IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

The efficiency of average linkage hierarchical clustering algorithm associated multi-scale bootstrap resampling in identifying homogeneous precipitation catchments

To cite this article: Zun Liang Chuan *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **342** 012070

View the article online for updates and enhancements.

IOP Conf. Series: Materials Science and Engineering **342** (2018) 012070 doi:10.1088/1757-899X/342/1/012070

The efficiency of average linkage hierarchical clustering algorithm associated multi-scale bootstrap resampling in identifying homogeneous precipitation catchments

Zun Liang Chuan[1], Noriszura Ismail[2], Wendy Ling Shinyie[3], Tan Lit Ken[4], Soo-Fen Fam[5], Azlyna Senawi[1] and Wan Nur Syahidah Wan Yusoff[1]

[1]Faculty Industrial Sciences & Technology, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang
[2]School of Mathematical Sciences, Faculty Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor
[3] Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
[4]Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak), 54100 Kuala Lumpur
[5]Centre of Technopreneurship Development, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka

Email: chuanzunliang@ump.edu.my

Abstract. Due to the limited of historical precipitation records, agglomerative hierarchical clustering algorithms widely used to extrapolate information from gauged to ungauged precipitation catchments in yielding a more reliable projection of extreme hydro-meteorological events such as extreme precipitation events. However, identifying the optimum number of homogeneous precipitation catchments accurately based on the dendrogram resulted using agglomerative hierarchical algorithms are very subjective. The main objective of this study is to propose an efficient regionalized algorithm to identify the homogeneous precipitation catchments for non-stationary precipitation time series. The homogeneous precipitation catchments are identified using average linkage hierarchical clustering algorithm associated multi-scale bootstrap resampling, while uncentered correlation coefficient as the similarity measure. The regionalized homogeneous precipitation is consolidated using K-sample Anderson Darling non-parametric test. The analysis result shows the proposed regionalized algorithm performed more better compared to the proposed agglomerative hierarchical clustering algorithm in previous studies.

1. Introduction

The projection accuracy of the magnitudes and frequency of extreme hydro-meteorological such as extreme precipitation events is highly dependent on the availability and quality of the historical precipitation records. However, the accuracy of the projection for extreme hydro-meteorological at East-

Coast region of Peninsular Malaysia based on at-site frequency analysis is always uncertain due to the limitation of historical precipitation records. In order to reduce the uncertainty of projection, extrapolation information from gauged to ungauged precipitation catchments based on regionalization is indeed much needed. For the past decades, several regionalization algorithms have been widely applied in hydro-meteorological studies, including the agglomerative hierarchical clustering algorithm [1-7], the canonical correlation algorithm [8], the neural network algorithm [9,10] and the principal component algorithm [11,12].

For the regional frequency analysis of hydro-meteorological studies carried out in Malaysia, the agglomerative hierarchical clustering are frequently applied as the regionalization algorithm. As examples, Ahmad et al. [5] studied the efficiency of 77 regionalized algorithms in identifying the homogenous precipitation catchments based on precipitation amount, which the 77 regionalized algorithms formed from all possible combinations of 7 agglomerative hierarchical clustering algorithms and 11 similarity measures. They concluded that the combination of complete linkage hierarchical clustering algorithm and correlation similarity measure is the most appropriate algorithm to identify the homogeneous precipitation catchments in Peninsular Malaysia. They also proposed to determine the optimum number of homogeneous regions when most of the internal clustering validation indices indicated the similar results.

Zakaria et al. [7] carried out another regional frequency analysis focused on East-Coast region, Malaysia. The Ward's minimum variance hierarchical clustering applied in their studies in identifying the homogenous precipitation catchments, which the topographic characteristics of precipitation catchments such as elevation, latitude and longitude are used as attributes of regionalization. Meanwhile, the regionalized homogenous precipitation catchments consolidated using discordant and heterogeneity tests.

