ADV MATH SCI JOURNAL

Advances in Mathematics: Scientific Journal **9** (2020), no.12, 10903–10913
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.12.74 Special Issue SMS-2020

QUANTILE MAPPING WITH THREE PARAMETERS OF GAMMA DISTRIBUTION FOR FUTURE PROJECTION OF CARDIOVASCULAR DISEASE MORTALITY RATE

M.E. AINA IZZATI[1] AND A.H. SYAFRINA

ABSTRACT. Global Climate Model (GCM) outputs are particularly useful for the projection of future climate changes, but they exhibit large systematic biases relative to observational datasets. One of the bias correction methods (BCM) namely quantile mapping (QM) technique is widely used to capture the evolution of the mean and variability of a GCM. Gamma distribution with shape and scale parameters is frequently applied in QM. However, including the location parameter in the Gamma distribution could better illustrate the data. Hence, the objectives of this study are (i) to develop the QM with shape, scale and location parameters of Gamma distribution to correct the biases of temperature series, (ii) to project the corrected future temperature series and (iii) to calculate the future mortality rate of cardiovascular disease (CVD) using the developed model. Results have shown that QM with shape, scale and location parameters was able to remove the biases and capture the variability and the mean of temperature very well. The average corrected values of future daily mean temperature are projected to be higher. Meanwhile, the number of deaths for the male group is higher than the number of deaths for the female group from 2006-2100.

1. INTRODUCTION

Extreme heat events are the most prominent cause of weather-related human mortality. There were 200 reported cases related to hot weather which is 52 cases

[1]*corresponding author*

2020 *Mathematics Subject Classification.* 18F60.

Key words and phrases. bias correction method, cardiovascular disease, gamma distribution, mortality, quantile mapping.

of heat cramps, 126 cases of heat exhaustion, 22 cases of heat stroke includes 2 deaths of heat stroke [1]. It is expected that extreme heat events will occur more frequently and will be of longer duration due to the ongoing climate change. Northern China showed an increase of 48% to 74% by 2050 and 69% to 134% by 2080 in the annual heat-related cardiovascular disease (CVD) deaths from May until September [2]. These increases support the projection in the Fifth Intergovernmental Panel on Climate Change (IPCC) report, which predicted increases in global mean surface temperature during 2046–2060 relative to 1986–2005 in the ranges of 0.4°C to 1.6°C (RCP2.6), 0.9°C to 2.0°C (RCP4.5), and 1.4°C to 2.6°C (RCP8.5). The annual number of temperature-related CVD deaths projected by the RCP2.6, RCP4.5, and RCP8.5 exhibited a gradually increasing trend before 2050. In the late 21st century, the number of temperature-related CVD deaths is projected to increase under the RCP4.5 and RCP8.5 while the number of temperature-related CVD deaths is projected to remain constant and slightly decreasing under the RCP2.6 due to lower emission of greenhouse gases.

2. Cardiovascular Disease

CVD is a group of disorders of the heart and blood vessels. Several risk factors may influenced CVD. Aging increases the risk of damaged by narrowing the arteries and weakened or thickened the heart muscle. Gender factor also contributes to the risk of CVD where males are generally at greater risk of heart disease while females are at greater risk after menopause. A family history also plays an important role where a family history of heart disease increases the risk of coronary artery disease in the family member. Heart attacks are found to be more common for smokers than nonsmokers due to the high level of nicotine contained in cigarettes. People who is taking unhealthy food are also prone to get heart disease because of their diet that contains high levels of fat, salt, sugar, and cholesterol which can lead to the growth of heart disease. Furthermore, the uncontrolled high blood pressure could also cause arteries to harden and thicken, and the vessels through which blood flows to widen.

One of the CVDs namely coronary heart disease (CHD) is a leading killer disease in Malaysia. According to the Department of Statistics Malaysia (DOSM), CHD was the principal causes of death in 2016 with 13.2%, followed by pneumonia with 12.5%, cerebrovascular diseases with 6.9%, transport accidents with 5.4%

and malignant neoplasm of trachea, bronchus, and lung with 2.2% [3]. In 2017, the percentage of deaths caused by CHD was 13.9%, followed by pneumonia with 12.7%, cerebrovascular diseases with 7.1%, transport accidents with 4.6%, and malignant neoplasm of trachea, bronchus, and lung with 2.3% [4]. CHD still remained as the principal cause of death with 15.6%, followed by pneumonia with 11.8%, cerebrovascular diseases with 7.8%, transport accidents with 3.7%, and chronic lower respiratory diseases with 2.6% in 2018 [5].

