Climate Projections of Future Extreme Events in Malaysia

A.H. Syafrina, M.D. Zalina and A. Norzaida

Department of Science in Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Gombak, 53100, Kuala Lumpur, Malaysia

UTM Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

Abstract: In Malaysia, extreme rainfall events are often linked to a number of environmental disasters such as landslides, monsoonal and flash floods. In response to the negative impacts of such disaster, studies assessing the changes and projections of extreme rainfall are vital in order to gather climate change information for better management of hydrological processes. This study investigates the changes and projections of extreme rainfall over Peninsular Malaysia for the period 2081-2100 based on the RCP 6.0 scenario. In particular, this study adopted the statistical downscaling method which enables high resolution, such as hourly data, to be used for the input. Short duration and high intensity convective rainfall is a normal feature of tropical rainfall especially in the western part of the peninsular. The proposed method, the Advanced Weather Generator model is constructed based on thirty years of hourly rainfall data from forty stations. To account for uncertainties, an ensemble multi-model of five General Circulation Model realizations is chosen to generate projections of extreme rainfall for the period 2081-2100. Results of the study indicate a possible increase in future extreme events for both the hourly and 24 h extreme rainfall with the latter showing a wider spatial distribution of increase.

Keywords: Climate Change, General Circulation Model, Advanced Weather Generator, High Temporal Resolution, Statistical Downscaling

Introduction

Extreme precipitation events can be defined as maximum values of precipitation or exceedance above pre-existing high threshold (Stephenson, 2008). Such events developed from the combination of various factors including seasonal variability, hence determining the causes of it can be difficult. The Intergovernmental Panel on Climate Change (IPCC, 2013) reported that globally, almost all regions are projected to experience higher intensity and frequency of extreme precipitation. Precipitation extremes are predicted to escalate beyond the mean and the intensity of precipitation. In addition, the occurrences of extreme precipitation events are also expected to rise in nearly all parts of the world based on Special Report Emission Scenarios (SRES) A2 and B2.

With regards to extreme rainfall, which is the focus of this paper, Ekström et al. (2005) discovered that a rise in event magnitude throughout the UK with increasing trends for higher return periods in some areas. Winter extremes were also predicted to be more prevalent in the future with return periods lesser than present-day return periods. Similar results were observed by a study in the southwest of United States by Gershunov et al. (2013), in which extreme rainfall is predicted to become more frequent and more severe in the wintertime. Likewise, results in a study at multiple locations in the United States by Zhu et al. (2013) concluded that the intensity of extreme rainfall is predicted to be higher at all locations although the rate of increase varies among locations. The same phenomenon is occurring in Malaysia. Massive floods had been plaguing the country, such as the events of 2006, 2012 and 2014. Extreme rainfall during the monsoon seasons and highly intense convective rainfall during intermonsoon seasons were observed to be more frequent (Chia, 2004).

Hazards caused by extreme rainfall often results in extensive evacuation and loss of lives not to mention the destruction of public infrastructure, crop yield damage and economic losses (Juneng et al., 2010). A study on historical data between the years 1975 and 2010 by Syafrina et al. (2015) shows increasing trends of extreme rainfall in Peninsular Malaysia. This is an indication that Malaysia will face a higher probability of huge floods from heavy rainfall. Thus, information on the projections of extreme rainfall and their future
behaviour is of importance to the relevant authorities in Malaysia. Kwan et al. (2013) projected an increase of the probability of extreme rainfall occurrences during September to November over the west coast of Peninsular Malaysia. Increase of rainfall extremes were also projected for the stations located over the East Malaysia, particularly during the second half of the year. There is also an indication of earlier shift of monsoon onset at certain regions over the East Malaysia. These projections were made based on United Kingdom Meteorological Office (UKMO) Providing Regional Climates for Impacts Studies (PRECIS) model outputs for SRES A1B scenario for the period 2070-2099. Loh et al. (2016) also used PRECIS model on projections of future rainfall in Malaysia. The results show that during the months of December to May, approximately 20-40% decrease of rainfall is projected over Peninsular Malaysia and Borneo, particularly for the A2 and B2 emission scenarios. During the summer months, rainfall is projected to increase by 20-40% across most regions in Malaysia, especially for A2 and A1B scenarios. Both studies applied dynamical downscaling and used daily historical rainfall data to make projection on extreme and average rainfall for Malaysia.