Most of the regional frequency analyses carried out in previous studies are mainly focused on the stationary precipitation time series. However, past studies showed that the regional phenomenon, such as the monsoon, El Nino-Southern Oscillation, Indian Ocean Dipole and Madden-Julian Oscillation has created non-stationary components in climate variability [13,14]. In addition, the discordant and heterogeneity tests more appropriate for low skewed data [15]. Therefore, an efficient regionalization algorithm which more suitable for non-stationary precipitation time series is absolutely necessary to result more trustworthy information for projection of extreme hydro-meteorological events.

The main objective of this study is to propose an efficient regionalization algorithm to identify the homogeneous precipitation catchments for non-stationary precipitation time series, which unify the average linkage hierarchical clustering algorithm and multi-scale bootstrap resampling. On the other hands, K-sample Anderson Darling non-parametric test is also proposed to consolidate the results of the proposed regionalization algorithm.

2. Study Area

Agriculture sector plays a substantial role in ensuring food security, economic growth, socioeconomic improvement, employment generation and poverty reduction of the nation in Malaysia [16,17]. Therefore, the 20 precipitation catchments (figure 1) located at Kuantan River Basin, Pahang are selected as the study area. This is due to the Kuantan River Basin is one of the significant tributaries that irrigates the majority of the rural, urban, agriculture and industrial areas of Kuantan District [18], and frequently exposed to risks of flood occurrence during the Northeast Monsoon. The topography characteristics and descriptive statistics of monthly precipitation amount for 20 precipitation catchments are illustrated in table 1. The monthly precipitation amount applied this study covers the period February 2010 until November 2014.

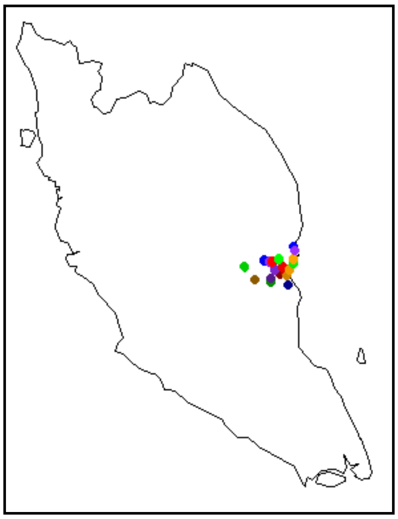


Figure 1. Location of 20 precipitation catchments in the Kuantan River Basin, Pahang.

Table 1. The topography and descriptive statistics of the monthly precipitation amount for 20 precipitation catchments in Kuantan River Basin, Pahang.

Catchment	Catchment Name	Elevation(m)	Latitude(N)	Longitude(E)	MN(mm)	CV(%)	SKEW	KURT
01	Sri Damai	14.9	03°44’47’’	103°13’20’’	90.5552	165.6909	3.3014	14.5292
02	Paya Bungor	34.7	03°41’30’’	102°56’00’’	158.1138	77.1298	1.0290	1.3189
03	Kampung Pulau Manis	37.4	03°39’10’’	103°07’10’’	181.5224	71.5749	2.0425	7.3836
04	Kampung Bahru	7.6	03°37’50’’	103°18’55’’	179.6224	92.8763	2.3708	6.9554
05	JKR Gambang	41.3	03°42’20’’	103°07’00’’	234.0069	66.1214	1.5544	3.3734
06	Paya Besar	6.0	03°46’20’’	103°16’50’’	162.8810	93.1208	2.2039	8.2500
07	Kampung Sungai Soi	11.9	03°43’50’’	103°18’00’’	210.9621	94.1730	2.6879	8.4473
08	Ladang Ulu Lepar	91.7	03°50’25’’	102°48’00’’	167.2017	65.7289	0.9717	1.0498
09	Ladang Mentiga	9.4	03°48’58’’	103°19’30’’	199.1931	73.1688	1.1191	2.0870
10	Panching	71.4	03°48’53’’	103°09’38’’	234.0707	86.9853	2.0654	6.9068
11	Paya Pinang	6.7	03°50’30’’	103°15’30’’	209.7328	93.7746	3.0494	13.5455
12	JPS Pahang	10.3	03°48’30’’	103°19’45’’	180.8603	97.8549	2.7916	12.2130
13	Ladang Jeram	-1.4	03°53’40’’	103°23’00’’	210.9414	119.548	3.3951	15.1796
14	Sungai Lembing	33.1	03°55’00’’	103°02’10’’	245.6690	62.2638	0.3866	-0.4114
15	Ladang Nada	16.9	03°54’30’’	103°06’20’’	227.9431	73.9733	1.6809	4.9231
16	Ladang Kuala Reman	29.9	03°54’00’’	103°08’00’’	201.8638	74.4579	0.7849	1.0727
17	Balok	4.1	03°56’40’’	103°23’00’’	220.8241	110.4599	2.6069	8.4019
18	Bukit Sagu	20.9	03°56’14’’	103°12’52’’	511.7517	79.6973	2.7206	12.0137
19	Kampung Cherating	9.0	04°05’35’’	103°22’50’’	221.2155	108.1321	2.5446	8.0246
20	Kampung Sungai Ular	58.5	04°30’00’’	103°23’40’’	228.7362	108.5308	3.0108	11.2957

