Jurnal
Teknologi

Full Paper

HOMOGENEITY ANALYSIS OF RAINFALL IN KELANTAN, MALAYSIA

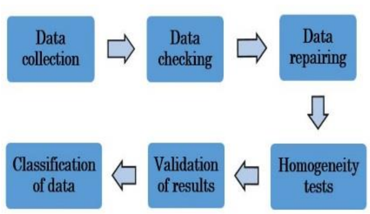
Article history
Received
15 July 2015
Received in revised form
2 August 2015
Accepted
26 August 2015

Ng Jing Lin[a*], Samsuzana Abd Aziz[a], Huang Yuk Feng[b], Aimrun Wayayok[a], Md Rowshon Kamal[a]

[a]Department of Biological & Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, Selangor, Malaysia
[b]Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia

*Corresponding author
jinglin.ng787@gmail.com

Graphical abstract



Abstract

Good quality of rainfall data is required for the hydrological studies, water resources planning and sustainable environmental management. Consequently, the assessment of the homogeneity of rainfall data at different region is becoming increasing popular in the past few decades. In this study, the homogeneity analysis of rainfall data was carried out in Kelantan River Basin, Malaysia. The methods, namely standard normal homogeneity test (SHNT), Buishand range test, Pettitt test and von Neumann ratio test were applied to the monthly, yearly and seasonal data. The historical rainfall data from 10 rainfall stations covering the study period from 28 to 60 years were selected. The four tests were applied to 120 monthly series, 10 yearly series and 40 seasonal series. 'Useful', 'doubtful' and 'suspect' were used to classify the results of the four tests. The results showed that 94.17% of the monthly rainfall series, 70% of yearly rainfall series and 97.5% of seasonal rainfall series are labelled 'useful'. There is 5% of monthly rainfall series, 30% of yearly rainfall series and 1% of seasonal rainfall series are classified as 'doubtful'. Meanwhile, there is only 0.83% of monthly rainfall series and no yearly rainfall series and seasonal rainfall series detected in the class 'suspect'. Overall, the percentage of inhomogeneity detected in the monthly, yearly and seasonal rainfall data series of Kelantan River Basin is very small, thus most of the data is suitable to be used for further hydrological and variability analysis.

Keywords: Homogeneity, rainfall, monthly, yearly, seasonal

1.0 INTRODUCTION

The reliable quality and homogeneous rainfall records are the essential foundation of the hydrological model that can be used for water resources and environmental management [1, 2, 10]. However, several factors often caused the inhomogeneity of rainfall records, such as changes of operational routines, instrument errors and shifting of rainfall gauges. Also, the environmental conditions, such as the increasing urbanization and the occurrence of natural hazards like flood and landslides may affect the homogeneity and the continuity of rainfall data. It ultimately influence the results of hydrologic models that use rainfall as input. The consequences resulted from the inhomogeneity of rainfall records are rarely fully incorporated in the hydrological analysis but it can be a potentially large source of error. Thus, the homogeneity of rainfall records should always be tested thoroughly.

The detection of inhomogeneity of rainfall data generally divided into two techniques, namely absolute and relative tests. The absolute tests involve only the individual station series where the tests are carried out in each rainfall stations independently. In contrast, the testing procedures of the relative tests are dependent on the neighboring stations. While both techniques are useful and convenient, they both have limitations. The absolute test is more appropriate for the sparse spatial density of the station networks. However, the key problem of absolute test is that the inhomogeneity is hard to distinguish from the real climate influences [8] . Therefore, the information regarding the station history is needed to test the departures from the homogeneity. The relative tests are preferred when the correlations between the candidate and its neighboring series are high. They assume that the climatic variability within a particular region are similar but problems will arise if there are simultaneous changes in the observational network and the measuring procedures [7].