Historical rainfall data and General Circulation Model (GCM) outputs of future forcing agents are invaluable input for projections of future extreme rainfall. GCMs are numerical models comprising of different earth frameworks and is widely used in providing outputs of global climate. Information on the significant processes concerning global and continental scale atmosphere can be projected by GCM for future atmosphere under different emission scenarios of forcing agents. Despite numerous uncertainties in the different GCMs (Chu et al., 2010), these outputs provide hydrologists with priceless information. However, the coarse resolution of GCMs may lead to mismatch between the model’s variables against observational variables for many climate change impact studies (Fowler et al., 2007; Hessami et al., 2008; Hashmi et al., 2009; Chu et al., 2010; Hashmi et al., 2010; Fatichi et al., 2011). The mismatch issues tend to produce inaccurate simulations of current regional climate for sub-grid scales (Chu et al., 2010; Guo et al., 2011). In order to match the scale between the GCM outputs and hydrological process at smaller scale, downscaling has been widely employed. Downscaling comprises of two different approaches, known as dynamical and statistical downscaling. Dynamical downscaling uses Regional Climate Model (RCM) to simulate climate variables with GCMs providing the boundary conditions (Fowler and Wilby, 2010). Statistical downscaling is an empirical method that defines the statistical relationships between the large-scale climate features and the hydrological variables (Wilby et al., 2004; Sunyer et al., 2011). There are various discussions and debates on these two approaches, however both practices are extensively used in determining the projections of future climate scenarios.

Statistical downscaling assesses relationships between large-scale atmospheric variables (predictors) and local-scale variables (predictands) in order to extrapolate future scenarios. There are two fundamental presumptions inside these strategies (IPCC, 2007) (i) the relations between predictors and predictands are presumed to be constant in the climate change context (stationary) and (ii) the selected predictors adequately represent the climate change signal for the predictand. The statistical relationship is used in conjunction with the change in the predictors to determine the future local climate. Basically, statistical downscaling approaches could be categorized into three main generic classes, namely regression models, weather generators and weather typing methods. Regression models could assess the relation between the climatic variables at local scale (e.g., rainfall) and a set of large-scale atmospheric variables (Fowler et al., 2007). On the other hand, weather generators are employed to reproduce time series of climatic variables such as rainfall, atmospheric humidity and wind speed. The Advance Weather Generator (AWE-GEN), developed by Fatichi et al. (2011), is an hourly weather generator which has the capacity to replicate climatic variables and crucial statistical properties of such variables. In addition, AWE-GEN is also able to reproduce extreme rainfall values. The third type, weather typing technique is based on the concept of gathering a fixed number of discrete weather types or “states” according to their synoptic similarity. GCMs are then adopted to evaluate the change in the frequency of weather types in order to estimate climate change (Burton et al., 2010; Quintana-Segui et al., 2011).

This study proposes the use of statistical downscaling method for future projections of extreme rainfall at hourly scale within Peninsular Malaysia. Such method is preferred because preceding studies which concentrated on dynamical downscaling, were using daily rainfall data as input, whereas, statistical downscaling, in particular the AWE-GEN model will enable smaller resolution, such as hourly data, to be used for the input. Smaller resolution data is greatly related to the high intensity convective rain, which is a common feature in tropical urban areas such as Malaysia (Syafrina et al., 2015; Norzaida et al., 2016). Secondly, statistical downscaling has the capacity to use an ensemble of numerous GCMs for the projections, which tend to match the overall observations better. Hence in this study, an ensemble of five GCMs under the RCP 6.0 forcing scenario is utilized. All these inputs will be incorporated into the AWE-GEN model to produce future projections of extreme rainfall for the period 2081-2100. The period of 2081-2100 will be used for future projections due to the fact that continuous greenhouse gas emissions at or beyond the present levels may lead to further warming
and could consequently alter the global climate system. This is in line with the report by IPCC (2007) which state that changes in global climate system in 21st century is expected to be bigger compared to the 20th century.

This paper investigates the trends of future projections of extreme rainfall at hourly scale with the use of AWE-GEN model in Peninsular Malaysia. In the next section, the adopted methodology is discussed in detail, followed by model validation to assess model’s capability in simulating rainfall series. Subsequently, Section 4 and 5 discussed the results of the future projections and draw conclusion on the expected trends of rainfall in Malaysia, especially with regards to extreme rainfall.

