Inhomogeneity detection in the rainfall series for the Mae Klong River Basin, Thailand

Alamgir Khalil¹

Received: 2 April 2021 / Accepted: 29 July 2021 / Published online: 10 August 2021
© The Author(s) 2021

Abstract

An accurate and complete rainfall record is prerequisite for climate studies. The purpose of this research study was to evaluate the homogeneity of the rainfall series for the Mae Klong River Basin in Thailand. Monthly rainfall data of eight stations in the Mae Klong River Basin for the period 1971–2015 were used. The double mass curve analysis was used to check the consistency of rainfall data, whereas the absolute homogeneity was assessed using the Pettitt test, standard normal homogeneity test, Buishand test, and von Neumann test at a 5% significance level. The results of these tests were qualitatively classified as ‘useful’, ‘doubtful’, and ‘suspect’ according to the null hypothesis. Results of the monthly time series indicated the rainfall data as ‘useful’ for 75% of the stations, while two stations’ data were classified as ‘doubtful’ (Stn130221) and ‘suspect’ (Stn376401). On an annual scale, seven out of eight stations data were classified as ‘useful,’ while one station (Stn376401) data were classified as ‘suspect’. Double mass curve analysis technique was used for the adjustment of inhomogeneous data. The results of this study can help provide reliable rainfall data for climate studies in the basin.

Keywords Rainfall · Inhomogeneity tests · Mae Klong River Basin

Introduction

The reliability of hydro-meteorological studies depends upon the accuracy and availability of long-term climate data. Due to the occurrence of extreme events such as droughts, floods, and extreme temperatures around the world, climate change studies have received increasing attention. A homogeneous climate series is only affected by natural factors of weather and climate. Often, non-climatic factors such as instrument replacement, change in position of the instrument, changes in observational procedures can seriously affect the accuracy of results. A homogeneous climate record is thus a prerequisite for climate change studies (Caloiero et al. 2020).

Two types of approaches have been used for checking the reliability of hydro-meteorological data series. The first approach is to compare the station data under consideration with the surrounding stations, which is known as relative homogeneity. An example of this approach is the

double mass curve analysis (DMCA) method (Alexandersson 1986). In the second approach, statistical methods have been applied to the single station data series for the detection of inhomogeneities which are termed absolute homogeneity tests. Commonly used statistical tests include the standard normal homogeneity test (SNHT) (Alexandersson 1986), the Pettitt test (Pettitt 1979), the Buishand range test (Buishand 1982), and the von Neumann ratio test (von Neumann 1941). Since the relative homogeneity test considers the spatial pattern with the neighboring stations, they are considered more reliable. However, if the stations are widely located and have significant differences in their topography, then an acceptable good correlation may not exist between them. Consequently, statistical methods are preferred (Ahmed et al. 2020).

Both the relative and absolute methods of assessing inhomogeneities in hydro-meteorological data have been used by many researchers. Arikan and Kahya (2019) conducted a homogeneity analysis of monthly precipitation records for 160 meteorological stations in Turkey for the period 1974–2014 using absolute homogeneity tests. Results indicated 44 stations to be inhomogeneous. Double mass curve analysis (DMCA) was used to correct the inhomogeneity in the precipitation records. Ahmed et al.

✉ Alamgir Khalil
alamgirkhalil@uetpeshawar.edu.pk

¹ Department of Civil Engineering, University of Engineering and Technology, Peshawar 25120, Pakistan

(2020) studied inhomogeneity detection in the precipitation series for Balochistan Province in Pakistan. The absolute homogeneity tests were applied at a 5% significance level to 14 climate stations for the period 1961–2009. Overall, greater portion of the time series data passed the SNHT but rejected the null hypothesis of homogeneous data for the von Neumann tests. Ay (2020) studied trends and homogeneity in temperature and rainfall series in the western Black Sea region, Turkey. The Runs test was used for homogeneity analysis of climate records. The monthly mean temperature for all stations showed no homogeneity. Kocsis et al. (2020) studied homogeneity and trends of precipitation time series in Keszthely, Western Hungary. The Pettitt test was employed for inhomogeneity detection in the annual, monthly, and seasonal precipitation records. Results indicated the importance of homogeneity tests which could help prevent the detection of false trends in the time series data. Suhaila et al. (2008) studied homogeneity analysis of daily rainfall series for 50 stations for the period 1975–2004 for peninsular Malaysia. Absolute homogeneity tests were applied due to sparse distribution of climate stations. Results showed that 10% of the stations had inhomogeneous data. Mair and Fares (2010) studied the homogeneity of the long-term rain gauges in Mākaha Valley, O‘ahu, Hawai‘i. Double mass analysis was used to investigate the data homogeneity. Pirnia et al. (2019) explored changes in streamflow due to climate and human activities in the Tajan River Basin in Iran using the Mann–Kendall test and double mass curve method. Gao et al. (2017) derived and used a double mass curve for hydrologic evaluation in the middle reaches of the Yellow River in China. The slope break was determined using the Pettitt test.

