

Statistical Analysis of Rainfall Data in Peninsular Malaysia using Strong Transformation of Cubic Spline Analysis

(Analisis Statistik Data Hujan di Semenanjung Malaysia menggunakan Transformasi dengan Analisis Splin Kubik)

Nur Arzilah Ismail^{a,b} & Wan Zawiah Wan Zain^b

^aJabatan Pendidikan Kejuruteraan, Fakulti Kejuruteraan dan Alam Bina, Universiti Kebangsaan Malaysia, Malaysia

^bJabatan Sains Matematik, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, Malaysia

Corresponding author; azila@ukm.edu.my

Received: 16 July 2022, Received in revised form 7th August 2022

Accepted 15th August 2022, Available online 15th November 2022

ABSTRACT

The constant rainfall is important for equatorial regions. The extreme rainfall shows our nature is facing climate change significantly and this needs to be paid attention by researchers hence the effects of disasters can be avoided as early as possible. Therefore, it is important to understand the rain pattern for the existing data so that forecasting can be done. This article discussed how to transform abnormal data into normal data and then analyze the data using the spline cubic analysis method. The study found that by using the transformation algorithm, the best r value is obtained. Cubic spline analysis shows that there are 30 rainfall stations in peninsular Malaysia where there are significant changes in rainfall patterns.

Keywords: Cube spline; log transformation; time series; rainfall data; seasonal pattern

ABSTRAK

Taburan hujan yang sekata dan malar adalah penting bagi kawasan khatulistiwa. Keadaan taburan hujan yang ekstrem menunjukkan alam ini menghadapi perubahan iklim dengan ketara dan ini perlu diberikan perhatian oleh para penyelidik supaya kesan bencana dapat dielakkan dengan lebih awal. Oleh itu adalah penting untuk melihat corak hujan bagi data sedia ada dan seterusnya ramalan dapat dilakukan. Kajian ini adalah untuk menganalisis kaedah transformasi data tidak normal menjadi data normal seterusnya menganalisis data tersebut dengan menggunakan kaedah analisis kubik splin. Kajian mendapati dengan menggunakan algoritma transformasi, nilai r yang paling baik diperolehi. Analisis kubik splin menunjukkan terdapat 30 stesen hujan di Semenanjung Malaysia yang berlakunya perubahan corak hujan yang signifikan

Kata kunci: Spline kiub; transformasi log; siri masa; data taburan hujan; corak bermusim

INTRODUCTION

The amount of water that falls as rain over the course of a week or a month is referred to as a rainfall. Since amounts may vary between sites and times, rainfall is estimated by gathering rainwater over a wide range of locations and times (Kwan *et. al.* 2013). Tropical weather prevails in Malaysia where the average yearly temperature in Malaysia is 25.4 °C. The average monthly temperature varies just slightly from season to season, fluctuating by one degree Celsius between a lowest of 24.9°C in January and a maximum of 25.9°C in May. These three months—April, May, and June—are the hottest of the year. Additionally, there is still a lot of rain, with an average annual rainfall of 3,085.5 millimetres (mm).

Along with remaining fairly consistent throughout the year, the average monthly precipitation ranges from about 200 mm in June and July to 350 mm throughout November and December. The Southwest Monsoon (April to September) and the Northeast Monsoon are the two monsoon seasons (October-March). Malaysia experiences roughly six hours of strong sunlight each day, with the likelihood of cloud cover increasing in the late afternoon and evening (Malaysia Meteorology Department, Annual Report)

To lessen the detrimental effects of extreme rainfall events, an accurate rainfall prediction is necessary. The ensemble technique, which aims to capture the uncertainty brought on by several elements, is one of the attempts to produce an accurate rainfall prediction (Jibrillaha 2018). As

has been done by many climate centres in developed nations, many scenarios are required to simulate future weather and climate projection using large-scale data. Hence, this article provides transformation technique to normalize the data and rainfall pattern is observed for 17 years period in 91 rain stations in Peninsular Malaysia. A study on rainfall during the southwest and northeast monsoons in Peninsular Malaysia showed that for most stations, the rainfall for the number of wet days has decreased significantly during extreme events (Jamrozik 2010). Therefore, it is important to take these two monsoons into account in studying climate change.