Data and Methodology

Data

In this study, the AWE-GEN model is constructed based on 30 years of historical data (1975-2005). The input data required by AWE-GEN are hourly rainfall, hourly temperature, hourly relative humidity and hourly wind speed. These input data are gathered from forty rainfall stations across Peninsular Malaysia. The simulation of rainfall series as well as the projection of future extreme rainfall will be the output for this study. These stations are chosen due to the quality of their data in terms of completeness and record length. Furthermore, stations are evenly distributed across the Peninsular. Stations with missing values greater than 2% of the total record hours within 1st January 1975 to 31st December 2010 were omitted. The process of choosing the stations also applied the Average Nearest Neighbour (ANN) to ensure that selected stations are sufficiently spaced out over the Peninsular (Syafrina et al., 2015).

If the $z_{ANN}$-score is less than 1, the stations are clustered. Otherwise, the stations are evenly spread throughout Peninsular Malaysia. The ANN test has been performed in this study, using 99% level of significance, the calculated $z_{ANN}$-score is found to be 2. This shows that the stations are evenly spread throughout the Peninsular Malaysia since the $z_{ANN}$-score is greater than 1. However, it is noted here that Selangor has a larger number of stations (Selangor is a state on the western region of the Peninsular). This is due to the fact that it is one of the most advanced state within Malaysia. As expressed by (Cowpertwait et al., 1996a; 1996b), the stations are evenly distributed across the Peninsular. Stations with missing values greater than 2% of the total record hours within 1st January 1975 to 31st December 2010 were omitted. The process of choosing the stations also applied the Average Nearest Neighbour (ANN) to ensure that selected stations are sufficiently spaced out over the Peninsular (Syafrina et al., 2015).

Realizations were acquired from the data pool in the World Climate Research Programme’s (WCRP’s), Coupled Model Intercomparison Project phase 5 (CMIP5). Models are selected according to the availability of data (availability of hourly rainfall time series as the main constraint) and the relative independence between models. The latter criterion is a necessity in using multi-model ensemble approach, which is the mutual independence between model realizations. Climatic models proposed by various groups could be assumed to be independent to a certain extent; nevertheless, these models may have similar elements or contain similar underlying theories for their parameterizations (Tebaldi and Knutti, 2007). To ensure the preservation of the relative independence among models, whenever multiple or revised versions of similar climate model are available, only a single version of each GCM is used. A single scenario, the RCP 6.0 scenario is adopted since it is an intermediary situation that relates to the median curve of global temperature increase among all considered scenarios.

Neyman-Scott Rectangular Pulse Model

In AWE-GEN, the intra-annual variability of rainfall is captured by the Neyman-Scott Rectangular Pulses (NSRP) model. Work by (Abas et al., 2014; Norzaida et al., 2016) indicated that the NSRP model is suitable to be used in Malaysia. As expressed by (Cowpertwait et al., 1996a; 1996b), $Y(t)$ is a random variable representing the rainfall intensity at time $t$ and $Y(ih)$ is the aggregated rainfall depth in the $i$th sampling interval of length $h$. Thus:

$$Y(ih) = \int_{(i-1)h}^{ih} Y(t) dt$$  \hspace{1cm} (1)

It is assumed that the rainfall time series, $\{Y(h): i = 1,2,…\}$ is stationary so that $E\left(Y(ih)\right) = E\left(Y(ih)\right)$ for all $i, j = 1,2,…$. Without loss of generality, the superscripts $i$ and $j$ may be omitted and the moments of $Y(h) = \int_{0}^{h} Y(t) dt$ will be considered. A general expression for the $n$th moment is given by:

$$E(Y(h)) = \int_{0}^{h} \cdots \int_{0}^{h} E(Y(t_1)Y(t_2)\ldots Y(t_n)) dt_1 dt_2 \ldots dt_n$$  \hspace{1cm} (2)

A full account of the mathematical expression is given by Rodriguez-Iturbe and Eagleson (1987). The arrival times of storm origins is assumed to follow a Poisson process with rate $\lambda$, each storm origin generating a random number $C$ of cell origins according to the geometrical distribution with the mean $\mu_c$. 

