Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 18, 2023, page 62 – 70

Statistical Trend Analysis for Rainfall Data in Pahang and Terengganu, Malaysia

Nurdiana Kamilia Mohd Faizal, Norazlina Ismail*
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: i-norazlina@utm.my

Abstract
The purpose of this research is to perform trend analysis on rainfall data from Peninsular Malaysia's east coast. This study used data from four rainfall stations in Pahang and Terengganu, spanning the years 1980 to 2021. The descriptive analysis was completed prior to the trend analysis. The mean, standard deviation, coefficient of variation, skewness, and kurtosis are the five main indices of descriptive analysis. From the results of the descriptive analysis, only one station had the lowest value of every index mentioned. Mann-Kendall and modified Mann-Kendall test were used to study trend analysis. To evaluate the performance of the methods used, the study was divided into two sections to evaluate the performance of the methods used: monthly and annual trend analysis. In monthly trend analysis, Mann-Kendall and modified Mann-Kendall test results show that only one station does not meet the criteria of trend existence. Nonetheless, there are a total of three stations in the Mann-Kendall test and only one station in the modified Mann-Kendall test that do not have an annually trend. The modified Mann-Kendall test yields more precise and comprehensive results when long-term persistence (LTP) is taken into account. Despite the fact that one of the stations is statistically rejected, all stations show a positive magnitude of trend in all monthly and annual rainfall indices. In the trend analysis, the only station with the lowest standard deviation in the descriptive analysis is also the only station that does not show trend significance.

Keywords: Rainfall; Descriptive Analysis; Mann-Kendall; modified Mann-Kendall; Trend

1. Introduction

Rainfall and temperature are frequently used as weather and climate parameters to predict global climate change. Over the period 1880-2012, the global temperature increased by the range of 0.65-1.06°C according to the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) (IPCC 2014). Across Southeast Asia, temperature has been rising at a rate of 0.14-0.20 °C per decade, accompanied by an increase in the number of hot temperatures and summer nights, and a decrease in colder seasons since the 1960s, (Sa'adi et al., 2019). It was discovered that rising sea surface temperatures as a result of climate change have resulted in significant changes in rainfall and rainfall extremes (Trenberth, 2011). Seeing as tropical regions are more vulnerable to natural disasters, with strong geographical diversity, time and frequency variability of climate plays an important role of many ecological and environmental mechanisms.

Malaysia's climate is distinguished by its consistent temperature, humidity levels, and abundant rainfall. The winds are usually light. Even during times of severe lack of rain, it is exceptionally uncommon to have a comprehensive day with particularly clear skies in the sub - tropical downturn. On the other hand, exceptional during the northeast monsoon seasons, it is uncommon to have several days without a sunlight. Rainfall distribution patterns across the country are determined by seasonal wind flow patterns in conjunction with local topographic features. Heavy rains are expected in exposed areas such as

Peninsular Malaysia's east coast, Western Sarawak, and Sabah's northeast coast during the northeast monsoon season. Meanwhile, inland locations or zones protected by hills and mountains are relatively free of its involvement.

Peninsular Malaysia's east coast, for example, is more susceptible to climate change than the west. Floods caused by heavy rainfall have become an almost integral part of modern society in Peninsular Malaysia's east coast states. For instance, Pahang, the nation's biggest and densely populated state, was stunned by unimaginable floods, the most devastating in half a decade, on December 19, 2022. The weekend's heavy downpours ended up causing waterlogging, flash floods towns and villages and rendering access streets inaccessible, trapping many vehicle owners in their vehicles for hours. In Terengganu, the flood situation worsened as the number of victims rose sharply to 11,415, compared with 3,785 in one night. The continual rain also submerged several areas with water levels soaring to lethal levels.

Descriptive analysis will be used in this study to determine the characteristics of rainfall distribution. The Mann-Kendall tests are used in the exploratory analysis to examine the trend of rainfall data. In comparison, the modified Mann-Kendall test was used to examine the differences when natural variability was factored into the calculation. As a result, the goal of this study is to investigate the rainfall pattern using exploratory and descriptive analysis of rainfall data. In addition, the Mann-Kendall and modified Mann-Kendall tests must be used to determine changes in the monthly and annual rainfall trend. Also, the Mann-Kendall and modified Mann-Kendall tests were used to compare the trend.