The Mae Klong River Basin is one of the 25 main river basins in Thailand. It has an abundance of water supplies, which are critical to Thailand’s economy. Administratively, the Mae Klong River Basin is composed of 25 districts belonging to eight provinces. Three provinces namely Ratchaburi, Samut Songkhram, and Kanchanaburi contribute around 70% of the basin area. The Electricity Generating Authority of Thailand (EGAT) has been given authority to manage all systems of three major dams as well as operate water supply from reservoirs. The Royal Irrigation Department (RID) is in charge of allocating water in the main irrigation system and meeting downstream water needs at Mae Klong Dam. The main aim of this study was to assess the presence of inhomogeneities in the rainfall series for the Mae Klong River Basin using relative and absolute homogeneity methods which can help provide reliable data for hydro-meteorological modeling of water resources in the basin.

Study area

The Mae Klong River Basin lies between 13°8′ and 16°23′ N in latitude and 98°11′–100°13′ E in longitude in the west of Thailand as shown in Fig. 1. The total area of the basin is 30,167 km². There are two main storage dams and two diversion dams. The two storage dams, Srinagarind (SNR) and Vajiralongkorn (VJK), are located on Khwae Yai and Khwae Noi Rivers, respectively. The two diversion dams, Tha Thung Na (TN) and Mae Klong (MK), are located on the Khwae Yai and Mae Klong Rivers, respectively. Two small tributaries Lam Taphoen River discharges to Khwae Yai River and Lampachi River drains to Khwae Noi River. The water inside the basin is supplied for domestic and industrial demands, hydel-power, and salinity control. Outside the basin, water is supplied to Bangkok Metropolitan Waterworks Authority (MWA) and to the neighboring Tha Chin Basin in the dry season. The most intensive water demand comes from the Greater Mae Klong Irrigation Project (GMKIP), which is located in the lower region of the basin (Khalil et al. 2018).

Materials and methods

Data

The monthly rainfall data of eight stations (Stn130013, 130042, 130053, 130211, 130221, 130571, 376401, and 470161) were obtained from the Royal Irrigation Department (RID) and the Thai Meteorological Department (TMD) for the period 1971–2015. The spatial distribution of rainfall stations is shown in Fig. 2a. The percentage of missing rainfall data varied from 0.19% (Stn130053) to 9.81% (Stn376401) as given in Table 1. As the rainfall data were not available for the consecutive months and some full years, the missing data were filled using two approaches. The rainfall data were obtained for 18 stations in the basin, but due to a large number of missing data, only eight stations were considered in this study. Firstly, the missing data for these stations were filled based on strong correlation with the nearby stations (Xia et al. 1999; Garcia et al. 2006). In the second approach, a long-term mean was used to fill the missing data for the corresponding stations. The spatial distribution of mean annual rainfall in the basin is shown in Fig. 2b. The mean annual rainfall of the stations was interpolated using the inverse distance weighting (IDW) technique.

Methodology

Double mass curve analysis (DMCA)

It is a graphical technique used to check the consistency of hydro-meteorological data. The inconsistency in rainfall

Fig. 1 The Mae Klong River Basin

data can be attributed to many reasons such as replacement of a rain gauge, shifting of the instrument to a new location, rainfall observation errors, calamities due to forest fires or landslides, etc. Inconsistency can be corrected using DMCA technique. The double mass curve is plotted by taking summation of a rainfall series for the rain gauge station whose consistency is to be checked against the summation of the average rainfall of the remaining rainfall gauge stations in the area under consideration. If

the rainfall data are consistent, then the double mass curve should follow a straight line. A break in the slope of the line would indicate that conditions have changed at one location but not at another as this technique is based on the hypothesis that when each recorded data comes from the same parent population, they are consistent. The inconsistent rainfall data for a station can be adjusted as follows (Arikan and Kahya 2019)