The number of deaths for males due to CHD recorded a higher percentage compared to other diseases in 2016, 2017, and 2018. In particular, the percentage of CHD deaths for male group is 15.3% in 2016. The percentage becomes higher in the next two years with 16.0% and 17.8%, respectively [3–5]. On the other hand, the percentage of CHD deaths for female group ranked second. In particular, the percentage of deaths for females due to pneumonia is 14.0%, followed by CHD with 9.9% in 2016. In 2017, the percentage of deaths for females due to pneumonia is 14.1%, followed by CHD with 10.5% and in 2018, the percentage of deaths for females due to pneumonia is 12.8%, followed by CHD with 12.2% [3–5].

3. Global Climate Model

The IPCC is the United Nations body that was established to assess the science related to climate change. The IPCC was created to provide the policymakers with regular scientific assessments on climate change, its implication, and potential future risks as well as to put forward adaptation and mitigation options. However, the resolution of GCMs are too coarse for climate change impact assessment and require downscaling to obtain finer resolutions. There are two types of downscaling which are statistical downscaling and dynamical downscaling. Statistical downscaling is a process consisting of the development of statistical relationships between local climate variables (predictands) and large-scale atmospheric variables (predictors). Statistical downscaling have advantages whereby they are often much less computationally, able to capture bias and has a potential to produce outputs at finer resolution compared to dynamical downscaling [6].

4. Bias Correction Method

Bias correction method (BCM) employs a transfer function to correct the projected raw daily GCM output using the differences in the mean and variability between GCM and observations in a reference period [6]. The advantages of BCMs includes the corrections of algorithm and its parameterization for current climate conditions that are valid for future conditions [7]. Quantile mapping (QM) is one of the BCMs. QM matches the statistical distribution function of the raw data to the statistical distribution function of observational data. It is used to adjust mean, standard deviation, quantiles, and preserves the extremes. The daily mean temperature in Selangor is assumed to follow Gamma distribution for both observations and GCM data because of the shape of the distribution which is positively skewed [8]. Identifying the suitable distribution for the random variables is essential for the accurate evaluation of the reliability of a structure. Addition of the location parameter could better illustrate the data. Therefore, the location parameter will be introduced to Gamma distribution as an additional parameter. Hence, the objectives of this study are (i) to develop the QM with three parameters of Gamma distribution to correct the biases between the observed and historical GCM (1976-2005) temperature series, (ii) to project the corrected future temperature series and (iii) to calculate the future mortality rate of CVD using the developed model (2006-2100).

5. Method and Data

5.1. **Data.**

5.1.1. *Meteorological Data.* Selangor is recognized as Malaysia's most developed state and one of the warmest regions with a daily mean temperature of 33°C. Selangor is used as a study area with longitude 100.775°N to 101.975°S and latitude 2.52°E to 3.875°W.

Observed mean daily temperature (1976-2005) was obtained from [9]. The GCM outputs from Model for Interdisciplinary Research on Climate (MIROC5) that was modeled under the Coupled Model Intercomparison Project Phase 5 (CMIP5) is used. In this study, the historical GCM daily mean temperature (1976-2005) and the future GCM daily mean temperature (2006-2100) under RCP8.5 were obtained. The RCP8.5 combines assumptions about high population and relatively