STUDY AREA & DATA CURATION

Rainfall is not only vital for life on our planet but is also a possible determinant of climate change (Kwan 2013). We illustrate methods for graphing daily rainfall totals reported from 91 meteorological stations (JPS) in the West Malaysia peninsula and using cubic spline functions to smooth and assess rainfall change over the 17 years from March 2000 to February 2017 inclusive. Figure 1 shows the distribution location of rainfall stations all over Peninsular Malaysia.



FIGURE 1. Distribution of Rainfall Stations in Peninsular Malaysia by State

Most meteorological stations in Malaysia have incomplete data and are mostly located in more densely populated areas (Mildrexler, 2011). These 91 stations are those that have the most complete data and cover the whole study region.

Figure 2 shows the most extreme data observed at station Semeling in the state of Kedah. The Q-Q plot observed high residual over the linear line. Hence logarithm transformation is introduced, and the smooth Q-Q plot data is observed. It is worth mentioned that the data are still skewed, so we use a stronger transformation, further shrinking value 10 mm by 50%.

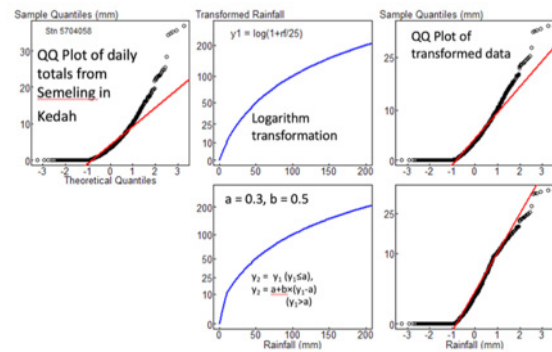


FIGURE 2. Quantile-Quantile (Q-Q) Plots

Figure 3 shows the seasonal patterns of North-East North-West coast stations. For the duration of 17 years. Years are divided into 52 periods each of duration 7 days, except for 8 days from January 1-8 and from days 359-366 in leap years. Results in Figure 3 portrayed the transformation data that fit into Logarithm transformation.

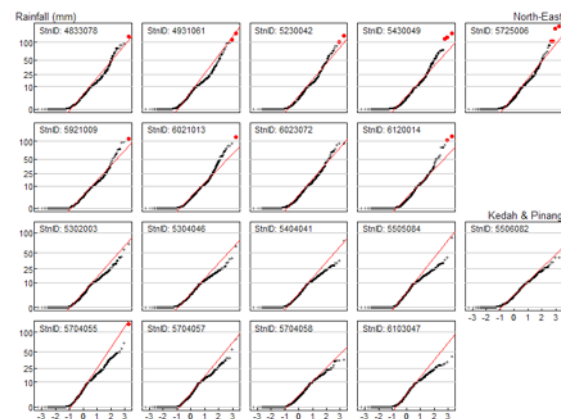


FIGURE 3. Plots Q-Q Plots for North-East & North-West Coast Stations

METHODOLOGY

The cubic spline function is proposed to be applied in this study because there is not much change in the seasonal pattern of rainfall data. The cubic spline function is one of the semi-parametric approaches in which one of the general assumptions is that both the response and predictor variables are continuous, therefore can be used to obtain a time series in a smooth and continuous form between each data point. A detailed description and procedure of the method used is given in (Wongsai, 2017). The cubic spline function is given in the following equation:

$$s(t) = a + bt + \sum_{k=1}^p c_k (t - t_k)_+^3 \tag{1}$$

where t is a time $t_1 < t_2 < \dots < t_p$ in which is $(t - x)_+$ a knot for $(t - x)$ given $t > x$ and otherwise 0. The smoothness of the connection between the points must

satisfy the continuity of the first and second derivatives of the above function given in the following equation:

$$s'(t) = bt + 3 \sum_{k=1}^p c_k (t - t_k)_+^2 \quad (2)$$

$$s''(t) = b + 6 \sum_{k=1}^p c_k (t - t_k)_+ \quad (3)$$

There are many types of cubic spline, but this study uses the original cubic spline function with the assumption that the second derivative is equal to zero at the end point as given in equation 4. To make the quadratic coefficient and the cubic spline function to be 0 for $t > T_0$ and $t < T_p$, a special annual periodic boundary conditions is required which is given in Equations 4 and 5, where t_1 and t_2 are the locations of the first knot and the last knot respectively.