2. Materials and methods

2.1 Data

The rainfall data used in this study are in the form of daily rainfall amounts, which are then classified as monthly and annual rainfall amounts. The rainfall data are obtained from the Department of Irrigation and Drainage Malaysia and the Malaysian Metrological Department for the period of 20 to 40 years ranging from 1980 to 2021. Four rainfall stations from the states of Pahang and Terengganu have been chosen for the evaluation process.

2.2 Descriptive statistics

Descriptive statistics provide a simple summary of the sample data series. This summary consists of a quantitative analysis which is the summary of descriptive statistics or a graphical interpretation of the data. This section will explain the mathematical formulation to obtain the sample mean, coefficient of variation, statistical dispersion values that consist of the standard deviation and the skewness and kurtosis of the data.

Mean,	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$	(1)
Standard deviation,	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$ where n is the number of observations and μ is the mean of the sample.	(2)

Coefficient of variation,	$CV(\%) = \frac{\sigma}{\mu} \times 100\%$ where σ is the standard deviation and μ is the mean of the sample.	(3)
Skewness,	$y = \frac{E(x-\mu)^3}{\sigma^3}$ where μ is the mean, σ is the standard deviation and E is the expectation operator.	(4)
Kurtosis,	$k = \frac{E(x-\mu)^4}{[E(x-\mu)^2]^2} = \frac{\mu_4}{\sigma^4}$ where E represents the expectation operator, σ is the standard deviation, μ is the mean and μ_4 is the 4th moment of the mean.	(5)

2.3 Mann-Kendall test

The nonparametric Mann-Kendall (MK) statistical test, also called Kendall's tau test (Mann 1945; Kendall 1975), has been applied in many studies to identify whether monotonic trends exist in hydro-meteorological data such as temperature, rainfall, and streamflow. This test is often used because of its property that no assumptions are needed about the data that need to be tested. In the trend test, the null hypothesis H_0 is that there is no trend in the population from which the dataset is drawn and the sample of data $\{x_j, j = 1,2,\dots,n\}$is independent and identically distributed. The alternative hypothesis H_1 is that a trend exists in the dataset. The test statistic, Kendall's S, is defined as follows:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} Sign\ (x_j - x_k) \tag{6}$$

where x_j and x_k are the sequential data values, n is the length of the dataset, and

$$Sign\ (x) = \begin{cases} 1 & \text{if x} > 0 \\ 0 & \text{if x} = 0 \\ -1 & \text{if x} < 0 \end{cases} \tag{7}$$

Under the null hypothesis, the statistic S is approximately normally distributed when $n \geq 8$ with zero mean and the variance is given as follows:

$$Var\ (S) = \frac{n(n-1)(2n+5) - \sum_t^n t(t-1)(2t+5)}{18} \tag{8}$$

where t is the extent of any given tie and denotes $\sum_t$ the summation over all ties.

The standardized test statistic Z is computed by

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var\ (S)}} & \text{for } S > 0 \\ 0 & \text{for } S = 0 \\ \frac{S+1}{\sqrt{Var\ (S)}} & \text{for } S < 0 \end{cases} \tag{9}$$

The significance of the trend is measured by using the value of Z. It follows a standard normal distribution. In a two-sided test for trend, H_0 should be accepted if $|Z| \leq 1.645$ at the 0.10 level of significance.

Otherwise, it is rejected if it is greater than the critical value of a chosen significance level. A positive Z value indicates an upward trend, whereas a negative one indicates a downward trend.

2.4 Modified Mann-Kendall test

In modified Mann-Kendall tests, the equivalent normal variants of rank of the de-trended series are obtained using the following equation:

$$Z_i = \phi^{-1}\left(\frac{R_i}{n+1}\right) \quad for\ i = 1 : n \tag{10}$$

where R_i is the rank of the de-trended series x_i^l, n is the length of the time series, and ϕ^{-1} is the inverse standard normal distribution function which is the mean is 0 and the standard deviation is 1.