394
The rectangular pulse \((L, X)\) is associated with each cell origin, where \(L\) and \(X\) are independent random variables corresponding to the lifetime and intensity of the pulse, respectively. The pulse represents a rain cell. \(X_{i,w}(u)\) is an independent random variable representing the rainfall intensity at time \(t\) due to a cell with starting time \(t-u\). \(\Delta M(t) = N(t, t+\delta)\) is the number of cell origins in the time interval \(N(t, t+\delta)\). The total intensity at time \(t\), \(Y(t)\), is the summation of the intensities of all cells alive at time \(t\) and can be written as:

\[
Y(t) = \int_{u=0}^{t} X_{i,w}(u) dN(t-u)
\]

Equation 2 can be estimated using Equation 3 as follows:

\[
E\{Y(t)\} = \sum_{u_{i-1}+u_{i}=t} \left\{ E\{X_{i-w}(u)\} X_{i-w}(u_{i}) \ldots X_{i-w}(u_{a}) \right\}
\]

\[
= \sum_{u_{i-1}+u_{i}=t} \sum_{u_{a} \geq 0} \left\{ E\{X_{i-w}(u)\} X_{i-w}(u_{i}) \ldots X_{i-w}(u_{a}) \right\}
\]

\[
\times E\{dN(t_{i}-u_{i}) dN(t_{i}-u_{i}) \ldots dN(t_{i}-u_{a})\}
\]

because \(X_{i,w}(u)\) and \(dN(t-u)\) are independent.

It is assumed that \(L\) is exponentially distributed with mean \(\eta^{-1}\). Cell origin waiting time after the occurrence of a storm origin is independently exponentially distributed with mean \(\beta^{-1}\), which means that no cell origin occurs at the storm origin. The properties of \(C\) which essentially follow from Equation 4 and 2, can be written as:

\[
\mu_{b} = E\{Y_{b}\} = \lambda \mu_{h} h / \eta
\]

\[
\gamma_{h,l} = Cov\{X_{i}^{h}, X_{i+1}^{l}\} = \lambda h \gamma_{-l} A(h, l)
\]

\[
= \frac{2 \mu_{b} E\{X^{2}\} + \frac{\mu_{h}^{2} \beta^{2} E\{C^{2} - C\}}{\beta^{2} - \eta^{2}} - \lambda \mu_{h}^{2} B(h, l) E\{C^{2} - C\}}{\beta^{2} - \eta^{2}}
\]

where, \(A(h, 0) = (h \eta + e^{-\beta h} - 1), B(h, 0) = (h \beta + e^{-\beta h} - 1)\) and for \(l\) a positive integer, \(A(h, l) = \frac{1}{2} (1 - e^{-\beta h}) e^{-\beta hl/2} \) and \(B(h, l) = \frac{1}{2} (1 - e^{-\beta h}) e^{-\beta hl/2} \).

In order to estimate the parameter, an objective function comprising of statistical properties of rainfall at different aggregation times is used. The selected statistical properties, are the coefficient of variation \(C_{h}(h) = \frac{\gamma_{h,l}}{\mu_{h}} \), the lag-1 auto-correlation \(\rho(h) = \gamma_{h,1} / \gamma_{h,0}\) the skewness \(\kappa(h) = \gamma_{h,1} / \gamma_{h,0}^{2}\) and the probability that an arbitrary interval of length \(h\) is dry. \(\phi(h)\). The parameters \(\mu_{h}, \gamma_{h,1}\) and \(\gamma_{h,0}\) represent the mean, the covariance and the third moment of precipitation process at a given aggregation time interval \(h\) and lag \(l\). Specifically, statistical properties of rainfall process at four different time scales \(h: 1, 6, 24\) and \(72\) h are used. Maximization of the objective function is achieved by applying the simplex method. A set of parameters are estimated on a monthly basis in order to account for seasonality.

Previous study by Wilks and Wilby (1999) indicated that variance of the generated series was smaller than the variance inferred from observed data due to the underlying stationarity assumption in weather generator. Therefore, the Auto-Regressive lag-1 (AR1) property is adopted to ensure the variance and the autocorrelation properties of the precipitation process at the annual scale are preserved. The AR1 model is:

\[
P_{\rho}(i) = \overline{P}_{\rho} + \rho_{\rho} (P_{\rho} - \overline{P}_{\rho}) + \eta(i) \sigma_{\rho} \sqrt{1 - \rho^{2}}
\]

where, \(\overline{P}_{\rho}\) [mm] is the average annual precipitation, \(\sigma_{\rho}\) is the standard deviation and \(\rho_{\rho}\) is the lag-1 autocorrelation of the process. The term \(\eta(i)\) represents random deviate of the process which is transformed according to Wilson-Hilferty approach. The parameters \(\overline{P}_{\rho}, \sigma_{\rho}, \rho_{\rho}\) and \(\gamma_{h,0}\) are determined from the annual observations. The rejection threshold \(\overline{P}\) is determined based on the information relating to observational errors of annual rainfall. The symbol \(M\) refers to the maximum number of iterations \(j\) allowed within a given year \(i\), when searching for the best match of the total annual rainfall between the NSRP and AR (1) models. The NSRP model is used to generate rainfall series at the hourly time scale for the period of one year. The obtained total rainfall will be compared with the annual