Fig. 2 Spatial distribution of **a** rainfall stations and **b** mean annual rainfall in millimeter for the period 1971–2015 in the Mae Klong River Basin (adapted from Khalil 2020a; Khalil 2020b)

Table 1 Rainfall stations with percentage of missing rainfall data during 1971–2015

Station code	Station name	Data available	% of missing data	Missing months
130013	A. Mueang, Kanchanaburi	1971–2015	1.11	2004 (Jul, Aug, Sep, Oct, and Nov), 2003 (Nov)
130042	A. Tha Maka, Kanchanaburi	1971–2015	2.59	2014 (May) 2013 (Mar, Sep), 2007 (July), 2003 (Apr), 2002 (Jun, Jul), 1999 (May), 1997 (May, Sep, Oct, Nov), 1995 (Jul), 1976 (Jun)
130053	A. Thong Pha Phum, Kanchanaburi	1971–2015	0.19	2003 (Nov)
130211	Ban Lum Sum, Kanchanaburi	1971–2015	4.26	2010 (all months), 2015 (Apr to Dec), 1991 (May, Jun)
130221	Huai Mae Nam Noi, Kanchanaburi	1971–2015	3.89	2010 (all months), 2015 (Apr to Dec)
130571	Ban Thung Pong, Kanchanaburi	1971–2015	4.07	2010 (all months), 2015 (Apr to Dec), 2002 (Jun), 1997 (Jun), 1995 (May)
376401	Umphang, Tak	1971–2015	9.81	1973 (Apr, Jul, Oct, Nov), 1977 (Jun, Sep to Dec), 1982 (May, Oct, Nov) 1988 (Apr, May, Aug, Sep, Oct), 1991 to 1993 (all months)
470161	Ban Bo, Ratchaburi	1971–2015	6.30	2010–2011 (all months), 2015 (Apr to Dec), 1997 (Jun)

$$R_a = \frac{b}{a} R_o \quad (1)$$

where

R_a = Adjusted rainfall.

R_o = Observed rainfall.

b : Slope of graph to which records are adjusted.

a : Slope of graph at time R_o was observed.

Rainfall data homogeneity tests

Homogeneity tests are used to assess the effect of non-climatic factors such as changes in instruments, observation practices, station relocations, and station environments on climate time series data (Toreti et al., 2011). The absolute homogeneity of the rainfall data for the eight stations was assessed by four commonly used tests: the Pettitt test, the standard normal homogeneity test (SNHT), the Buishand test, and the von Neumann test. These tests were conducted at a 5% significance level using XLSTAT software version 2020.5.1 (Addinsoft, 2021). By $X_i (i = 1, 2, \dots, T)$, we refer to a series of T variables for which we observe $x_i (i = 1, 2, \dots, T)$ at T successive times. Let $\hat{\mu}$ be the mean of the T observed values and let $\hat{\sigma}$ be the biased estimator of their standard deviation (we divide by T) (XLSTAT_help 2020).

Pettitt test Pettitt test (Pettitt 1979) is a nonparametric test and is independent of the distribution of data. It is based on the ranked-based Mann–Whitney test which is used to detect the shift in the time series. In this test, the null hypothesis is H_0 : The T variables follow one or more distributions that have the same location parameter. For a two-tailed test, the alternative hypothesis is H_a : There exists a time t from which the variables change of location parameter.

The statistic used for the Pettitt test is computed as follows: Let

$$\begin{aligned} D_{ij} &= -1 \text{ if } (x_i - x_j) < 0 \\ D_{ij} &= 0 \text{ if } (x_i - x_j) = 0 \\ D_{ij} &= 1 \text{ if } (x_i - x_j) > 0 \end{aligned} \quad (2)$$

We then define

$$U_{i,T} = \sum_{j=1}^i \sum_{j=i+1}^T D_{ij} \quad (3)$$

The Pettitt statistic for the alternative hypothesis is

$$K_T = \max_{1 \leq t < T} |U_{t,T}| \quad (4)$$

The p value can be calculated by $p = 2 \exp\left(\frac{-6K_T^2}{T^3 + T^2}\right)$ when the value of p is smaller than the significance level (α), then the null hypothesis is rejected. The time t when K_T occurs is the change point.

Standard Normal Homogeneity Test (SNHT) The SNHT (standard normal homogeneity test) was developed by Alexandersson (1986) to detect a change in a series of rainfall data. The test is applied to a series of ratios that compare the observations of a measuring station with the average of several stations. The ratios are then standardized. The series

of X_i corresponds here to the standardized ratios. The null and alternative hypotheses are determined by:

H_0 : The T variables X_i follow an $N(0, 1)$ distribution.