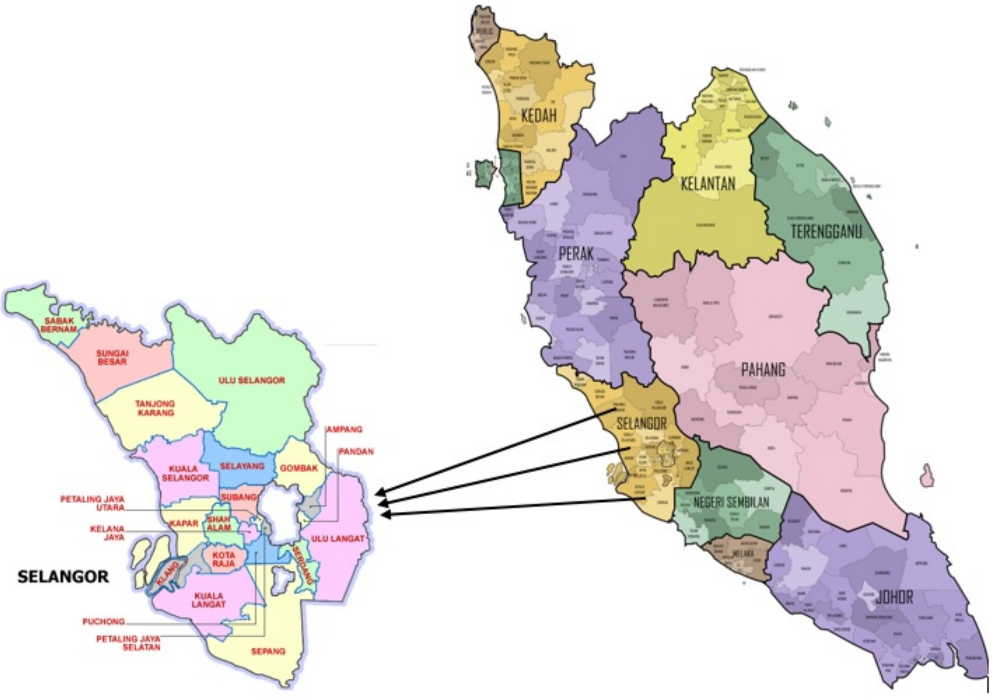


FIGURE 1. Map of Selangor

slow income growth with modest rates of technological change and energy intensity improvements, leading in the long term to high energy demand and GHG emissions in absence of climate change policies. It is a concentration pathway that leads to concentrations in 2100 which produces a change in forcing of 8.5 Wm^2, CO_2 equivalent of 1370 $p.p.m$, temperature anomaly of 4.9°C, and a rising pathway.

5.1.2. *Mortality Data.* Heart attack is the most common symptom of CHD. In this study, mortality data due to heart attack being recorded (2006-2015) were obtained from the National Cardiovascular Database (NCVD) reports. The population value set by the DOSM is used (6,298,500).

5.2. **Method.**

5.2.1. *Quantile Mapping.* QM defines a transfer function for the historical GCM output to match with certain statistical properties of the observed data. This transfer function is then used to correct the historical GCM output for the future GCM output [6]. The transfer function of QM is

$$h = \frac{1}{CDF_y}(CDF_x \times x), \tag{5.1}$$

where h is the corrected values, x is the observed daily mean temperature, CDF_x and CDF_y is the cumulative distribution function of observed and historical GCM daily mean temperature respectively. The projection of future daily mean temperature under RCP8.5 is calculated using equation (5.1). The observed values (x) were replaced with future GCM values (v) whereas the historical GCM values (y) were replaced with the corrected values (h). Let p_1 represents the projection of future temperature under RCP8.5.

$$p_1 = \frac{1}{CDF_h}(CDF_v \times v).$$

Gamma distribution with three parameters have probability density function which is

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

where, α the shape, β the scale, θ the location and $\gamma(\alpha)$ gamma function. The $\ln L$ is

$$\ln L = -n\alpha \ln\beta - n\ln\gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n}\ln(x_i-\theta) - \frac{\sum_{i=1}^{n}(x_i-\theta)}{\beta}. \tag{5.2}$$

By taking the the partial derivatives for equation (5.2) with respect to α, β, and θ and and letting the equation equals to zero,

$$\frac{\delta \ln\alpha}{\delta\alpha} = \sum_{i=1}^{n}\ln(x_i-\theta) - n\ln\beta - n\psi(\alpha) = 0, \tag{5.3}$$

$$\frac{\delta \ln\beta}{\delta\beta} = \frac{\sum_{i=1}^{n}(x_i-\theta)}{\beta^2} - \frac{n\beta}{\alpha} = 0, \tag{5.4}$$

$$\frac{\delta \ln\theta}{\delta\theta} = (\alpha-1)\sum_{i=1}^{n}\frac{1}{x_i-\theta} + \frac{n}{\beta} = 0. \tag{5.5}$$