Many countries adopted different homogenization procedure depending on the climatic variables tested and the spatial and temporal resolution of the data. Wijngaard *et al*. [11] applied standard normal homogeneity test, the Buishand range test, the Pettitt test and the Von Neumann ratio test to the European daily temperature and rainfall series for the period 1901-1999. Fewer breaks were detected for the rainfall series due the higher variability. Costa and Soares [3] proposed a geostatistical approach that used a direct sequential simulation to assess the homogeneity of rainfall in southern region of Portugal. They compared the performance with standard normal homogeneity test, the Buishand range test and the Pettitt test. The resulted showed that the geostatistical approach performed very well and it is able to recognize the change points at the beginning and the end of the series. However, the exact number of simulated realization required to infer the local probability density functions accurately is not known. A recent study by Hosseinzadeh *et al.* [4] involved the Bayesian, Cumulative Deviations and von Neumann tests in evaluating the quality of rainfall data in Iran. They reported that the outputs of the homogeneity analysis had discrepancies due to the varying sensitivity level of the tests. Also, the von Neumann test is more sensitive than the other two tests.

The assessments of the homogeneity of weather data were also done in Malaysia. Kang and Yusof [5] employed the standard normal homogeneity test, Buishand Range test, Pettitt test and von Neumann ratio tests to test the homogeneity of rainfall in Peninsular Malaysia. The testing variables used are the annual mean, annual maximum and annual median. A similar study was also carried by Ahmad and Deni [1] in Peninsular Malaysia, Sabah and Sarawak by testing the homogeneity of annual rainfall amount and annual number of wet days. The previous homogeneity study conducted in Malaysia are mainly applied to the daily, monthly and annual data. So far, however, there has been little discussion about the inhomogeneity detection on the monsoon seasonal data in Malaysia.

Accordingly, the purpose of this study is to evaluate the homogeneity of monthly, yearly and seasonal rainfall data in Kelantan River Basin, Malaysia. The two way approach proposed by Wijngaard *et al.* [11] are adopted in this study. Absolute tests for homogeneity testing is chosen instead of relative tests because the location of rainfall stations in Malaysia are randomly scattered. The rainfall data of 10 stations are acquired for the period from 28 to 60 years. In total, there are 120 monthly series, 10 yearly series and 40 seasonal series.

2.0 STUDY AREA AND DATASETS

Kelantan River Basin is the study area of this study and it extends from 4° 40' N to 6° 12' N latitudes and 101° 20' E to 102° 20' E longitudes located at the north eastern part of the Peninsular Malaysia. The climate of Kelantan River Basin is equatorial-monsoon with high temperature and high humidity. The rainfall in this area is affected by monsoon and inter-monsoon seasons, which are Southwest monsoon (May to August), inter-monsoon1 (September to October), inter-monsoon2 (March to April) and Northeast monsoon (November to February). The historical rainfall data of 10 rainfall stations were acquired from Malaysian Meteorological Department (MMD). The rainfall data were checked to ensure that there is no missing values. If there is presence of missing values, the linear regression method was used to estimate the missing gaps. The geographical locations of all the stations are shown in Table 1and Figure 1.