$$\sum_{k=1}^p c_k = 0 \quad (4)$$

$$\sum_{k=1}^p c_k t_k = 0 \quad (5)$$

However, there is a problem that most commonly arises in using this method which is to identify the appropriate number of knots and their location. The number and location of knots is determined by the user and it depends on changing rain patterns. Changes in rainfall patterns imply a high number of knots and vice versa (Jamrozik, 2010). Therefore, the determination of the number of knots is important because this increases the accuracy of the estimate. Several assumptions were considered to allocate knots at the beginning and end of the year and the stability of the data to maintain seasonal patterns between years. In addition, when studying modeling, there are several assumptions that need to be emphasized. In general, when data is taken from the same source, there will be a correlation between the data, or in other words, we call it autocorrelation. One of the assumptions that must be satisfied in modeling is that the correlation between data points must be independent. Therefore, in this study the autocorrelation was adjusted in the cubic spline function.

Figure 4 shows the cubic splines analysis after the data is transformed. Figure 4 exhibit the seasonal patterns of daily rainfall in North-East and North-West Malaysia for March 2000 to February 2017. It is highly expected that the cubic spline analysis enables to estimate the seasonal patterns and the smooth median is observed for the northern region of Peninsular Malaysia. Using four equally spaced knots, Figure 5 shows the time series trends estimated by fitting cubic splines to season-adjusted weekly averages.

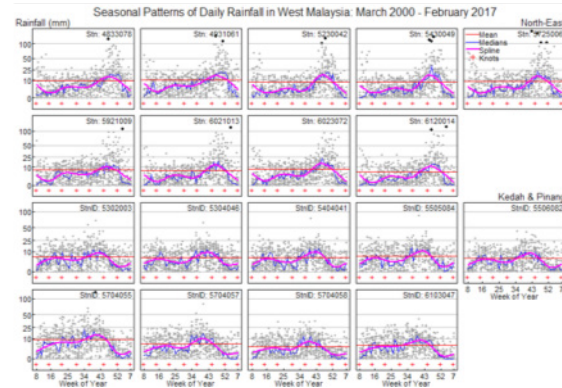


FIGURE 4. Seasonal Patterns of Daily Rainfall in West Malaysia

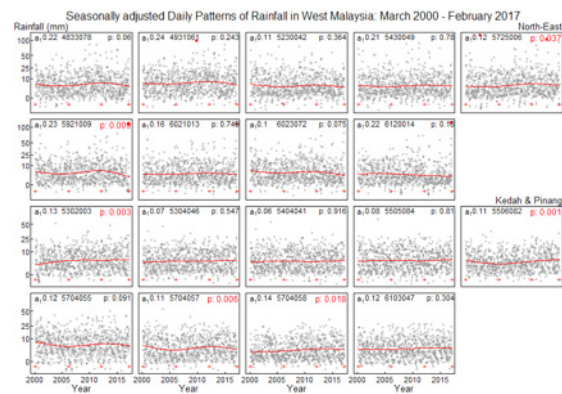


FIGURE 5. Time Series Trends: Rainfall in West Malaysia (March 2000 - February 2017)

Figure 6 shows the statistical model of natural cubic spline function. The statistical model is assumed to be a natural cubic spline function with four equally spaced knots spanning the period and errors driven by a single-parameter auto-regression. P-values test the null hypothesis that rainfall at a station did not change over the 17-year period.