The scaling coefficient or Hurst coefficient, H is obtained by maximizing log likelihood function in McLeod and Hipel (1978). This estimate of H is approximately normally distributed for the uncorrelated case when true H is 0.5. The correlation matrix for a given Hurst coefficient, H is derived using the following equation:

$$C_n\ (H) = \left[\rho_{|j-i|}\right], \quad \text{for } i = 1 : n, \qquad j = 1 : n, \tag{11}$$

$$\rho_l = \frac{1}{2}(|l+1|^{2H} - 2|l|^{2H} + |l-1|^{2H}), \tag{12}$$

where ρ_l is the autocorrelation function of lag l for a given H, and is independent of the time scale of aggregation for the time series.

The value of H is obtained by maximizing the log likelihood function of H as given below:

$$\log L(H) = -\frac{1}{2}\log|C_n\ (H)| - \frac{Z^\tau [C_n\ (H)]^{-1} Z}{2\gamma_0} \tag{13}$$

where $C_n\ (H)$ is the determinant of correlation matrix $[C_n\ (H)]$, Z^τ is the transpose vector of equivalent normal variates $Z[C_n\ (H)] - 1$ is the inverse matrix, and γ_0 is the variance of Z_i. Equation (7) can be solved numerically for different values of H, and the value for which $\log L(H)$ is maximum is taken as the H value for the given time series x_i. In this study, the value of H is solved between 0.50 and 0.98 with an incremental step of 0.01.

A significance level of H is determined using the mean (μH) and standard deviation (σH) when $H = 0.5$ (normal distribution) as given by the following equations (Hamed 2008):

$$\mu_H = 0.5 - 2.87 n^{-0.9067}, \tag{14}$$

$$\sigma_n = 0.7765 n^{-0.5} - 0.0062 \tag{15}$$

The significance of H is determined using μH and σn in Eq. (9) and (10). In this study, 5% and 10% significance levels were used for rainfall to determine the significance of H. The Hurst coefficient provides a measure for long term memory or persistence in the series. The method proposed by Hamed (2008) was used to find the significance of H value for the data in this study. If H is found to be significant, the variance of S is calculated using the following equation for given H:

$$V(S)^{H'} = \sum_{i<j} \cdot \sum_{k<1} \frac{2}{\pi} sin^{-1}\left(\frac{\rho|j-i| - \rho|i-l| - \rho|j-k| + \rho|i-k|}{\sqrt{(2-2\rho|i-j|)(2-2\rho|k-l|)}}\right) \tag{16}$$

where ρ_l is calculated using Eq. (7) for given H and $V(S)^{H'}$ is the biased estimate. The unbiased estimate

(17)

$V(S)^H$ is calculated by multiplying by a bias correcting factor B as follows.

$$V(S)^H = V(S)^{H'} \times B,$$

where B is a function of H as shown below.

$$B = a_0 + a_1 H + a_2 H^2 + a_3 H^3 + a_4 H^4 \quad (18)$$

The coefficients, a_0, a_1, a_2, a_3, and a_4 in Eq. (18) are functions of the sample size n. The values of the coefficients can be found in Hamed (2008). The significance of the Mann Kendall test is computed using $V(S)^H$ in place of $V(S)$ in Eq. (7).

3. Results and discussion

This chapter describes the findings of the analysis of both annual and seasonal rainfall throughout the east coast region of Peninsular Malaysia. The results are separated into two subchapters: descriptive analysis and trend analysis. Trend analysis consists of two topics which are monthly rainfall analysis and annual rainfall analysis.

3.1 Descriptive analysis

Statistical values such as mean, standard deviation, coefficient of variation (CV), skewness and kurtosis are obtained to be a guide towards this study.