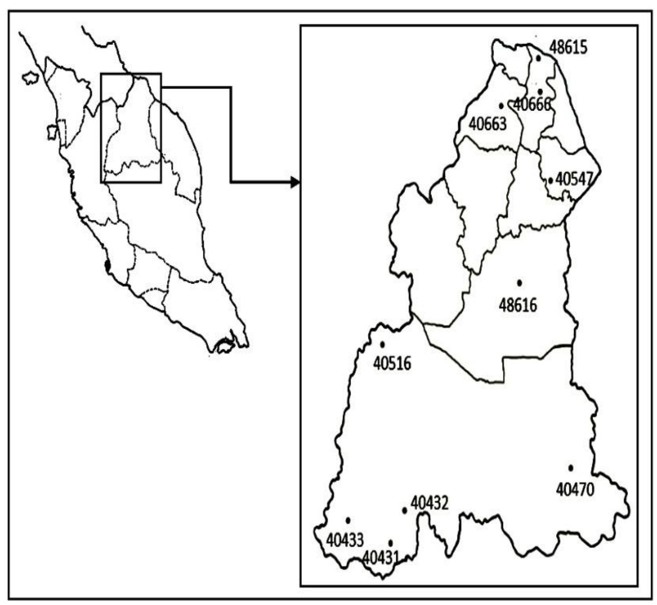


Figure 1 Geographical location of rainfall stations in Kelantan River Basin

Table 1 List of rainfall stations used in this study

Station code	Station name	Record Period	Duration	Latitude	Longitude
40431	Pos Blau	1978-2013	36	04° 39' N	101° 41' E
40432	RPS Kuala Betis	1974-2013	40	04° 42' N	101° 45' E
40433	Pos Hau	1978-2013	36	04° 42' N	101° 32' E
40470	Pos Lebir	1978-2013	36	04° 56' N	102° 23' E
40516	Pos Gob	1978-2013	36	05° 17' N	101° 38' E
40547	Mardi Jeram Pasu	1984-2013	30	05° 48' 46'' N	102° 20' 40'' E
40663	Pusat Pertanian Pasir Mas	1976-2013	38	06° 02' N	102° 07' E
40666	Mardi Kubang Keranji	1982-2013	32	06° 05' N	102° 17' E
48615	Kota Bharu	1954-2013	60	06° 10' N	102° 18' E
48616	Kuala Krai	1986-2013	28	05° 32' N	102° 12' E

3.0 METHODOLOGY

Four homogeneity tests, which are standard normal homogeneity test (SNHT), the Buishand range (BR) test, Pettitt test, and von Neumann ratio (VNR) tests are selected to test the homogeneity of the rainfall data. The testing variables used are the monthly, annual and seasonal rainfall data.

3.1 Standard Normal Homogeneity Test (SNHT)

The standard normal homogeneity test assumes that the testing variables are normally distributed and it is able to detect inhomogeneity near the beginning and end of the series. The statistic can be expressed in the equation:

$$\mathrm{T_k} = \mathrm{k}\overline{\mathrm{z}_1^2} + (\mathrm{n}-\mathrm{k})\overline{\mathrm{z}_2^2} \qquad \mathrm{k} = 1,2,3,\ldots,\mathrm{n}$$

where

$$\overline{\mathrm{z}_1} = \frac{1}{\mathrm{k}}\sum_{\mathrm{i}=1}^{\mathrm{k}}(\mathrm{P_i}-\overline{\mathrm{P}})/\mathrm{s} \quad \text{and } \overline{\mathrm{z}_2} = \frac{1}{\mathrm{n}-\mathrm{k}}\sum_{\mathrm{i}=\mathrm{k}+1}^{\mathrm{n}}\frac{(\mathrm{Y_i}-\overline{\mathrm{Y}})}{\mathrm{s}}$$

and $\mathrm{P_i}$ is the time series to be tested, $\overline{\mathrm{P}}$ is the mean and s is the standard deviation. $\mathrm{T_k}$ reaches its peak value near the year k=K if a break is located at the year K. The test statistic is expressed as:

$$\mathrm{T_0} = \max_{1\le k\le n} \mathrm{T(k)}$$

If $\mathrm{T_0}$ is greater than the critical value, the null hypothesis is rejected and there is a significant break present within the time series.

3.2 Buishand Range Test

The Buishand range test is capable to detect breaks in the middle of a time series. The adjusted partial sums are defined as

$$S_0^* = 0 \; and \; S_K^* = \sum_{i=1}^{k}(P_i - \bar{P}) \quad k = 1,2,3,\ldots,n$$

