Frequency of Rainy Days as One of Parameter Changes in Rainfall Behavior

Najwa Zainal*, Firdaus Mohammad Hamzah* & Othman Jaafar*

*Department Of Civil And Structural Engineering, Faculty Of Engineering And Built Environment, National University Of Malaysia, 43600 Ukm Bangi, Selangor, Malaysia.

Abstract

Daily precipitation data of 27 years from six research area were collected around Peninsular Malaysia where four in west regions and another two in east region. The purpose is to study one of the parameter that affected the rainfall behavior; frequency of rainy days affect amount of rainfall. Those data will undergo three stages of analysis which started by using Pearson correlation and the result were then strengthen by using Smooth Regression to determine the linearity relationship between frequency of rainy days and amount of precipitation. Finally, daily precipitation data were used to look at the distribution of rainfall on research area using the daily precipitation concentration index (CI).

Keywords: Daily Precipitation, Frequency of rainy days Pearson'r Correlation, Concentration Index (CI), Smooth Regression, Linearity Test

Kata kunci: Daily Pemendakan, Kekerapan hari hujan Pearson'r Korelasi, Indeks Kepekatan (CI), Regresi Smooth, Keleburusan Test

INTRODUCTION

Malaysia is Southeast Asia country as in Fig.1 [1]. Malaysia is a country where the climate is strongly influenced by natural climate variability with oceans as geographically located in between of two large ocean which is Pacific ocean to east and Indian Ocean to west until our rainfall distribution were known as monsoon rains [2], [3].

From the report [4], the records of earth surface temperature showed that the climate of the earth is globally warming as the temperature has increased about 0.74°C in global average surface temperature for the past 100 years. Due to the changes, extreme change in rainfall behavior occurred such as extreme flood events in Ireland [5], Southern African (Ambrosino et al 2011) and South Korea, [7]. Extended decadal droughts and even heat wave had hit in Australia (Greenville et al 2012), China (Zhai et al. 2005) and Italy [10].

Due to these severe of climate change, a study on the distribution of rainfall has been done in a variety of factors to find the cause and the evidence. This can be look over through research such as in Ghana, Orissa, East African and Nigeria where a significant rainfall trend shown [11]–[14] Beside that, research on these matter also had been covered such as in Bangladesh [15], South Africa [16], Spain [17], India [18], Northeastern US [19], Europe [20], Western Saudi Arabia [21], Portugal [22] and Hawaii [23]. Not left behind, research on trends in Peninsular Malaysia during monsoon by [24] had been done.

However, according to Tangang et al. 2012, Malaysia still lack of knowledge on the intensity and frequency of future extreme events (drought and flood). Thus, this paper focus on a relationship between frequency of rainy days and amount of rainfall using a better approach to describe the frequency as one of parameters that affect the rainfall variability by considering the Pearson’s correlation and also the contribution of days of greatest rainfall to the total amount which is known as rainfall concentration index (CI) is used [25]. Then, to strengthen up the evidence whether this frequency of rainy days had a big impact on changes of rainfall distribution, a model of linearity is formed using Sm.regression analysis.

MATERIALS AND METHODS

Study Area :

A map that indicates the location of selected stations that will be used in the analysis is shown in Fig.2. The rainfall stations were chosen as represent two regions in Peninsular Malaysia which is the east and west as the area also is the former rural area in which then were developed into thriving city that became the focus of the crowd. Around 27 years (1985-2014) of daily rainfall data for about six rainfall station were obtained.
from Malaysian Meteorological Department and School of Social Science, National University of Malaysia. Due to rainfall data is secondary data, homogeneity test were done and the result shown that all the data from six station were homogen and valid for the next analysis (N. H. Ahmad et al. 2013; Firat et al. 2010; Ho, & Yusof 2013; Hosseinzadeh Talae et al. 2014; Ng et al. 2015; and Wijngaard et al. 2003).

\[
r = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sqrt{\left(\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}\right) \left(\sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N}\right)}}
\]

(1)

Where, \(N\) = represents the number of pairs of data \((X,Y)\) in which the overbar will indicate the average values over the sample of the size \(N\) [34]. As value \(x\) and \(y\) is the individual value of each variables (frequency of rainfall and amount of rainfall) [35].