Table 1: Monthly descriptive statistics for rainfall stations

Station Name	Mean (mm)	Std. Dev. (mm)	CV	Skewness	Kurtosis
Sungai Lembing PCCL Mill	7.2022	5.9547	82.6798	2.0187	7.2777
Kampung Sungai Yap	5.1427	3.4963	67.9850	0.9522	2.2264
Kampung Dura	8.6416	9.0263	104.4521	2.2548	6.0854
Al-Muktafi Billah Shah	8.9874	8.9631	99.7291	2.8151	12.3374

The highest mean value is 8.9874mm for Al-Muktafi Billah Shah station. This station indicated that the standard deviation is 8.9631mm and the data skewed to right as the skewness value is positive. However, the highest standard deviation was recorded in Kampung Dura station with a value of 9.0263mm which states that the rainfall distribution was highly dispersed in the rainfall pattern.

Table 2: Annual descriptive statistics for rainfall stations

Station Name	Mean (mm)	Std. Dev. (mm)	CV	Skewness	Kurtosis
Sungai Lembing PCCL Mill	7.2254	2.3965	33.1679	-0.7585	1.7377
Kampung Sungai Yap	5.1543	1.5236	29.5588	-1.1413	2.3431
Kampung Dura	8.6669	2.7257	31.4494	0.0409	1.4244
Al-Muktafi Billah Shah	9.0219	2.8249	31.3122	-0.2757	0.6103

The greatest mean value for Al-Muktafi Billah Shah station is 9.0219mm. The standard deviation at this station is 2.8249mm, and the data is skewed to the left because the skewness value is negative. Because the kurtosis value is less than three, the coefficient of variation (CV) is the second highest among all of the stations with flatter distributions. In contrast, the smallest standard deviation was determined at the Kampung Sungai Yap station, with a value of 1.5236mm, indicating that the rainfall distribution was evenly distributed in the rainfall pattern.

3.2 Trend analysis

3.2.1 Monthly rainfall analysis

The results of the Mann-Kendall test and modified Mann-Kendall test for monthly rainfall analysis are shown below.

Table 3: Monthly total amount of rainfall for Mann-Kendall test

Station Name	Variance	τ value	p-value
Sungai Lembing PCCL Mill	0.1426×10^{8}	0.1099	0.2325×10^{-3}
Kampung Sungai Yap	0.1426×10^{8}	0.0315	0.2922
Kampung Dura	0.1426×10^{8}	0.1177	0.8137×10^{-4}
Al-Muktafi Billah Shah	0.1426×10^{8}	0.0854	0.4222×10^{-2}

The monthly total amount of rainfall for Mann-Kendall test shows that the p-values for all the stations except Kampung Sungai Yap are smaller than 0.05. This concludes that Sungai Lembing PCCL Mill, Kampung Dura, and Al-Muktafi Billah Shah stations have the proof of the significant trend existence with an increasing trend. The variance values for all stations are the same with each other which is 0.1426×10^{8}.

Table 4: Monthly total amount of rainfall for modified Mann-Kendall test

Station Name	New variance	τ value	p-value
Sungai Lembing PCCL Mill	0.4328×10^{8}	0.1097	0.2363×10^{-10}
Kampung Sungai Yap	0.1104×10^{8}	0.0314	0.2313
Kampung Dura	0.5628×10^{7}	0.1174	0.3553×10^{-9}
Al-Muktafi Billah Shah	0.5329×10^{7}	0.0853	0.2857×10^{-5}

Kampung Sungai Yap station is the only station that has p-value greater than 0.05 with 0.2313. By that, the stations that have p-value smaller than 0.05 which are Sungai Lembing PCCL Mill, Kampung Dura, and Al-Muktafi Billah Shah stations indicates that there is a significant trend in the areas. The states also have an increasing trend with positive τ value.

The clear difference between Mann-Kendall and modified Mann-Kendall tests is the variance values. For the Mann-Kendall test, all the variance values are approximately the same for all of the stations. Yet, the variance values for each station differ from each other in modified Mann-Kendall test.

3.2.2 Annual rainfall analysis

The results of the Mann-Kendall test and modified Mann-Kendall test for annual rainfall analysis are shown below.