No systematic deviations of the P_i values with respect to their mean will appear and the values of S_K^* will fluctuate around zero if the series is homogeneous. The 'rescaled adjusted range' R can be used to test the significance of the change in the mean. The value of R is given as

$$R = (\max_{0\le k\le n} S_k^* - \min_{0\le k\le n} S_k^*)/s$$

3.3 Pettitt Test

The Pettitt test is a non-parametric test that evaluate the homogeneity of a series based on the ranks of the elements. The statistic is defined as

$$X_k = 2\sum_{i=1}^{k} r_i - k(n+1) \qquad k = 1,2,3,\ldots,n$$

where $r_1, r_2, r_3, \ldots, r_n$ are the ranks of the tested variables that arranged in ascending order. The value of statistic will reach its maximum or minimum near the year k=E if there is a shift of mean level in year E. The maximum value is given as

$$X_E = \max_{1\le k\le n}|X_k|$$

3.4 Von Neumann Ratio Test

The Von Neumann ratio test is a non-location specific test that evaluate whether the series is randomly distributed. It is sensitive to departures of homogeneity within a time series. It uses the ratio of mean square successive (year to year) difference to the variance. The test statistic is shown as follows:

$$N = \frac{\sum_{i=1}^{n-1}(P_i - P_{i+1})^2}{\sum_{i=1}^{n}(P_i - \bar{P})^2}$$

The value will lower than 2 if the sample has a break and rapid variations in the sample will make the value of N higher than 2. The critical values of each homogeneity tests are given in Table 2.

3.5 Validation of Results

All four tests are tested at 5% significant level. The outputs generated from all the four tests are classified into three classes, namely useful, doubtful and suspect. The classification is defined as 'useful' when one or none of the tests reject the null hypothesis; 'doubtful' when the series reject two null hypotheses out of four tests; 'suspect' when three or all hypothesis of the tests are rejected.

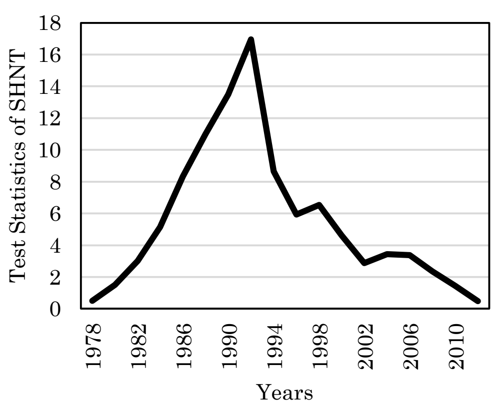


Figure 2 Test statistics of SHNT of yearly rainfall at station 40516

Table 2 The 5% critical values for the four homogeneity tests as a function of n

n	T_0 (SNHT)	$R/\sqrt{n}$ (BR)	X_E (PET)	N (VNR)
20	6.95	1.43	57	1.30
30	7.65	1.50	107	1.42
40	8.10	1.53	167	1.49
50	8.45	1.55	235	1.54
70	8.80	1.59	393	1.61

4.0 RESULTS AND DISCUSSION

All the four homogeneity tests are applied on the 120 monthly series, 10 yearly series and 40 seasonal series of all the rainfall stations. The results of the homogeneity tests are shown in Table 3. The overall results of monthly time series indicated that there were 113 out of 120 monthly series (94.17%) were "useful", 6 series (5%) were labeled as "doubtful" and the remaining (0.83%) were found to be "suspect". Table 4 presents the results of homogeneity tests for October time series of all the rainfall stations and their respective 5% critical values. It is apparent that almost all the October time series of the rainfall stations are grouped as class "useful" except the station 40663 that rejected two null hypothesis at 5% significance level and categorized under the class "doubtful". Interestingly, the rejection of the null hypothesis of Von Neumann ratio test was found in almost all of the October time series. As such, it indicated that the October time series are not randomly distributed.

For the yearly time series, the results of SNHT, Buishand range and Pettitt tests for Station 40516 are shown in Figure 2, 3 and 4, respectively. It is noticeable from Figure 2 that there is a break or extreme in year 1992 for the SHNT, thus there was a rejection of null hypothesis at the 5% level of significance. The similar results are drawn from the Buishand range tests and Pettitt tests where the test statistics of both tests pass a break in the year 1992. From this standpoint, it stands to reason that there is inhomogeneity around year 1992 due to the significant break detected by three of the homogeneity tests. As shown in Table 3, there were 7 out of 10 yearly series (70%) were categorized as "useful", 3 of the series (30%) were "doubtful" and there was no series labelled as "suspect".