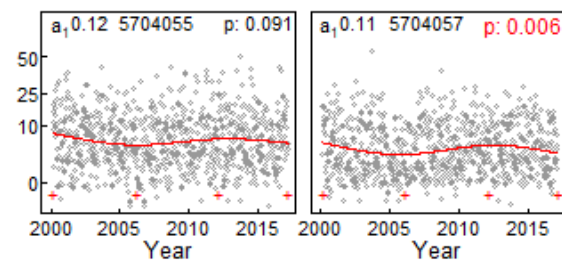


FIGURE 6. Statistical Model for Natural Cubic Spline Function (Northern Region)

Based on results in Figure 1 till Figure 6, the repetition of the Logarithm transformation and Cubic Spline analysis are done for the rest of the data in Peninsular Malaysia. Based on the statistical model, Figure 7 is produced. It is noticed that, based on Figure 7, out of the 91 stations in the sample, 30 had p-values below 0.05. If the rainfall pattern did not change overall, the expected number of stations with p-values less than 0.05 would be 0.05×91 , that is, 4.5. So, the rainfall pattern in the West Malaysia Peninsular changed over the 17 years from March 2000 to February 2017

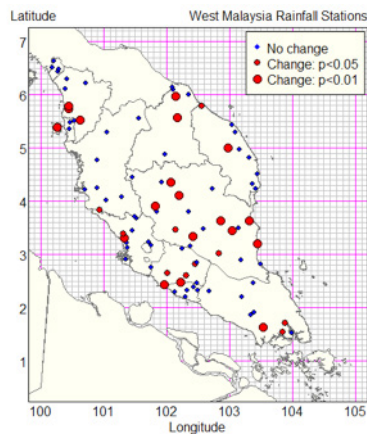


FIGURE 7. Statistical Model for Natural Cubic Spline Function (all Peninsular Malaysia)

CONCLUSION

It is observed that the pattern of the rainfall data exhibited non normal observation and was skewed. However strong Logarithm transformation is required to successful convert the data into normal distribution. The usage of cubic spline analysis enables to estimate the seasonal pattern with smooth medians. The overall statistical model indicates a major change in rainfall pattern where 30 stations record significant values less than 0.05, thus indicating the climate change warning. Hence the major change in rainfall pattern in almost 20 years duration could alarm us about evolutionary nature which is similarly conclude by Pin et al, 2013, Suhaila, 2010, and Tangang, 2012.

REFERENCES

- Jamrozik, J., Bohmanova, J., and Schaeffer, L.R. 2010. Selection of locations of knots for linear splines in random regression test-day models. *J Anim Breed Genet.* 127(2): 87-92.
- Jibrillaha, A.M., Jaafar, M., and Choy, L.K. 2018. Climate Change Awareness and Adaptations Among the Farming and Animal Rearing Communities of the Central Sokoto Close-Settle Zone, North-Western Nigeria. *Jurnal Kejuruteraan SI* 1(6): 71-79.
- Kwan, M.S., Tangang, F.T., and Juneng, L. 2013. Projected Changes of Future Climate Extremes in Malaysia. *Sains Malaysiana* 42(8): 1051–1059.
- Malaysia Meteorology Department Annual Report 2016. Malaysia: [accessed 27 Jun 2018]. Available: <http://www.met.gov.my>
- Mildrexler, D.J., Zhao, M., and Running, S.W. 2011. A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *J Geophys Res.* 116: 1-15.
- Pin, K.F., Pereira, J.J., and Aziz, S. 2013. Platforms of climate change: An evolutionary perspective and lessons for Malaysia. *Sains Malaysiana* 42(8): 1027–1040.
- Suhaila, J., Deni, S.M., Wan Zin, W.Z., and Jemain, A.A. 2010. Trends in Peninsular Malaysia rainfall data during the southwest monsoon and northeast monsoon seasons: 1975–2004. *Sains Malaysiana* 39(4): 533–542.
- Tangang, F.T., Juneng, L., Salimun, E., Sei, K.M., Le, L.J., and Muhamad, H. 2012. Climate change and variability over Malaysia: Gaps in science and research information. *Sains Malaysiana* 41(11): 1355–1366.
- Tomlinson, C. J., Chapman, L., Thornesb, J.E., and Bakera, C. 2011. Remote sensing land surface temperature for meteorology and climatology: a review. *Meteorol Appl.* 18: 296–306.
- Wongsai, N., Wongsai, S., and Huete, A.R. 2017. Annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime MODIS LST data. *Remote Sens.* 9: 1254.