**Table 1.** GCMS that will be used for future projections

<table>
<thead>
<tr>
<th>Modelling centre</th>
<th>GCM</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOAA Geophysical Fluid Dynamics Laboratory, United States</td>
<td>GFDL-CM3</td>
<td>2.0°×2.5°</td>
</tr>
<tr>
<td>Institut Pierre-Simon Laplace, Paris</td>
<td>IPSLCM5A-LR</td>
<td>1.875°×3.75°</td>
</tr>
<tr>
<td>Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan</td>
<td>MIROC5</td>
<td>1.406°×1.4°</td>
</tr>
<tr>
<td>Meteorological Research Institute, Japan</td>
<td>MRICGCM3</td>
<td>1.125°×1.121°</td>
</tr>
<tr>
<td>Norwegian Climate Centre, Norway</td>
<td>NorESM1-M</td>
<td>2.5°×1.895°</td>
</tr>
</tbody>
</table>
values estimated with the autoregressive model in Equation 7. If the disparity between the two values is higher than a certain percentage $p$ of the measured long-term mean annual rainfall, the generated hourly series of one year length will be discarded. New series will then be generated and the comparison process is repeated. However, if the difference between the two values is within threshold range, the generated series will be accepted. The whole process will be repeated until every annual values generated with model in Equation 7 have matching hourly series generated with the NSRP model.

**Factor of Change**

Factor of change method is applied for simulation of future hourly series of rainfall. Factors of change are used specifically to perturb the statistically derived time series to generate statistical expressions of future hourly time series (Wilby et al., 2004; Fowler et al., 2007; Fatichi et al., 2011). The climate statistical properties for a given station are calculated from the observations as well as from the GCMs. In particular, the mean, variance, lag-1 autocorrelation, skewness and the frequency of no-precipitation of the observed rainfall data are estimated at different aggregation of 24, 48, 72 and 96 h. These statistical properties of GCM are also calculated for both control and future period at different aggregation of 24, 48, 72 and 96 h. The equation representing the product factor of change for a statistical property of rainfall at the time aggregation $h$ is:

$$ S(h)^{FUT} = \frac{S(h)^{GCM,FUT}}{S(h)^{GCM,CTS}} S(h)^{OBS} $$

(8)

where, $FUT$ denotes the future scenario, $OBS$ denotes observations and $CTS$ denotes the control scenario while the $GCM$ denotes the climate model:

$$ t_{\text{mon}}^{FUT} = t_{\text{mon}}^{OBS} + \left( (t_{\text{mon}}^{GCM,FUT} - t_{\text{mon}}^{GCM,CTS}) \right) $$

(9)

**Bayesian Approach**

The multi-model ensemble technique being implemented is taken from the work by Tebaldi et al. (2007), which proposed the Bayesian statistical model. Information from several GCMs and observations (i.e., factors of change) are merged to find the Probability Density Functions (PDFs) of future changes for a particular climatic variable at the regional scale. According to Bayesian approach, all unspecified quantities are modelled as random variables, with a priori probability distributions. Assumptions comprise of the specific requirement of conditional distributions for the data (likelihood), given the parameters and the prior distributions for all the parameters of the Bayesian framework. Through Bayes’ theorem, prior distributions and likelihood are combined into a posteriori distributions of parameters.

**Likelihood Functions**

Let Gaussian distributions for $X_i$ and $Y_i$:

$$ X \sim N(\mu, \nu^{-1}) $$

(10)

$$ Y \sim N(\nu, \omega \upsilon^{-1}) $$

(11)

where, $N(\mu, \nu^{-1})$ denotes a Gaussian distribution with mean $\mu$ and $1/\nu$ variance Variables $\mu$ and $\nu$ are representing true values of present and future temperature in a particular region and season. A key parameter of interest will be $\Delta T = \nu - \mu$ representing the expected temperature change. Meanwhile, the parameter $\upsilon$ reciprocal of the variance, is referred to as the precision of the distribution of $X$. The distribution will be parameterized by the product $\omega \upsilon$ where $\omega$ is an additional parameter, common to all GCMs in order to allow for the possibility that $Y$ has different precision from $X$.