Next, Table 5 revealed that the inter-monsoon2 time series of all the rainfall stations are assigned to class "useful "because there was only one rejection of null hypothesis under four tests at 5% significance level. It is remarkable to note that most of the inter-monsoon2 time series also reject the null hypothesis of Von Neumann ratio test. The overall results of homogeneity tests as show in Table 3 indicated that the homogeneity break is not apparent from the seasonal times series where 39 out of 40 times series (97.5%) were "useful" and only one time series was "doubtful".

The overall results of this study show that most of the rainfall data in the chosen rainfall stations of the Kelantan River Basin is homogeneous with only small percentage of break detected in monthly, yearly and seasonal time series. Those "useful" time series can be used for further hydrological modelling. For the "doubtful" series, careful considerations should be taken because there is an indication of inhomogeneity within the series. The "suspect" series should be strictly discarded from further analysis because the level of inhomogeneity has exceed the normal limit and therefore is not suitable for any analysis. These results are consistent with the findings of Suhaila *et al.* [9] and Ahmad and Deni [1] which showed that the time series of the east coast Peninsular Malaysia are sufficiently homogeneous for further analysis. The possible reason for this is the slower regional development of the east coast Peninsular Malaysia compared to the west coast Malaysia [1]. The states located in the east coast Peninsular

Malaysia, especially the Kelantan state recorded the lowest growth rate of Gross Domestic Product [6]. Therefore, it is possible that the changes of environmental conditions are less detected and contributing to a higher percentage of homogeneity. Nevertheless, the station networks of this study are less dense where only ten stations of Kelantan state were utilized. Besides, the reference stations and the neighboring stations are not proven to be sufficiently correlated, thus, the absolute tests appear to be the appropriate choice to detect the breaks of observational routines.

Table 3 The classification for class 'useful', 'doubtful' and 'suspect' for monthly, yearly and seasonal rainfall data

	Monthly	Yearly	Seasonal
Class 1 'useful'	113 (94.17%)	7 (70%)	39 (97.5%)
Class 2 'doubtful'	6 (5%)	3 (30%)	1 (2.5%)
Class 3 'suspect'	1(0.83%)	0 (0%)	0 (0%)

Table 4 Results of four homogeneity tests for October time series at 95% level of significance (bolded value represents that the time series is inhomogeneous)

Station code	SNHT	Buishand	Pettitt	Von Neumann
40431	2.6427	1.2328	83	1.1087
40432	1.8236	0.8669	68	**2.7032**
40433	4.0636	0.9747	106	**2.0916**
40470	2.0260	0.7550	72	**2.3983**
40516	3.1165	0.9312	117	**2.2242**
40547	1.8617	0.8394	71	**1.7144**
40663	4.1079	1.0381	**121**	**1.7442**
40666	1.3424	0.8489	59	**2.0803**
48615	5.8131	1.2461	291	**1.8574**
48616	1.4678	0.8652	77	**2.1530**

Table 5 Results of four homogeneity tests for inter-monsoon2 time series at 95% level of significance (bolded value represents that the time series is inhomogeneous)

Station code	SNHT	Buishand	Pettitt	Von Neumann
40431	3.7436	**1.6642**	106	1.2527
40432	3.2250	1.2046	122	**2.3278**
40433	5.2050	1.1174	101	**2.0749**
40470	5.2767	1.1305	118	**2.0010**
40516	3.7165	1.0620	109	**1.9630**
40547	1.0834	0.7203	45	**2.2739**
40663	3.3661	0.8504	54	**1.8096**
40666	1.0590	0.8131	45	**1.9745**
48615	4.0933	1.0598	225	**2.0685**
48616	2.7699	0.7856	48	**2.4405**

Figure 3 Test statistics of BR test of yearly rainfall at station 40516

Figure 4 Test statistics of Pettitt test of yearly rainfall at station 40516

5.0 CONCLUSION

Overall, the detection of inhomogeneity that might present within the rainfall data was carried out successfully. Four homogeneity tests, namely SNHT, Buishand Range test, Pettitt test and Von Neumann ratio test have been applied to the monthly, yearly and seasonal rainfall of 10 rainfall stations in Kelantan River Basin., Malaysia. The absolute tests were selected instead of relative tests due to the sparse density of rainfall stations in Kelantan River Basin. Among all the testing variables, seasonal rainfall data showed the highest percentage of homogeneity (97.5 %). In general, almost all the time series tested are homogeneous and considered useful for further analysis. Those series that contain inhomogeneity should be examined properly before applying them in any hydrological application.