The equations (5.3), (5.4) and (5.5) cannot be estimated directly because of the $\sum_{i=1}^{n}\frac{1}{x_i-\theta}$. The α, β, and θ parameters can be estimated in a closed form and do not require the simultaneous solution of the nonlinear equations [10]. The formula to estimate θ is

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

where

$$\theta_1 = 1 + (1 - 1/n)^n$$

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

x is the first order statistics of daily mean temperature value. The formula to estimate α and β are

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

$$\beta = \frac{\bar{x} - \theta}{\alpha},$$

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

5.2.2. *Attributable Daily Deaths.* The estimates of future (2006-2100) impact of high temperature on CVD mortality rates were calculated using the attributable daily deaths (ADD) formula which is,

$$ADD = y_0 \times ERC \times POP,$$

where y_0 is the baseline daily mortality, ERC is the attributable change in mortality for change in temperature, and POP is population. The value of POP and y_0 are constant.

6. Results and Discussion

Table 1 shows the estimated parameters values $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ for observed data, historical GCM, corrected values, and future GCM (RCP8.5). The values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ were best estimated for observed data, historical GCM, corrected values, and future GCM (RCP8.5) since the 95% confidence interval shows the values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ are in the range which shows the estimated parameters are highly significant.

Figure 2 shows the average monthly mean temperature from January until December for observed data, historical GCM data, and corrected values. The trend of temperature series is different between the observations and the historical GCM

TABLE 1. The estimated parameter values of the shape $\hat{\alpha}$, scale $\hat{\beta}$, and location $\hat{\theta}$ for observed data, historical GCM, corrected values, and future GCM (RCP8.5).

Data	Parameter	Estimated value (SE)	95% Confidence interval
Observed	$\hat{\alpha}$	411.1721 (0.19)	(411.1719, 411.1723)
	$\hat{\beta}$	0.1011 (0.000047)	(0.101099, 0.101100)
	$\hat{\theta}$	-15.5684 (-0.0072)	(-15.56839, -15.56841)
Historical GCM	$\hat{\alpha}$	1680.0231 (0.78)	(1680.0224, 1680.0238)
	$\hat{\beta}$	0.0278 (0.000012)	(0.027799, 0.027800)
	$\hat{\theta}$	-21.0113 (-0.0097)	(-21.01129, -21.01130)
Corrected values	$\hat{\alpha}$	411.1721 (0.19)	(411.1719, 411.1723)
	$\hat{\beta}$	0.0945 (0.00004)	(0.09449, 0.09450)
	$\hat{\theta}$	-14.5449 (-0.0067)	(-14.54491, -14.54486)
Future GCM (RCP8.5)	$\hat{\alpha}$	1312.9293 (0.61)	(1312.9288, 1312.9298)
	$\hat{\beta}$	0.0373 (0.00002)	(0.03729, 0.03730)
	$\hat{\theta}$	-21.6427 (-0.0099)	(-21.64270, -21.64269)

due to the existing bias in the data. However, when the biases are removed, the trend of the corrected temperature values is similar to the observations. The average monthly mean temperature recorded the lowest value in January and December while the highest value was recorded in May. This is because January and December correspond to the northeast monsoon rainfall where the weather is much cooler during this season. Peninsular Malaysia (including Selangor) experiences southwest monsoon rainfall in May where this area receives minimal rainfall during this period and considered as a dry period. Such phenomenon explains the highest recorded temperature in that month.

Figure 3 shows the average of monthly mean temperature for future raw GCM and corrected future GCM under RCP8.5 from January until December. The average daily mean temperature is lower in January. The average daily mean temperature starts to increase in March and remains constant until June. The trend is declining from the end of June until December. Even though the temperature values are slightly different, the pattern of the trend is similar between the raw and corrected GCM.