***Note:** **MN**: Monthly average precipitation amount; **CV**: Coefficient of variation; **SKEW**: Skewness; **KURT**: Kurtosis

3. Methodology

Figure 2 illustrated the overview of the proposed regionalization in identifying the homogenous precipitation catchments for non-stationary precipitation time series.

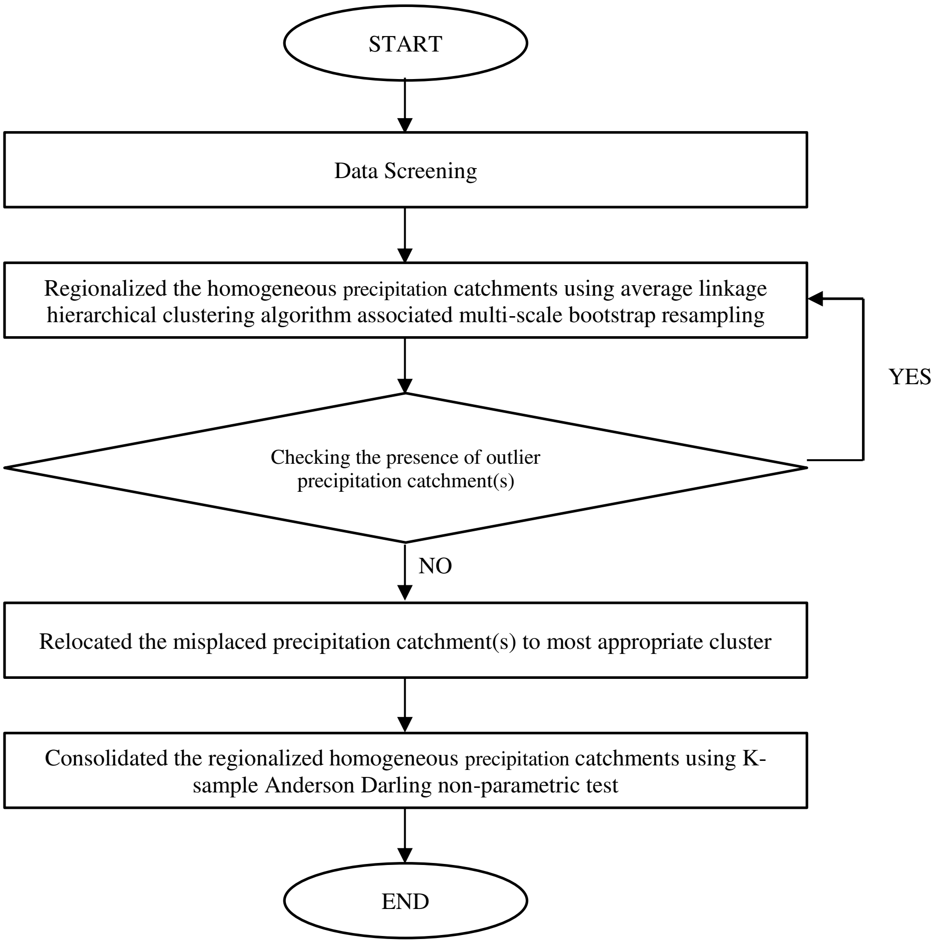


Figure 2. The overview of the proposed regionalized algorithm for non-stationary time series.