Table 5: Annual total amount of rainfall for Mann-Kendall test

Station Name	Variance	τ value	p-value
Sungai Lembing PCCL Mill	0.8514×10^4	0.1986	0.06542
Kampung Sungai Yap	0.8514×10^4	0.02207	0.8453
Kampung Dura	0.8514×10^4	0.2985	0.0055
Al-Muktafi Billah Shah	0.8514×10^4	0.1312	0.2248

The annually total amount of rainfall for Mann-Kendall test shows that the p-values for all the stations except Kampung Dura are larger than 0.05. This concludes that Kampung Dura station has proof of the significant trend existence with an increasing trend. Nevertheless, Sungai Lembing PCCL Mill, Kampung Sungai Yap and Al-Muktafi Billah Shah stations do not have trend with p-value more than 0.05. The variance values for all stations are the same with each other which is 0.8514×10^4.

Table 6: Annual total amount of rainfall for modified Mann-Kendall test

Station Name	New variance	τ value	p-value
Sungai Lembing PCCL Mill	0.1925×10^4	0.1986	0.1067×10^{-3}
Kampung Sungai Yap	0.2397×10^4	0.02207	0.7132
Kampung Dura	0.2559×10^4	0.2985	0.4171×10^{-6}
Al-Muktafi Billah Shah	0.3102×10^4	0.1312	0.0443

Kampung Sungai Yap station is the only station that has p-value greater than 0.05 with 0.7132. By that, the stations that have p-value smaller than 0.05 which are Sungai Lembing PCCL Mill, Kampung Dura, and Al-Muktafi Billah Shah stations indicates that there is a significant trend in the areas. All the states have an increasing trend with positive τ value.

For the Mann-Kendall test, all the variance values are approximately the same for all the stations. Nevertheless, some of the stations do not achieve 5% significance level of p-value. Yet, the variance values for each station differ from each other in modified Mann-Kendall test. Also, only one station that is rejected due the p-value more than 0.05.

Conclusion

According to monthly descriptive analysis, the highest monthly rainfall is found at Al-Muktafi Billah Shah station. The highest standard deviation was recorded at the Kampung Dura station. The maximum average for annual descriptive analysis was measured at the Al-Muktafi Billah Shah station. In contrast, the smallest standard deviation was determined at the Kampung Sungai Yap station. In monthly rainfall analysis, the Mann-Kendall test showed that Sungai Lembing PCCL Mill, Kampung Dura, and Al-Muktafi Billah Shah stations have a significant increasing trend. The modified Mann-Kendall test confirmed the above findings by displaying the exact output. For annual rainfall analysis, results from Mann-Kendall tests conclude that Kampung Dura station has the proof of a significant trend existence. Ironically, modified Mann-Kendall test showed the stations of Sungai Lembing PCCL Mill, Kampung Dura, and Al-Muktafi Billah Shah stations indicates that there is a significant trend. The variance values and p-values clearly differ between the Mann-Kendall and modified Mann-Kendall tests, with some stations failing to achieve the 5% significance level of p-value. To summarize the findings of the study, the only one station which is Kampung Sungai Yap, located in the state of Pahang does not satisfy the requirement of trend existence with the lowest standard deviation. For recommendations of this study, the use of various rainfall indices for describing the main characteristics of rainfall and seasonal factors in the methodology may be beneficial in determining the rainfall characteristics that influence trend existence.

References

[1] Bari, S. H., Rahman, M. T. U., Hoque, M. A., & Hussain, M. M. (2016). Analysis of seasonal and annual rainfall trends in the northern region of Bangladesh. Atmospheric Research, 176–177, 148–158. https://doi.org/10.1016/j.atmosres.2016.02.008

[2] Bojago, E., & Yaya, D. (2021a). Trend analysis of seasonal rainfall and temperature pattern in Damota Gale districts of Wolaita Zone, Ethiopia. https://doi.org/10.21203/rs.3.rs-757668/v1

[3] Hamed, K. H. (2008). Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. Journal of Hydrology, 349(3–4), 350–363. https://doi.org/10.1016/j.jhydrol.2007.11.009

[4] Jayawardene, H. K. W. I., Sonnadara, D. U. J., & Jayewardene, D. R. (2005). Trends of Rainfall in Sri Lanka over the Last Century. In Sri Lankan Journal of Physics (Vol. 6).

