabayashi Y., Hanasaki N., Nishimori M., "Contributions of different bias‐correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes," Journal of Geophysical Research: Atmopheres, vol. 122, no. 15, pp. 7800-7819, 2017.

[37] Pastén-Zapata E., Jones J. M., Moggridge H., Widmann, M., "Evaluation of the performance of Euro-CORDEX Regional Climate Models for assessing hydrological climate change impacts in Great Britain: A comparison of different spatial resolutions and quantile mapping bias correction methods," Journal of Hydrology, vol. 584, pp. 124653, 2020.

[38] Wayne G. P., "The beginner's guide to representative concentration pathways". Skeptical science, vol. 24, 2014.

[39] Ahmad W. A. W., Kui-Hian S., "Annual report of the acute coronary syndrome (acs) registry 2007-2008," National Cardiovascular Disease Database, 2008.

[40] Ahmad W. A. W., Kui-Hian S., "Annual report of the acute coronary syndrome (acs) registry 2009-2010," National Cardiovascular Disease Database, 2010.

[41] Ahmad W. A. W., Kui-Hian S., "Annual report of the acute coronary syndrome (acs) registry 2011-2013," National Cardiovascular Disease Database, 2013.

[42] Ahmad W. A. W., Kui-Hian S., "Annual report of the acute coronary syndrome (acs) registry 2014-2015," National Cardiovascular Disease Database, 2015.

[43] Kappenman R. F., "Estimation for the three-parameter Weibull, lognormal, and gamma distributions," Computational Statistics & Data Analysis, vol. 3, pp. 11-23, 1985.

[44] Yang X., Wood E. F., Sheffield J., Ren L., Zhang M., Wang Y., "Bias correction of historical and future simulations of precipitation and temperature for China from CMIP5 model," Journal of Hydrometeorology, vol. 19, no. 3, pp. 609-623, 2018.

[45] Gasparrini A., Armstrong B., Kenward M. G., "Distributed lag non-linear models," Statistics in medicine, vol. 29, no. 21, pp. 2224-2234, 2010.

[46] De'Donato F.K., Leone M., Scortichini M., De Sario M., Katsouyanni K., Lanki T., Basagaña X., Ballester F., Åström C., Paldy A., Pascal M., "Changes in the effect of heat on mortality in the last 20 years in nine European cities. Results from the PHASE project," International journal of environmental research and public health, vol. 12, no. 12, pp. 15567-15583, 2015.

[47] Chen K., Fiore A.M., Chen R., Jiang L., Jones B., Schneider A., Peters A., Bi J., Kan H., Kinney P.L., "Future ozone-related acute excess mortality under climate and population change scenarios in China: A modeling study," Plos medicine, vol. 15, no. 7, pp. 1002598, 2018.

[48] Dutta P.N., Karlo T. and Dutta P., "Some Features of Surface Air Temperature: A Statistical Viewpoint," Environ. Ecol. Res., vol. 5, no. 5, pp. 367-376, 2017. DOI: 10.13189/eer.2017.050506

[49] Syafrina A. H., Zalina M. D., Juneng L., "Historical trend of hourly extreme rainfall in Peninsular Malaysia," Theoretical and Applied Climatology, vol. 120, no. 1-2, pp. 259-285, 2015.