**Table 2:** The classification of Pearson’s Correlation Coefficient

<table>
<thead>
<tr>
<th>Correlation ((r))</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.5-1.0</td>
</tr>
<tr>
<td>Medium</td>
<td>0.3-0.5</td>
</tr>
<tr>
<td>Low</td>
<td>0.1-0.3</td>
</tr>
</tbody>
</table>

Table 2 show the scale of measurement as to know the relationship whether the two variables are strongly correlated or not To strengthen the values from Pearson, significant level with confidence interval of \(p<0.05\) is used. When \(p\) less than 0.05, null hypothesis which is no relationships between the variables is rejected.

**Smooth Regression:**

Smooth regression can be used for nonparametric regression estimation model which is takes from as Eq 2

\[
y = m(x) + \varepsilon
\]

(2)

Where, \(y\) as response variable, \(x\) the explanatory variable and \(\varepsilon\) as independent error with variance equal to \(\sigma^2\) and zero mean [36]. We want to estimate \(f\), the trend or smooth. Assume the data are ordered so \(x_1 < x_2 < \ldots < x_n\). If we have multiple observations at a given \(x_i\) we introduce a weight \(w\) [37]. Thus, a model of linearity as to look deeper of relationship between the variables, a simple kernel estimator of \(\hat{m}\) (local mean estimator) express as:

\[
\hat{m} = \frac{\sum_{i=1}^{n} w(x_i-x/h)y_i}{\sum_{i=1}^{n} w(x_i-x/h)}
\]

(3)

Where \(w(z; h)\) smooth positive function which is the \(w\) is an arbitrary density or weight function and \(z\) is the bandwidth where as \(|z|\) increase in size gives weight decrease monotonically while \(h\) is smoothing parameter [38].

For a better estimator, for near the edge of the region over which the data been collected, is known as the local linear estimator and this involving least square problem, a estimation start by taking evaluation point \(x\) in the value of \(\alpha\) in Eq 8 as:

\[
\hat{m} = \frac{\sum_{i=1}^{n} w(x_i-x/h)y_i}{\sum_{i=1}^{n} w(x_i-x/h)}
\]

(3)

**Pearson’s Correlation:**

According to Abdi (2007), Pearson’s Correlation Coefficient \((r)\) is a measure of the strength of the association strength of linear dependence between the two variables which in between frequency of rainy days and annual amount of rainfall. As mention by Hamzah et al. (2016) express in Eq.1 where a total of frequency of rainy days in each year were correlated with annual amount of rainfall were used in this study [32]. [33] introduced the following computational procedure:

**Figure 2:** Location of Study Area

Table 1 show the details of the selected station in this study where a daily data and frequency of rainy days were used and calculated.

**Table 1:** Description of weather stations used in the study

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Latitude (°N)</th>
<th>Longitude (°E)</th>
<th>Start Year</th>
<th>End Year</th>
<th>Region Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKM</td>
<td>02° 40'</td>
<td>101° 44'</td>
<td>1985</td>
<td>2011</td>
<td>West</td>
</tr>
<tr>
<td>Porim</td>
<td>02° 54'</td>
<td>101° 47'</td>
<td>1988</td>
<td>2014</td>
<td>West</td>
</tr>
<tr>
<td>Semenyih</td>
<td>03° 04'</td>
<td>101° 52'</td>
<td>1989</td>
<td>2014</td>
<td>West</td>
</tr>
<tr>
<td>Subang</td>
<td>03° 07'</td>
<td>101° 33'</td>
<td>1988</td>
<td>2014</td>
<td>West</td>
</tr>
<tr>
<td>Kuantan</td>
<td>03° 47'</td>
<td>103° 13'</td>
<td>1988</td>
<td>2014</td>
<td>East</td>
</tr>
<tr>
<td>Terengganu</td>
<td>06° 10'</td>
<td>102° 17'</td>
<td>1988</td>
<td>2014</td>
<td>East</td>
</tr>
</tbody>
</table>

**Table 2:** The classification of Pearson’s Correlation Coefficient
these two extension of two dimension is a straightforward,

$$\min_{\alpha, \beta} \sum_{i=1}^{n} (y - \alpha - \beta(x_i - x))^2 w(x_i - x; h)$$  \hspace{1cm} (4)

This smooth function help in order to make sure that between variables there is a smooth relationship which no others factor effecting the relationship [38]. Thus, by taking null hypothesis as linear and alternative as smooth function, if value of p-value less than 0.05, null hypothesis will be rejected and thus vice versa when p-value more than 0.05.