GCM responses are assumed to have a symmetric distribution, whose centre is the “true value” of temperature, but with an individualistic variability viewed as a measure of how well each GCM approximates the climate response to the given set of natural and anthropogenic forcings (Tebaldi et al., 2004). This assumption has been supported by the Coupled Model Intercomparison Project (CMIP) studies (Meehl et al., 2000) where mean of a super ensemble has better validation properties as compared to the individual members. Other assumption is that the single Atmospheric Ocean General Circulation Model (AOGCM’s) realizations are centered around the true value could be easily modified in the presence of additional data. For example, if single-model ensembles are available, then an AOGCM-specific random effect could be incorporated.

The likelihood model of the observations of current climate as $X \sim N(\mu, \upsilon)$, $\upsilon_0$ is known as natural variability specific to the season, region and time average applied to the observations where it is different from $\upsilon_1, \ldots, \upsilon_k$ where $k$ is the number of GCM. This measures of model-specific precision and depend on the numerical approximations, parameterizations, grid resolutions of each GCM. The value of $\upsilon_0$ is fixed using estimates of regional natural variability from Giorgi and Mearns (2002). It could also be treated as a random variable as well if the data contained a long record of observations that could be used for its estimation. A normal prior and likelihood are being used in this study since the result is just the same as the posterior distribution obtained from the single observation of the
mean $\bar{X}$, since we know that $X \sim N\left(\frac{\sigma^2}{n}\right)$ and the above formula are the ones we had before $\frac{\sigma^2}{n}$ replaced with $\sigma^2/n$ and $X$ by $\bar{X}$ (Box and Tiao, 2011).

Prior Distributions

Precision parameters $\upsilon_i$, $i = 1, \ldots, 5$ are distributed according to Gamma prior densities which is $\text{Ga}(a, b)$ of the form:

$$b^a \Gamma(a)^{-1} e^{-b \upsilon}$$  \hspace{1cm} (12)

With $a$, $b$ known and chosen to ensure that the distribution will have a large variance over the positive real line. Similarly, $\omega \sim \text{Ga}(c, d)$ with $c$, $d$ known. A Gamma probability distribution as tested by Weibull has also been tested and compared with Gamma, however, no significant difference is observed.

Posterior Distributions

Bayes’s theorem is applied to the likelihood and priors specified above. The resulting joint posterior density for the parameters $\mu, \upsilon, \upsilon_1, \ldots, \upsilon_5$ is given by, up to a normalizing constant:

$$\prod_{i=1}^{5} \frac{\upsilon_i}{\bar{X}_i}^{\upsilon_i} \times e^{\frac{\upsilon_i}{2} \left(\left(X_i - \mu\right)^2 + \omega \left(Y_i - \upsilon\right)^2\right)} \times e^{\frac{\omega}{2} \left(\left(X_i - \mu\right)^2 \right)}$$  \hspace{1cm} (13)

Inference cannot be drawn from its analytical form due to the distribution in Equation 13 is not a member of any known parametric family. Therefore, Markov Chain Monte Carlo (MCMC) simulation is used to generate a large number of sample values from Equation 13 for all parameters and approximate all the summaries of interest from sample statistics. The distribution of $\mu$ fixing all other parameters is a Gaussian distribution with mean:

$$\bar{\mu} = \frac{\sum_{i=1}^{5} \upsilon_i X_i}{\sum_{i=1}^{5} \upsilon_i}$$  \hspace{1cm} (14)

and variance:

$$\left(\sum_{i=1}^{5} \upsilon_i\right)^{-1}$$  \hspace{1cm} (15)

Similarly, the conditional distribution of is Gaussian with mean:

$$\bar{\upsilon} = \frac{\sum_{i=1}^{5} \upsilon_i Y_i}{\sum_{i=1}^{5} \upsilon_i}$$  \hspace{1cm} (16)

and variance:

$$\left(\sum_{i=1}^{5} \upsilon_i\right)^{-1}$$  \hspace{1cm} (17)