Figures 4 shows the percentage of mortality rate under RCP8.5 for male and female group. As seen, the mortality rate for males is increasing from 2006-2100

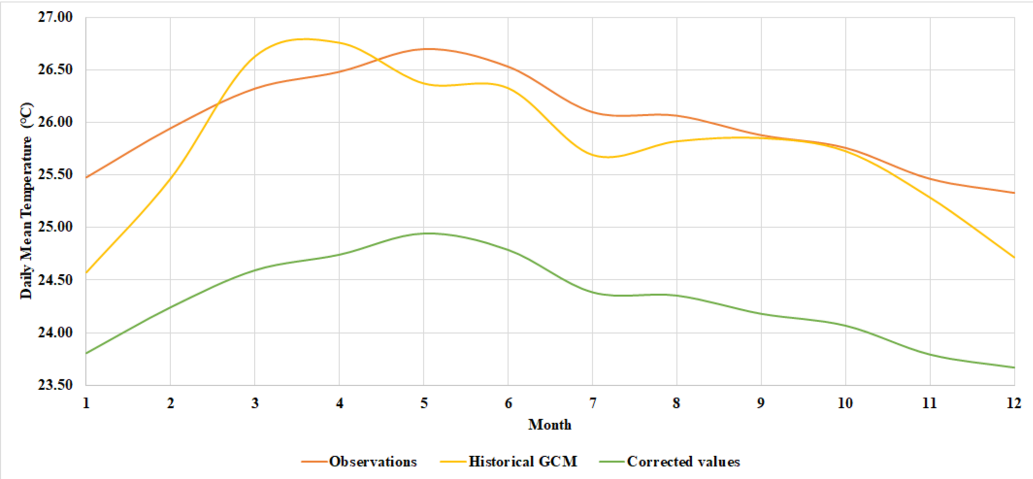


FIGURE 2. The average monthly mean temperature from January until December for observed data, historical GCM data, and corrected values.

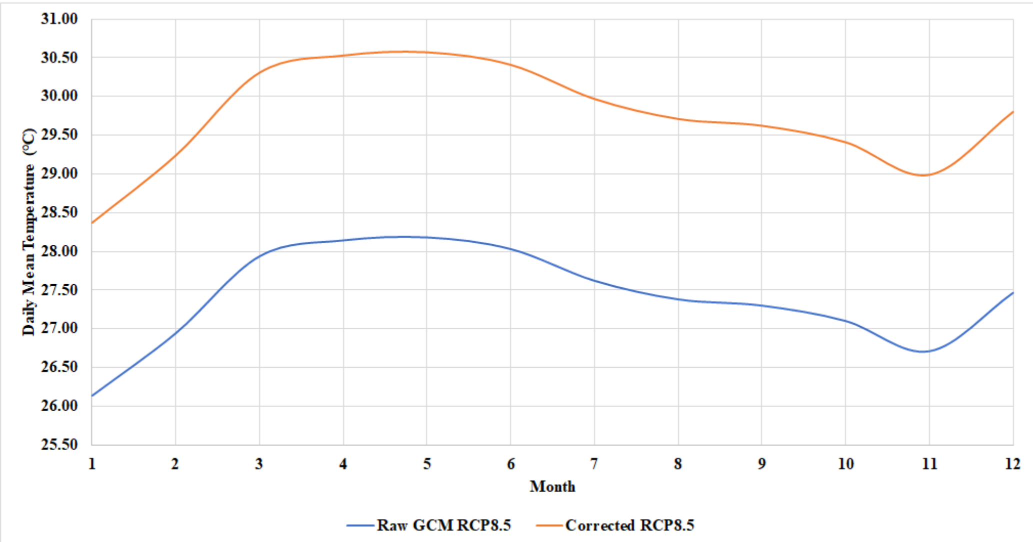
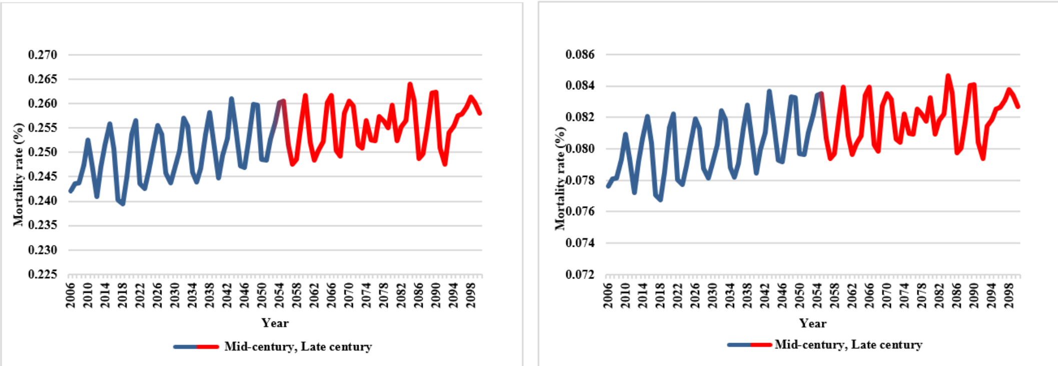


FIGURE 3. The average monthly mean temperature for future raw GCM and corrected GCM under RCP8.5 from January until December.

which ranges between 0.24% and 0.265%. Similarly, the mortality rate for females is increasing from 2006-2100 but with a lower percentage which ranges between 0.076% and 0.085%.