3.1 Data screening

Suppose that $\mathbf{Y} = \left[y_{ij} \right]_{MN}$ represents the *ith* month precipitation amount for *jth* precipitation catchments, such that $i,(j) = 1,\ 2,\ \ldots,\ M,(N)$, and $y_{ij} \in \mathbf{Y}$, the available monthly precipitation amount, $y_{ij}^{miss} \in \mathbf{Y}$, the missing precipitation amount. In this study, the y_{ij}^{miss} can be imputed based on the algorithm as below:

$$y_{ij}^{miss} = \left(\prod_{i=1}^{m_i} y_{ij} \right)^{\frac{1}{m_i}} \tag{1}$$

which is the superior imputation missing precipitation amount proposed by Saeed et al. [19]. Therefore, a complete $\hat{\mathbf{Y}} = \left[\hat{y}_{ij} \right]_{MN}$ is resulted. On the other hands, it should be highlighted that $\hat{\mathbf{Y}}$ is independent of the inherent assumptions of stationary time series after testing using Mann-Kendall test [20]. This is due to the Mann-Kendall test show that there is significant serial correlation over time presented in $\hat{\mathbf{Y}}$ although pre-whitening filter is applied.

3.2 Regionalized homogeneous precipitation catchments

The average linkage agglomerative hierarchical clustering algorithm is selected in this study due to this algorithm yielded more reasonable results compared to the single linkage and Ward's minimum variance agglomerative hierarchical algorithm, which the topography and descriptive statistics of monthly precipitation amount (table 1) as attributes of regionalization, and uncentered correlation coefficient as the similarity measures. In principle, the average linkage is performed $N-1$ successive fusion by agglomerating the closest pair of precipitation catchments based on the predetermined similarity measure, until N precipitation catchments fused as a single dendrogram.

Suppose that $\theta\left(\hat{\mathbf{Y}}_{a_0}, \hat{\mathbf{Y}}_{b_0}\right)$ represents the minimum value of predetermined similarity measure for a single cluster, $\hat{\mathbf{Y}}_a$, constitutes the fused pair of $\hat{\mathbf{Y}}_{a_0}$ and $\hat{\mathbf{Y}}_{b_0}$ clusters. A new dendrogrammatic similarity measure, $\tilde{\theta}\left(\hat{\mathbf{Y}}_a, \hat{\mathbf{Y}}_b\right)$, between a new single cluster, $\hat{\mathbf{Y}}_a$, and remaining infused clusters, $\hat{\mathbf{Y}}_b$ is yielded by updating the dendrogrammatic similarity function [21] given as

$$\tilde{\theta}\left(\hat{\mathbf{Y}}_a, \hat{\mathbf{Y}}_b\right) = \frac{n_{a_0}}{n_{a_0} + n_{b_0}} \tilde{\theta}\left(\hat{\mathbf{Y}}_{a_0}, \hat{\mathbf{Y}}_a\right) + \frac{n_{b_0}}{n_{a_0} + n_{b_0}} \tilde{\theta}\left(\hat{\mathbf{Y}}_{b_0}, \hat{\mathbf{Y}}_b\right) \tag{2}$$

where n_{a_0} and n_{b_0} represents the number of precipitation catchments in the cluster $\hat{\mathbf{Y}}_{a_0}$ and $\hat{\mathbf{Y}}_{b_0}$, and the uncentered correlation coefficient as the similarity measure, $\bar{R}$, are given as

$$\bar{R} = 1 - \frac{\sum_{i=1}^{M}\left(\hat{y}_{ij}\hat{y}_{ik}\right)}{\sqrt{\sum_{i=1}^{M}\left(\hat{y}_{ij}\right)^2 \sum_{i=1}^{M}\left(\hat{y}_{ik}\right)^2}}; \; \hat{y}_{ij} \in \hat{\mathbf{Y}}_a, \hat{y}_{ik} \in \hat{\mathbf{Y}}_b \tag{3}$$

Since identifying the homogeneous precipitation catchments based on a single dendrogram is very uncertain, therefore this study suggested the use of multi-scale bootstrap to reduce the uncertainty in identifying the homogeneous precipitation catchments and outlier precipitation catchment(s). The multi-scale bootstrap resampling is an approximately unbiased test based on the bootstrap probabilities.