[5] Khan, N., Pour, S. H., Shahid, S., Ismail, T., Ahmed, K., Chung, E. S., Nawaz, N., & Wang, X. (2019). Spatial distribution of secular trends in rainfall indices of Peninsular Malaysia in the presence of long-term persistence. Meteorological Applications, 26(4), 655–670. https://doi.org/10.1002/met.1792

[6] Mahmoud M. Smadi, A. Z. (2006). smadi2006. A Sudden Change In Rainfall Characteristics In Amman, Jordan During The Mid 1950s.

[7] Nalley, D. (2012). Analyzing Trends In Temperature, Precipitation And Streamflow Data Over Southern Ontario And Quebec Using The Discrete Wavelet Transform.

[8] Olusegun Mayowa, O., Hadi Pour, S., Shahid, S., Mohsenipour, M., bin Harun, S., Heryansyah, A., & Ismail, T. (2015). Trends in rainfall and rainfall-related extremes in the east coast of peninsular Malaysia.

[9] Ongoma, V., & Chen, H. (2017). Temporal and spatial variability of temperature and precipitation over East Africa from 1951 to 2010. Meteorology and Atmospheric Physics, 129(2), 131–144. https://doi.org/10.1007/s00703-016-0462-0

[10] Pal, A. B., Khare, D., Mishra, P. K., & Singh, L. (2017). Trend Analysis Of Rainfall, Temperature And Runoff Data: A Case Study Of Rangoon Watershed In Nepal. International Journal of Students' Research in Technology & Management, 5(3), 21–38. https://doi.org/10.18510/ijsrtm.2017.535

[11] Pour, S. H., Harun, S. bin, & Shahid, S. (2014). Genetic programming for the downscaling of extreme rainfall events on the east coast of peninsular Malaysia. Atmosphere, 5(4), 914–936.

https://doi.org/10.3390/atmos5040914

[12] Rahman, M. A., Yunsheng, L., & Sultana, N. (2017). Analysis and prediction of rainfall trends over Bangladesh using Mann–Kendall, Spearman's rho tests and ARIMA model. Meteorology and Atmospheric Physics, 129(4), 409–424. https://doi.org/10.1007/s00703-016-0479-4

[13] Sa'adi, Z., Shahid, S., Ismail, T., Chung, E. S., & Wang, X. J. (2019a). Trends analysis of rainfall and rainfall extremes in Sarawak, Malaysia using modified Mann–Kendall test. Meteorology and Atmospheric Physics, 131(3), 263–277. https://doi.org/10.1007/s00703-017-0564-3

[14] Sonali, P., & Nagesh Kumar, D. (2013). Review of trend detection methods and their application to detect temperature changes in India. Journal of Hydrology, 476, 212–227. https://doi.org/10.1016/j.jhydrol.2012.10.034

[15] Suhaila, J., Mohd Deni, S., Zawiah Wan Zin, W., & Aziz Jemain, A. (2010). Trends in Peninsular Malaysia Rainfall Data During the Southwest Monsoon and Northeast Monsoon Seasons. In Sains Malaysiana (Vol. 39, Issue 4).

[16] Trenberth, K. E. (2011). Changes in precipitation with climate change. Climate Research, 47(1–2), 123–138. https://doi.org/10.3354/cr00953

[17] Wong, C. L., Venneker, R., Uhlenbrook, S., Jamil, A. B. M., & Zhou, Y. (2009). Variability of rainfall in Peninsular Malaysia Variability of rainfall in Peninsular Malaysia. In Hydrol. Earth Syst. Sci. Discuss (Vol. 6). www.hydrol-earth-syst-sci-discuss.net/6/5471/2009/

[18] Yao, C., Qian, W., Yang, S., & Lin, Z. (2010). Regional features of precipitation over Asia and summer extreme precipitation over Southeast Asia and their associations with atmospheric-oceanic conditions. Meteorology and Atmospheric Physics, 106(1), 57–73. https://doi.org/10.1007/s00703-009-0052-5

[19] Yue, S., & Wang, C. (2004). The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. In Water Resources Management (Vol. 18).