Spatial Distribution:

According to [33], to look over how distribution of rainfall using daily precipitation, a concentration index can be used as contribution of the greatest rainfall of the days to the total amount, is presented using Eq.5

$$Y = aX \exp(bX)$$  \hspace{1cm} (5)

Where adjust the accumulated percentages of precipitation Y contributed by the accumulated percentage of rainy days X. Then, the coefficients a and b are constant and estimated by the least-squares method in

$$\ln a = \frac{\sum x_i^2 \sum \ln y_i + \sum x_i \sum \ln x_i - \sum \sum \sum \ln x_i \ln x_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{N \sum x_i \sum \ln y_i + \sum x_i \sum \ln x_i - N \sum x_i \ln x_i - \sum x_i \sum \ln y_i}{N \sum x_i^2 - (\sum x_i)^2}$$  \hspace{1cm} (6)

Once after constants a and b have been determined, the definite integral of the exponential curve between 0 and 100 is the area under the curve

$$S = \left[ \frac{a}{b} e^{bx} \left( x - \frac{1}{b} \right) \right]$$  \hspace{1cm} (8)

Based on S, the area S’ compressed by the curve, the equidistribution line and X = 100 is the difference between 5,000 and the value of S as in (Figure 3).

$$S' = 5000 - S$$  \hspace{1cm} (9)

From this value of the precipitation concentration, which resembles that of Gini, the normalized daily precipitation concentration index can be defined as follows:

$$CI = \frac{S'}{5000}$$  \hspace{1cm} (10)

Where the CI value is S’ divide by area lower surface of the triangle is bounded by the equidistribution line as shown in Fig.3. Then, the value of CI will categorized as if CI>0.60 is high, 0.55<CI<0.60 is moderate and 0.55<CI is low.

This CI will help in looking out how daily precipitation in area where its display an information of extreme precipitation as this extreme are in close relation to flood event [39].

RESULT AND DISCUSSION

Analysis Of Annual Data:

The relationship between annual rainfall and frequency of rainy days were observed through graph in Fig.4. in general, if the frequency of rainy days increases, the amount of rainfall will also increases. From this graph the correlation of annual rainfall and frequency of rainy days were calculated and the result gave a high positive relationship in-between r = 0.62-0.92 for all six stations which is the value in range of high correlation, r >0.5. Furthermore, the correlation for all stations show significantly in correlation as p-value less than 0.05. Thus, from these result, the total amount of rain that falls should be influenced by the number of rainy days over the year. The purpose of these analysis as to due to climax changes and also alots of flash flood event in western of Peninsular Malaysia. As these can be shown in Segamat, Johor where almost recently around end of Disember 2006, flash flood had hit this town and 16 deaths and as many as 104,023 people were evacuated [40].

<table>
<thead>
<tr>
<th>Station</th>
<th>Pearson’s Correlation (r)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKM</td>
<td>0.71</td>
<td>0.00004</td>
</tr>
<tr>
<td>Porim</td>
<td>0.92</td>
<td>0.0000000004</td>
</tr>
<tr>
<td>Semenyih</td>
<td>0.86</td>
<td>0.000000003</td>
</tr>
<tr>
<td>Subang</td>
<td>0.67</td>
<td>0.00012</td>
</tr>
<tr>
<td>Kuantan</td>
<td>0.62</td>
<td>0.00058</td>
</tr>
<tr>
<td>Terengganu</td>
<td>0.72</td>
<td>0.00003</td>
</tr>
</tbody>
</table>

Daily Precipitation Data: Smooth Regression:

By having a model of linearity test using smooth regression analysis, it help in investigating the evidence of whether there are changes in between relationship of the frequency of rainy days with amount of rainfall [41], [42]. The value of h= 4 as
reference band [43] is used to minimized the cross-validator, give the linearity between the variables. This band in the graph shown in blue color (Fig 5-10). From result in Table 4, show that all stations give p-value >0.05 which is not linearly significant. From these result, the amount of rainfall is not only affected by frequency of rainy days and as expected by [44], precipitation is not a clear indicator of the time compression of precipitation. As mention by [45], precipitation is increasing, there are a variety of ways in which such an increase could have occurred and this also make Malaysia facing the same problem on hydrologic cycle.