The weights $\upsilon_1, \ldots, \upsilon_5$ in Equations 14 and 15 are random quantities and account for the uncertainty in their estimation. Such uncertainty will inflate the width of the posterior distributions of $\upsilon$, $\mu$ and thus also $\Delta T$. An approximation to the mean of the posterior distribution of the $\upsilon_i$ for $i = 1, \ldots, 5$ is:

$$E\left(\upsilon_i \mid \{X_1, \ldots, X_5, Y_1, \ldots, Y_5\}\right) \approx \frac{a + 1}{b + \frac{1}{2} \left(\left(X_i - \bar{\mu}\right)^2 + \omega \left(Y_i - \bar{\upsilon}\right)^2\right)}$$  \hspace{1cm} (18)

Specifically, sample values from the posterior will be generated by MCMC using Metropolis-Hastings algorithms to get an accurate empirical estimation of its features. The posterior mean of $\mu$ is approximately:

$$\bar{\mu} = \frac{\sum_{i=1}^{5} \upsilon_i X_i}{\sum_{i=1}^{5} \upsilon_i}$$  \hspace{1cm} (19)

a weighted average of the observation and model output, with weights $\upsilon_0, \upsilon_1, \ldots, \upsilon_5$.

The posterior mean of $\upsilon$ is approximately:

$$\bar{\upsilon} = \frac{\sum_{i=1}^{5} \upsilon_i Y_i}{\sum_{i=1}^{5} \upsilon_i}$$  \hspace{1cm} (20)

a weighted average of the 5 model responses, with weights $\upsilon_0, \upsilon_1, \ldots, \upsilon_5$.

$$\Delta P = 100 \frac{\upsilon - \bar{\upsilon}}{\bar{\mu}},$$  \hspace{1cm} (21)

the percent precipitation change, is a derived quantity and its posterior mean is similarly a weighted average of the individual models’ precipitation change signals, with weights a function of $\upsilon_0, \upsilon_1, \ldots, \upsilon_5$.

Each $\upsilon_i$’s posterior mean is approximately:

$$\bar{\upsilon}_i \approx \frac{1}{\left[X_i - \bar{X}\right] + \theta |Y_i - \bar{Y}|}$$
Validation of the Model

In order to validate the model, the simulated hourly rainfall is divided into two non-overlapping periods of (i) 1975 to 1989 and (ii) 1990 to 2005. The 1975 to 1989 is used as the reference period where the multiplicative factor is calculated based on the simulation output and the high resolution observational data. The changing factors are then used to correct the biases of the simulation output from 1990 to 2005. The corrected hourly rainfall is then compared to the observation from the identical period of 1975-1989.

Results and Discussion

The initial stage of assessing model’s performance involved comparing generated results to historical data within the period of 1975-2005. Rainfall series are simulated for all 40 rainfall stations independently. The location of 40 rainfall stations is shown in Fig. 1 while the list of stations is listed in Table 2. Results of station Loji Air Kuala Kubu Bharu Selangor (station 3516022) is discussed in this study. Fig. 2 gives the comparison of statistical properties between the simulated and observations. As can be seen in this figure, the monthly statistics are well reproduced at each period. Similar results are also observed for other remaining stations used in this study. Fig. 3 shows the simulations of extreme rainfall against the observations. AWE-GEN has shown good performance in simulating the extreme rainfall up to the return periods of 20-30 years (Fig. 3a and b) as well as extremes wet spell durations and extreme dry spell duration (Fig. 3c and d).
The projections of future hourly extreme rainfall are shown in Fig. 4. Both hourly and 24 h extremes rainfall seems to be on rise in future especially in the higher return periods. The 40 year return period for the hourly extreme exceed 100 mm while the 24 year extreme exceed 200 mm for the same return period (Fig. 4a and b). This is parallel to the study by MMD, which reported that the intensity and frequency of extreme events in Malaysia are increasing based on long-term historical data from 1951 to 2009 (Diong et al., 2010). Meanwhile, extreme dry spell is projected to decline in future with less than 50 consecutive days, whereas extreme wet spell is expected to hold constant in future with 20 consecutive days (Fig. 4c and d). Results of 40 stations are summarized in Fig. 5, comparison of future and observed hourly extreme rainfall ((a)-(c)) and 24 h extreme rainfall ((d)-(f)) for 10, 20 and 40 years return periods. In general, 67.5% of the stations show an increase in hourly extreme rainfall whilst 82.5% shows an increase in 24 h extreme rainfall for all return periods. Most of the stations depicting an increase in hourly extreme are located on the western part of the Peninsular where convective rainfall is more prominent. During the two inter-monsoon seasons (March-April and Sept-Oct), the late thunderstorm driven by local convection and land-sea breezes is frequently occurred on the west coast.
Fig. 2. Comparison of statistical properties between observed and simulated data at aggregation time of 1, 24 and 48 h for station Loji Air Kuala Kubu Bharu, Selangor (station 3516022), (a) Mean, (b) Variance, (c) Lag-1 autocorrelation, (d) Skewness, (e) Probability of no rainfall, (f) Transition probability wet-wet
Fig. 3. A comparison between Observations (OBS) and simulations for the control period (CTS) values of extreme rainfall at (a) 1 and (b) 24 h aggregation periods; (c) extremes of dry and (d) wet spell durations for station Loji Air Kuala Kubu Bharu, Selangor (station 3516022)