H_a : Between times 1 and ν , the variables follow an $N(\mu_1, 1)$ distribution, and between $\nu + 1$ and T , they follow an $N(\mu_2, 1)$ distribution.

The Alexandersson statistic is defined by:

$$T_0 = \max_{1 \leq t < T} \left[\nu \bar{z}_1^2 + (n - \nu) \bar{z}_2^2 \right] \quad (5)$$

$$\text{With } \bar{z}_1 = \frac{1}{\nu} \sum_{i=1}^{\nu} x_i \quad (6)$$

$$\bar{z}_2 = \frac{1}{n - \nu} \sum_{i=\nu+1}^T x_i \quad (7)$$

The statistic T_0 derives from a calculation comparing the likelihood of the two alternative models. The model corresponding to H_a implies that μ_1 and μ_2 are estimated while determining the ν parameter maximizing the likelihood.

Buishand test The Buishand test (1982) can be used on variables following any type of distribution. But its properties have been particularly studied for the normal case. In his study, Buishand focuses on the case of the two-tailed test. In the case of the Q statistic, the null and alternative hypotheses are given by:

H_0 : The T variables follow one or more distributions that have the same mean.

Two-tailed test: H_a : There exists a time τ from which the variables change of mean.

$$\text{We define } S_o^* = 0 \quad (8)$$

$$S_k^* = \sum_{i=1}^k (x_i - \hat{\mu}), k = 1, 2, \dots, T \quad (9)$$

$$\text{and } S_k^{**} = \frac{S_k^*}{\hat{\sigma}} \quad (10)$$

The Buishand's Q statistics are computed as follows:

$$Q = \max_{1 \leq k < T} |S_k^{**}|, \text{ for the two - tailed test case.} \quad (11)$$

In the case of the R statistic (R stands for range), the null and alternative hypotheses are given by:

H_0 : The T variables follow one or more distributions that have the same mean.

Two-sided test: H_a : The T variables are not homogeneous for what concerns their mean.

The Buishand's R statistic is computed as:

$$R = \max_{1 \leq k < T} (S_k^{**}) - \min_{1 \leq k < T} (S_k^{**}) \quad (12)$$

von Neumann test The von Neumann ratio is defined by:

$$N = \frac{1}{T\hat{\sigma}} \sum_{i=1}^{T-1} (x_i - x_{i+1})^2 \quad (13)$$

where the expectation of N is 2 when the X_i have the same mean.

This test does not allow for the detection of time at which change occurs.

For the four tests, XLSTAT evaluates the p value and an interval around the p value by using a Monte Carlo method.

Qualitative Interpretation of the homogeneity tests

The results of the homogeneity tests were interpreted following the approach suggested by Wijngaard et al. (2003). Based on the results of the four tests for inhomogeneity assessment, the data series of rainfall stations were divided into three classes (Wijngaard et al. 2003).

Class U: 'useful'—one or zero tests reject the null hypothesis at the 5% significance level.

Class D: 'doubtful'—two tests reject the null hypothesis at the 5% significance level.

Class S: 'suspect'—three or four tests reject the null hypothesis at the 5% significance level.

Results and discussion

Consistency check of Rainfall data

The double mass curve analysis (DMCA) technique was used to check the consistency of rainfall series for the Mae Klong River Basin before the application of the statistical tests used for absolute homogeneity analysis. The double mass plots for the eight stations are shown in Fig. 3. It can be seen that for the six stations (Stn130013, 130053, 130211, 130221, 130571, and 470161) the data follow a straight line which indicates the consistency of rainfall data for these stations. For station 130042, there are slight deviations about the straight line, while for the stn376401, there is a clear break in the slope of the double mass curve in the year 2000 (determined by visual observation of time series) which clearly indicates inconsistency of rainfall data.

Absolute Homogeneity tests

Inhomogeneity detection in each calendar month

The monthly time series of eight stations were grouped into each calendar month (January to December) which resulted

in 45 values for each rainfall station based on available data for the period 1971–2015. The four tests (Pettitt test, SNHT, Buishand test, and von Neumann test) were applied to each month for each station. These tests were applied at a 5% significance level.