This study involved only ten rainfall stations in Kelantan River basin, thus, the results produced cannot give a full indication of the whole river basin. Further studies should be conducted by including more rainfall stations in Peninsular Malaysia to provide more valuable information. Also, this study detected the inhomogeneity present within the rainfall data, but it do not involve the adjustments to the inhomogeneous series. The adjustment method requires the detailed information of a particular study area, such as the observational routines of instruments, station history, environmental characteristics and many others. Further work needs to be done to investigate the possible causes of the inhomogeneity and subsequently, suitable adjustment techniques can be used to improve the quality of hydrological time series.

Acknowledgement

The authors are thankful to the Malaysian Meteorological Department (MMD) for providing the daily precipitation data and the Ministry of Education Malaysia (MOE) for the financial support. The authors also would like to acknowledge the sincere appreciation to the reviewers for their valuable comments.

References

[1] Ahmad, Norlee Hainie, and Sayang Mohd Deni. 2013. Homogeneity Test on Daily Rainfall Series for Malaysia. *MATEMATIKA*. 141-150.

[2] Buishand, T. A., G. De Martino, J. N. Spreeuw, and T. Brandsma. 2013. Homogeneity of Precipitation Series in the Netherlands and Their Trends in the Past Century. *International Journal of Climatology*. 33(4): 815-833.

[3] Costa, Ana Cristina, and Amílcar Soares. 2008. Homogenization of Climate Data: Review and New Perspectives Using Geostatistics. *Mathematical Geosciences*. 41(3): 291-305.

[4] Hosseinzadeh Talaee, P., Mahdi Kouchakzadeh, and B. Shifteh Some'e. 2013. Homogeneity Analysis of Precipitation Series in Iran. *Theoretical and Applied Climatology*. 118(1-2): 297-305.

[5] Kang, Ho Ming, and Fadhilah Yusof. 2012. Homogeneity Tests on Daily Rainfall Series in Peninsular Malaysia. 7(1): 9-22.

[6] Krimi, M., Z. Yusop, and L. Hook. 2010. Regional Development Disparities in Malaysia. *Journal of American Science*. 6(3): 70-78.

[7] Peterson, Thomas C., David R. Easterling, Thomas R. Karl, Pavel Groisman, Neville Nicholls, Neil Plummer, Simon Torok, *et al*. 1998. Homogeneity Adjustments of In Situ Atmospheric Climate Data: A Review. *International Journal of Climatology*. 18(13): 1493-1517.

[8] Sahin, Sinan, and H. Kerem Cigizoglu. 2010. Homogeneity Analysis of Turkish Meteorological Data Set. *Hydrological Processes*. 24(8): 981-992.

[9] Suhaila, Jamaludin, Sayang Mohd Deni, and Abdul Aziz Jemain. 2008. Detecting Inhomogeneity of Rainfall Series in Peninsular Malaysia. *Asia-Pacific Journal of Atmospheric Sciences*. 44(4): 369-380.

[10] Vicente-Serrano, Sergio M., Santiago Beguería, Juan I. López-Moreno, Miguel A. García-Vera, and Petr Stepanek. 2009. A Complete Daily Precipitation Database for Northeast Spain: Reconstruction, Quality Control, and Homogeneity. *International Journal of Climatology.* 30(8): 1146-1163.

[11] Wijngaard, J. B., a. M G Klein Tank, and G. P. Können. 2003. Homogeneity of 20th Century European Daily Temperature and Precipitation Series. *International Journal of Climatology*. 23: 679-692.