<table>
<thead>
<tr>
<th>Stations</th>
<th>P-value</th>
<th>Significant</th>
<th>Linearity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKM</td>
<td>0.94</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Porim</td>
<td>0.76</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Semenyih</td>
<td>0.66</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Subang</td>
<td>0.49</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Kuantan</td>
<td>0.34</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Terengganu</td>
<td>0.35</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Daily Precipitation Data: Spatial Distribution Analysis:
Spatial analysis of rainfall will indicate the different behavior of rainfall pattern between east and west region. In this paper, four out of six stations will accounted for west region while another two is from east region where all stations experiencing a rapid development.

CI is used to measures the irregularity of the rainfall distribution by determining the percentage of rain amount contributed by rainy days in each rainfall amount class interval [25], [39], [46], [47]. Thus, all the stations were calculated as shown in Table 5 using equation 2 to form exponential curves.

<table>
<thead>
<tr>
<th>Station</th>
<th>a</th>
<th>b</th>
<th>CI</th>
<th>Rainfall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKM</td>
<td>0.040</td>
<td>0.032</td>
<td>0.58</td>
<td>69.20</td>
</tr>
<tr>
<td>Porim</td>
<td>0.060</td>
<td>0.002</td>
<td>0.55</td>
<td>65.40</td>
</tr>
<tr>
<td>Semenyih</td>
<td>0.026</td>
<td>0.036</td>
<td>0.62</td>
<td>71.80</td>
</tr>
<tr>
<td>Subang</td>
<td>0.022</td>
<td>0.038</td>
<td>0.68</td>
<td>72.80</td>
</tr>
<tr>
<td>Kuantan</td>
<td>0.015</td>
<td>0.041</td>
<td>0.67</td>
<td>76.20</td>
</tr>
<tr>
<td>Terengganu</td>
<td>0.013</td>
<td>0.042</td>
<td>0.68</td>
<td>77.10</td>
</tr>
</tbody>
</table>

Generally, the value of 25% of rainiest days will contributes to nearly 70% of the total rain [39], [44]. Therefore, most of the research stations experienced high percentage of total rainfall with maximum of 77% of total precipitation within 25% of rainiest day. Kuantan and Terengganu’s stations clearly give high total precipitation that indicates irrelevance of daily amount of rainfall. Whereas, UKM, Porim, Semenyih and Subang’s station give the percentage between 65%-72%. This can be seen in Fig 11, an example of Lorenz Curve for UKM and Porim’s stations on how to get the percentage of total precipitation.

From the previous studies by Suhaila, & Jemain (2012), west region received more regular daily amount compared to the eastern areas with CI values between 0.50 and 0.54. However, the annual CI index resulted in this studies ranges between 0.55-0.68. In terms of geographical area, east region will always receive an irregular rainfall due to heavy rainfall at the end of the year. From this studies that consists of the past 27 years of data, shown that west region also faced irregularity amount of daily rainfall that brought to an extreme rainfall event. This indicates the fact a very few rainy days bring high percentage of annual rainfall.

CONCLUSION
From this analysis, it clearly showing the amount of rainfall is not influenced by the frequency of rainy days. These are proved from all three analysis where smooth regression shows linearly insignificant although Pearson’s correlation r give strongly positive correlation for all station. These result then been supported by CI precipitation test, where all stations shown an irregular distribution. Thus give high rainfall intensity observed at the eastern Peninsular Malaysia with the high rainfall concentration. As these may due to more land use which makes intense to the area that give high concentration index [25]. These rejected the fact from previous research [33] that if the intensity is high, the amount of rainfall are high. These gives a sources understanding on hydrologic cycle has varied and changed. To be strengthened further, again as the result of the study from [45], this paper also found that, rainfall pattern changed in many ways such as increasing in total rainfall are strongly due to increase both in frequency and intensity of heavy and extreme rainfall event. Furthermore, its can be intensity of rainfall had increasing from heavy and extreme rainfall in days only and also from more moderate rainfall causing high proportion of total annual rainfall.

ACKNOWLEDGEMENT
The authors wish to thanks School of Social Science, National University of Malaysia and Malaysian Meteorology Department who have provided us with secondary data and valuable advice.
REFERENCES