Pettitt Test The results of p values for the Pettitt test are given in Table 2. Any p value less than the significance level of $\alpha = 0.05$ will indicate inhomogeneity in the corresponding time series. For five stations (Stn130013, 130042, 130211, 130221, and 470161), the null hypothesis was satisfied for all the months. Three stations (Stn130053, 130571, and 376401) were found to be inhomogeneous as the p values were less than 0.05. The data for Stn130053 were inhomogeneous for the month of April. Stn130571 had p values less than the threshold for 2 months (April and December). The data of Stn376401 were found to be inhomogeneous for 4 months (March, April, July, and September). The results varied between different months. For the month of April, three stations' rainfall data were found to be inhomogeneous.

Standard Normal Homogeneity Test (SNHT) Table 3 shows the p values obtained from SNHT. The null hypothesis of homogeneous data was rejected for p values less than 0.05. Five stations showed homogeneous data for all months. The data of Stn130221 were inhomogeneous in the months of January and March. Stn376401 showed inhomogeneous data for 5 months (March, July, August, September, and December). The months of February, April, May, June, October, and November satisfied the null hypothesis of homogeneous data for all stations.

Buishand test Results of p values obtained for the Buishand test are given in Table 4. The data for the three stations (Stn130013, 130053, and 470161) were found to be homogeneous for all months. The months of April and July could not pass the test for three stations each. Stn376401 indicated inhomogeneous data for 4 months (March, April, July, and September) with p values less than the threshold of 0.05.

Von Neumann test Table 5 shows the results of p values obtained for the von Neumann test. The month of October showed inhomogeneous data for two stations (Stn130053 and Stn130221) with p values less than 0.05. Stn376401 showed inhomogeneous data for the month of September. For the case of Stn130042, the null hypothesis was rejected in the month of May, while at Stn130571, the null hypothesis was rejected for the month of April. Overall, in this test, all stations indicated homogeneous data for the majority of the months.

For qualitative interpretation of the four tests, the results were evaluated qualitatively using the approach

Fig. 3 Double mass curve plots of rainfall stations for the Mae Klong River Basin

Table 2 The p values obtained from Pettitt test for each calendar month and station

Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Stn130013	0.421	0.657	0.580	0.505	0.586	0.930	0.200	0.706	0.627	0.266	0.614	0.946
Stn130042	0.564	0.814	0.157	0.200	0.649	0.571	0.129	0.612	0.191	0.621	0.624	0.953
Stn130053	0.653	0.474	0.215	0.017	0.825	0.230	0.384	0.509	0.240	0.427	0.268	0.843
Stn130211	0.908	0.792	0.486	0.106	0.435	0.697	0.631	0.676	0.228	0.823	0.475	0.587
Stn130221	0.451	0.764	0.776	0.127	0.601	0.268	0.109	0.809	0.846	0.131	0.563	0.050
Stn130571	0.221	0.342	0.179	0.009	0.614	0.608	0.170	0.206	0.112	0.731	0.439	0.035
Stn376401	0.694	0.061	0.000	0.009	0.750	0.155	0.035	0.272	0.015	0.852	0.927	0.307
Stn470161	0.544	0.930	0.487	0.610	0.112	0.341	0.378	0.413	0.926	0.820	0.098	0.740

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

Table 3 The p values obtained from SNHT for each calendar month and station

Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Stn130013	0.435	0.278	0.365	0.760	0.563	0.721	0.991	0.630	0.461	0.892	0.426	0.804
Stn130042	0.372	0.426	0.088	0.954	0.205	0.299	0.081	0.539	0.035	0.763	0.526	0.901
Stn130053	0.410	0.941	0.293	0.996	0.073	0.125	0.353	0.578	0.793	0.791	0.876	0.804
Stn130211	0.079	0.803	0.079	0.153	0.489	0.069	0.947	0.288	0.956	0.615	0.557	0.266
Stn130221	0.035	0.203	0.040	0.211	0.403	0.146	0.208	0.708	0.416	0.085	0.255	0.988
Stn130571	0.171	0.259	0.584	0.081	0.558	0.168	0.360	0.289	0.842	0.598	0.765	0.203
Stn376401	0.227	0.418	0.001	0.069	0.320	0.840	0.025	0.049	0.016	0.924	0.271	0.007
Stn470161	0.453	0.481	0.578	0.671	0.587	0.076	0.324	0.463	0.360	0.423	0.196	0.269

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

Table 4 The p values obtained from Buishand test for each calendar month and station

Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Stn130013	0.615	0.083	0.273	0.502	0.563	0.506	0.920	0.693	0.781	0.923	0.223	0.541
Stn130042	0.155	0.215	0.443	0.926	0.311	0.232	0.044	0.391	0.088	0.542	0.295	0.779
Stn130053	0.622	0.867	0.131	0.967	0.409	0.059	0.129	0.880	0.770	0.562	0.656	0.645
Stn130211	0.268	0.756	0.206	0.025	0.479	0.332	0.813	0.366	0.793	0.340	0.320	0.154
Stn130221	0.357	0.259	0.088	0.083	0.491	0.064	0.045	0.603	0.192	0.044	0.610	0.957
Stn130571	0.411	0.079	0.333	0.003	0.269	0.164	0.196	0.179	0.973	0.465	0.859	0.304
Stn376401	0.368	0.765	< 0.0001	0.001	0.164	0.736	0.010	0.191	0.003	0.882	0.646	0.417
Stn470161	0.835	0.954	0.899	0.421	0.929	0.050	0.272	0.324	0.678	0.233	0.075	0.929

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

Table 5 The p values obtained from von Neumann test for each calendar month and station

Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Stn130013	0.757	0.756	0.291	0.179	0.897	0.861	0.723	0.494	0.253	0.189	0.779	0.427
Stn130042	0.265	0.168	0.101	0.180	0.029	0.463	0.392	0.348	0.000	0.051	0.466	0.274
Stn130053	0.723	0.471	0.968	0.891	0.060	0.098	0.143	0.406	0.702	0.006	0.206	0.756
Stn130211	0.134	0.718	0.315	0.304	0.230	0.164	0.976	0.791	0.244	0.187	0.524	0.478
Stn130221	0.122	0.373	0.263	0.419	0.945	0.114	0.524	0.771	0.235	0.010	0.747	0.882
Stn130571	0.518	0.348	0.911	0.019	0.987	0.216	0.636	0.170	0.984	0.668	0.483	0.777
Stn376401	0.861	0.552	0.316	0.493	0.858	0.475	0.073	0.312	0.012	0.209	0.746	0.077
Stn470161	0.510	0.867	0.747	0.591	0.384	0.130	0.166	0.057	0.371	0.582	0.482	0.761

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

Table 6 Qualitative classification of each calendar month and station based on the four homogeneity tests

Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Stn130013	U	U	U	U	U	U	U	U	U	U	U	U
Stn130042	U	U	U	U	U	U	U	U	U	U	U	U
Stn130053	U	U	U	U	U	U	U	U	U	U	U	U
Stn130211	U	U	U	U	U	U	U	U	U	U	U	U
Stn130221	U	U	U	U	U	U	U	U	U	D	U	U
Stn130571	U	U	U	S	U	U	U	U	U	U	U	U
Stn376401	U	U	S	D	U	U	S	U	S	U	U	U
Stn470161	U	U	U	U	U	U	U	U	U	U	U	U

Table 7 The p values obtained from the homogeneity tests for the annual rainfall

Stations	Pettitt test	SNHT	Buishand test	von Neumann test	Class
Stn130013	0.878	0.145	0.310	0.533	U
Stn130042	0.248	0.100	0.074	0.007	U
Stn130053	0.353	0.695	0.962	0.508	U
Stn130211	0.944	0.356	0.559	0.119	U
Stn130221	0.754	0.777	0.450	0.474	U
Stn130571	0.782	0.641	0.492	0.797	U
Stn376401	0.000	< 0.0001	< 0.0001	0.004	S
Stn470161	0.829	0.316	0.525	0.634	U

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

of (Wijngaard et al. 2003) and are shown in Table 6. It can be seen that the rainfall data for five stations can be classified into class U (useful) for all the months. The data at Stn130571 can be classified into class S (suspect) in the month of April. Stn130221 data were classified into class D (doubtful) in October. The data at Stn376401 showed D in April while indicated S for 3 months (March, July, and September).