(A) Male group (B) Female group

FIGURE 4. The percentage of mortality rate under RCP8.5 for male and female group.

7. CONCLUSION

QM with three parameters of Gamma distribution has been fitted to the temperature series. It was found that QM was able to remove the biases and could capture the variability and the mean of temperature very well. The average corrected values of daily mean temperature for future GCM under RCP8.5 are projected to be higher than the observations. The mortality rate of CVD and temperature were directly proportional. The number of deaths for the male group is higher (0.24%-0.265%) than the number of deaths for the female group (0.076%-0.085%) from 2006-2100. The significant positive trend of death for the male group could be influenced by factors such as lifestyle, medical status, and family history. However, there are some caveats that need to be addressed in this study. Firstly, this study only corrects temperature in the present study, without consideration of any other biases from different sources, for instance, in selecting sample sizes and other potential sources of bias. Secondly, the population in the future is assumed to be constant. Finally, other causes of CVD and different groups of patients such as age and race need to be considered in a future study to achieve more significant results. All these issues deserve to be further studied. Nevertheless, the results are useful for impacts and mitigation assessment measures in disease prevention and health promotion.

ACKNOWLEDGMENT

The observed gridded data were sourced from [9]. This work was supported by the Malaysian Ministry of Higher Education Fundamental Research Grant Scheme (FRGS/1/2018/STG06/UPM/02/5) and Universiti Putra Malaysia, Putra-IPM grant (9619400).

REFERENCES

[1] MINISTRY OF HEALTH MALAYSIA (MOH): *Planning division Health Informatics Centre*, MOH/S/RAN/17.16, (AR)(August 2016).

[2] G. LI, Q. GUO, Y. LIU, Y. LI, X. PAN: *Projected temperature-related years of life lost from stroke due to global warming in a temperate climate city, Asia: disease burden caused by future climate change*, Stroke, **49**(4) (2018), 828–834.

[3] H.M. KEI: *Statistics Causes of Deaths, Malaysia in 2017*, 2017.

[4] H.M. KEI: *Statistics Causes of Deaths, Malaysia in 2018*, 2018.

[5] R. MOHAMAD: *Statistics Causes of Deaths, Malaysia in 2019*, 2019.

[6] M. LUO, T. LIU, F. MENG, Y. DUAN, A. FRANKL, A. BAO, P. DE MAEYER: *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, **10**(8) (2018), art.no.1046.

[7] S.T. NGAI, F. TANGANG, L. JUNENG: *Bias correction of global and regional simulated daily precipitation and surface mean temperature over Southeast Asia using quantile mapping method*, Global and Planetary Change, **149** (2017), 79–90.

[8] G.J. HUSAK, J. MICHAELSEN, C. FUNK: *Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications*, International Journal of Climatology: A Journal of the Royal Meteorological Society, **27**(7) (2007), 935–944.

[9] C.L. WONG, R. VENNEKER, A.B.M. JAMIL, S. UHLENBROOK: *Development of a gridded daily hydrometeorological data set for Peninsular Malaysia*, Hydrological Processes, **25**(7) (2011), 1009–1020.

[10] W.-K. PANG, S.-H. HOU, W.T. BOSCO, K.W.K. LI: *A simulation based approach to the parameter estimation for the three-parameter gamma distribution*, European Journal of Operational Research, **155**(3) (2004), 675–682.

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, UNIVERSITI PUTRA MALAYSIA, 43400 UPM SERDANG, MALAYSIA.

Email address: `ainaizzati05@gmail.com`

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, UNIVERSITI PUTRA MALAYSIA, 43400 UPM SERDANG, MALAYSIA.

Email address: `syafrina@upm.edu.my`