Fig. 4. A comparison between Observations (OBS), simulations of the control period (CTS) and simulations of future period (FUT) values of extreme rainfall at (a) 1 and (b) 24 h aggregation periods; (c) extremes of dry and (d) wet spell durations for station Loji Air Kuala Kubu Bharu, Selangor (station 3516022)

Higher rainfall intensity in the future is modulated by the enhanced local convective activities in a warmer climate. During this period, the east coast appears to be less affected by the warming which resulted a decrease in future extreme rainfall (Kwan et al., 2013). This is consistent with the results of this study where (Fig. 5 and 6) few stations on the eastern coast experience a decrease for both hourly and 24 h extremes.
Fig. 5. Comparison of future and observed extreme rainfall for the 40 stations for 10, 20 and 40 years return periods: (a-c) hourly; (d-f) 24 h

However, most parts of the peninsular including parts of the eastern region show a future increase for the 24 h extreme. This could be explained by the longer duration of rainfall events during the NEM in the eastern regions (Suhaila et al., 2010). Moreover, the MMD (2009) reported an increase in tropical storms in the South China Sea which contributed to more extreme rainfall events in eastern regions. Another factor that contributed to the rise of extreme rainfall events is the increase in concentration of Greenhouse Gases (GHG) in the atmosphere. The IPCC (2007) indicates that warming and increase of water vapour over most land areas as well as stronger La Niña may have contributed to the increase in extreme rainfall across the region. Figure 6 presents the percentage of change between CTS and FUT scenario in the mean monthly rainfall for each station. The mean monthly rainfall for each station has been compared between CTS and FUT. The percentage of change is calculated as $(\text{FUT}-\text{CTS})/\text{CTS} \times 100\%$. As seen in Fig. 6, generally, the mean monthly rainfall is expected to increase in future especially during the NEM (Nov-Feb) where eastern region is most affected. An increase of mean future monthly rainfall is also expected during March and April (MA) and SWM (May-August) where western region is most affected. The mean of future monthly rainfall on the northern region is expected to decrease during the NEM. Similarly, the mean of future monthly rainfall on the southern region is also expected to decrease during the SWM.
Conclusion

In summary, results from this study has shown a future increase in both hourly and 24 h extremes with the latter showing a wider spatial distribution of increase. Overall, 62.5% of stations show future increase in both hourly and 24 h extremes. This is in agreement with the current scenario where Malaysia experiences positive trends of extreme rainfall since the 19th century. The results on analysis of future projections of extreme rainfall are also consistent with the current trends in extreme events where both intensity and frequency of extreme events are forecasted to increase in most parts of the world. However, this study could be extended to include additional stations with adequate data length to evaluate spatially weighed trends. Digitization of these data could have a positive impact on the projection of long-term climate change.

Acknowledgement

This research is funded by the Ministry of Education Fundamental Research Grant (FRGS) vote 4F120, Research Initiative Grant Scheme (Project ID: RIGS16-079-0243) and the Ministry of Higher Education, Malaysia. Authors would also like to thank Drainage and Irrigation Department Malaysia for the rainfall data and Simone Fatichi for the downscaling future projections MATLAB program codes.
Author's Contributions

A.H. Syafrina: Simulation work, data analysis and manuscript writing.

M.D. Zalina and A. Norzaida: Data analysis and interpretation and manuscript writing.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that co-authors have read and approved the manuscript and no ethical issues are involved.

References


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. Cambridge University Press, Cambridge, United Kingdom and New York, USA.