Inhomogeneity detection in annual rainfall series

For detection of inhomogeneity in the rainfall data on an annual scale, the monthly data were aggregated into total annual rainfall for all stations during the period 1971–2015. The results of the four tests are given in Table 7. It is observed that the results of seven stations can be classified in class U (useful) as these stations pass three or more tests by satisfying the null hypothesis. Stn376401 could not pass any of the four tests and was classified into class S (suspect). This result is consistent

Table 8 The p values obtained from the homogeneity tests for the monthly rainfall series during 1971–2015

Stations	Pettitt test	SNHT	Buishand test	von Neumann test	Class
Stn130013	0.374	0.849	0.855	< 0.0001	U
Stn130042	0.579	0.159	0.065	< 0.0001	U
Stn130053	0.104	0.900	0.996	< 0.0001	U
Stn130211	0.679	0.883	0.878	< 0.0001	U
Stn130221	0.038	0.975	0.845	< 0.0001	D
Stn130571	0.727	0.879	0.900	< 0.0001	U
Stn376401	0.065	0.007	0.002	< 0.0001	S
Stn470161	0.443	0.869	0.903	< 0.0001	U

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

with the double mass curve analysis (Fig. 3) which indicated a clear break in slope in the year 2000.

Inhomogeneity detection in monthly rainfall series during 1971–2015

The absolute homogeneity tests were applied to the monthly rainfall series of all stations for the period 1971–2015. The results obtained from these tests are given in Table 8. Six stations data were classified into class U (useful) as they satisfied the null hypothesis of homogeneous data for three out of four tests. The monthly data for Stn130221 were classified into D (doubtful) as it could not pass two of the four tests with p values less than the threshold of 0.05. Stn376401 was classified into class S (suspect) as it failed to satisfy the null hypothesis of homogeneous data in three tests. Results also showed that all the stations failed the null hypothesis for the von Neumann test.

The DMCA technique was applied to adjust the rainfall data for the Stn376401. As shown in the DMCA plot in Fig. 4, there is a clear break in the slope of the line during the year 2000. This is also confirmed by SNHT and Buishand range tests (March 2000). However, the Pettitt test indicated the point of change in March 1999. The slope of

Fig. 4 Double mass curve plots and SNHT results for the Stn376401 before and after homogenization of the rainfall series during 1971–2015

Table 9 The p values obtained from the homogeneity tests for the monthly rainfall series during 1971–2015 for the Stn376401 after rainfall adjustment

Stations	Pettitt test	SNHT	Buishand test	von Neumann test	Class
Stn376401	0.09	0.677	0.999	<0.0001	U

Bold number represents $p < 0.05$ indicating inhomogeneity in the corresponding time series at a 5% significance level

the line before the break was calculated to be 0.139 and the slope of the line after the break was found to be 0.179. The correction was calculated as $0.179/0.139 = 1.29$. This factor was multiplied with the rainfall data before the break (the year 2000) to reflect the new conditions (after the year 2000) (Arikan and Kahya 2019). The author did not have access to the metadata for the Stn376401 which could help determine the reasons for inhomogeneity in the rainfall series. After adjustment of the rainfall data, the four homogeneity tests were applied and the results are given in Table 9. It can be seen that the rainfall series has passed the three tests at a 5% significance level with p values above the threshold of 0.05. The graphical representation of SNHT before and after the rainfall data homogenization

along with the double mass curve is shown in Fig. 4. Since the data have passed three tests, it can be classified as ‘useful’. Overall, the data for the eight stations can be considered useful for hydro-meteorological studies in the Mae Klong River Basin.

Conclusions

This study examined the homogeneity of the rainfall series for each calendar month, monthly time series, and annual rainfall data for eight stations in the Mae Klong River Basin for the period 1971–2015. Both the relative and absolute homogeneity tests were used. The relative homogeneity of the rainfall data was assessed through double mass curve analysis (DMCA), whereas the absolute homogeneity was investigated through the four commonly used statistical methods (the Pettitt test, SNHT, Buishand test, and von Neumann test) at a 5% significance level. Results showed that for each calendar month, rainfall data of five stations (Stn130013, 130042, 130053, 130211, and 470161) were found to be ‘useful’ in all months. The rainfall data for Stn130221 were found to be ‘doubtful’ in the month of October. The data of Stn130571 were found to be ‘suspect’

in April. Stn376401 were classified into ‘doubtful’ for the month of April, while ‘suspect’ for the months of March, July, and September. For annual rainfall, all the stations data were classified as ‘useful’ except for Stn376401 which were classified as ‘suspect.’ For the monthly time series, the data for Stn130221 were classified as ‘doubtful’ and for Stn376401 as ‘suspect’, while for the remaining stations, the rainfall data were classified as ‘useful’. The double mass curve analysis method was used to adjust the monthly rainfall data of Stn376401. After adjustment, the data of Stn376401 passed the null hypothesis of homogeneous data for three out of four homogeneity tests and were classified as ‘useful.’ Overall, the rainfall time series data can be considered useful for hydro-meteorological studies of water resources in the basin.

Acknowledgements The author is grateful to the Royal Irrigation Department (RID) and Thai Meteorological Department (TMD) for providing data for this study. Comments by anonymous reviewers and the journal editor are gratefully acknowledged.

Author contributions The author contributed to the study conception and design, material preparation, data collection, and analysis and wrote the manuscript.

Funding No funding was received for this study.

Declarations

Conflict of interest The author declares no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Addinsoft (2021) XLSTAT statistical and data analysis solution. New York, USA. <https://www.xlstat.com>
- Ahmed K, Nawaz N, Khan N, Rasheed B, Baloch A (2020) Inhomogeneity detection in the precipitation series: case of arid province of Pakistan. *Environ Develop Sustain* 23(5):7176–7192
- Alexandersson H (1986) A homogeneity test applied to precipitation data. *J Climatol* 6(6):661–675

- Arikan BB, Kahya E (2019) Homogeneity revisited: analysis of updated precipitation series in Turkey. *Theoret Appl Climatol* 135(1–2):211–220
- Ay M (2020) Trend and homogeneity analysis in temperature and rainfall series in western Black Sea region. *Turk Theor Appl Climatol* 139(3):837–848
- Buishand TA (1982) Some methods for testing the homogeneity of rainfall records. *J Hydrol* 58(1–2):11–27
- Caloiero T, Filice E, Coscarelli R, Pellicone G (2020) A homogeneous dataset for rainfall trend analysis in the Calabria Region (Southern Italy). *Water* 12(9):2541
- Gao P, Li P, Zhao B, Xu R, Zhao G, Sun W, Mu X (2017) Use of double mass curves in hydrologic benefit evaluations. *Hydrol Process* 31(26):4639–4646
- Garcia BIL, Sentelhas PC, Tapia L, Sparovek G (2006) Filling in missing rainfall data in the Andes region of Venezuela, based on a cluster. *Rev Bras Agrometeorologia* 14(2):225–233
- Khalil A (2020b) Drought characterization in the Mae Klong River Basin, Thailand, using standardized precipitation index. *Arab J Geosci* 13(14):1–15
- Khalil A, Rittima A, Phankamolsil Y (2018) The projected changes in water status of the Mae Klong Basin, Thailand, using WEAP model. *Paddy Water Environ*, 16(3):439–455
- Khalil A (2020a) Rainfall trend analysis in the Mae Klong River Basin, Thailand. *Songklanakarin J Sci Technol* 42(4):879–888
- Kocsis T, Kovács-Székely I, Anda A (2020) Homogeneity tests and non-parametric analyses of tendencies in precipitation time series in Keszthely. *West Hung Theor Appl Climatol* 139(3–4):849–859
- Mair A, Fares A (2010) Assessing rainfall data homogeneity and estimating missing records in Mākaha valley, O ‘ahu, Hawai ‘i. *J Hydrol Eng* 15(1):61–66
- Pettitt AN (1979) A non-parametric approach to the change-point problem. *J Roy Stat Soc: Ser C (appl Stat)* 28(2):126–135
- Pirnia A, Golshan M, Darabi H, Adamowski J, Rozbeh S (2019) Using the Mann-Kendall test and double mass curve method to explore stream flow changes in response to climate and human activities. *J Water Clim Change* 10(4):725–742
- Suhaila J, Deni SM, Jemain AA (2008) Detecting inhomogeneity of rainfall series in Peninsular Malaysia. *Asia-Pac J Atmos Sci* 44(4):369–380
- Toreti A, Kuglitsch FG, Xoplaki E, Della-Marta PM, Aguilar E, Prohom M, Luterbacher J (2011) A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. *Int J Climatol* 31(4):630–632
- von Neumann J (1941) Distribution of the ratio of the mean square successive difference to the variance. *Ann Math Stat* 12:367–395
- Wijngaard JB, Klein Tank AMG, Können GP (2003) Homogeneity of 20th century European daily temperature and precipitation series. *Int J Clim: A J Royal Meteorol Soc* 23(6):679–692
- Xia Y, Fabian P, Stohl A, Winterhalter M (1999) Forest climatology: estimation of missing values for Bavaria. *Ger Agric For Meteorol* 96(1–3):131–144
- XLSTAT_help (2020) The XLSTAT help documentation. <https://www.xlstat.com>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.