3.3 Consolidated the regionalized precipitation catchments

The K-sample Anderson Darling non-parametric test, λ_{AD}^2, is a rank test, which independent on any assumption of statistical assumption [15]. In this study, the K-sample Anderson Darling non-parametric test is selected due to the distribution for most of precipitation amount (table 1 and figure 3) is highly positively skewed, therefore the discordant and heterogeneity tests might be yielded a misleading result.

Suppose that $\hat{\mathbf{Y}}$ is regionalized into Q regions of homogeneous precipitation catchments, which each region can be represented by $\mathbf{C}_q = \left[\hat{y}_{il}\right]_{MQ}$; $i,(l) = 1,2,3,\ldots,\mathrm{M},(Q)$ and $\mathbf{V} = \left[v_r\right]_{(MQ)} = vec\left(\mathbf{C}_q\right)$; $r = 1,\ 2,\ \ldots,\ MQ$ is a vector of the pooled precipitation amount, such that $v_1 < v_2 < \ldots < v_{MQ}$, which $vec(\cdot)$ is vectorization. The $\mathbf{C}_q$ is considerable statistical heterogeneity if

$$\left(\phi_{\mathbf{C}_q} = \frac{\lambda_{AD}^2 - MQ + 1}{\sqrt{\sigma^2_{\lambda^2_{AD}}}}\right) \geq \phi_{(MQ-1,\alpha)} \tag{4}$$

where λ_{AD}^2 and its variance, $\sigma^2_{\lambda^2_{AD}}$, is given in equations (5) and (6), respectively.

$$\lambda_{AD}^2 = \frac{1}{(MQ)^2}\sum_{r=1}^{MQ-1}\left(\frac{\left(\tilde{n}(MQ) - rM\right)^2}{r(MQ-r)}\right) \tag{5}$$

$$\sigma^2_{\lambda^2_{AD}} = \frac{\Gamma(MQ-3)}{\Gamma(MQ)}\left(\omega_1 + \omega_2(MQ) + \omega_3(MQ)^2 + \omega_4(MQ)^3\right) \tag{6}$$

The $\tilde{n}$ represents the number of precipitation amount in the lth precipitation catchments that are not greater than v_r, and ω_1, ω_2, ω_3 and ω_4 are defined in equations (7), (8), (9) and (10), respectively.

$$\omega_1 = 2Q(\gamma Q - 2\gamma + 3Q) \tag{7}$$

$$\omega_2 = 2\gamma\left(\frac{Q^2(4\delta - 1) + 3Q + 2\delta}{\delta}\right) + 2\beta(\gamma - 3) \tag{8}$$

$$\omega_3 = 2\delta\left(Q^2 + \beta + 2\right) + 2\left(4\gamma(Q-1) - \beta(7\gamma - 2) + 3 - 2Q^2\right) \tag{9}$$

$$\omega_4 = 2\delta(2Q - 3\beta - 2) - 2(3Q + 2\gamma - 3) \tag{10}$$

where $\beta = \frac{1}{QM}$, $\gamma = \sum_{r=1}^{QM-1}\frac{1}{r}$, and $\delta = \sum_{i=1}^{QM-2}\sum_{l=1+1}^{QM-1}\left(\frac{1}{l(QM-i)}\right)$.

4. Analysis Results

The five-number summary of monthly precipitation for 20 precipitation catchments located Kuantan River Basin, Pahang are illustrated as figure 3. On average, it can be observed (figure 3) that Catchments 1 and 18 respectively received the lowest and highest monthly precipitation compared to other 18 precipitation catchments. The description also can be consolidated by the result of nonparametric pairwise comparison, which indicates that the average monthly precipitation amount of Catchment 18 is significantly higher, while the average monthly precipitation amount of Catchment 01 is significantly lower compared to other precipitation catchments. On the other hands, the non-parametric multiple comparison tests also indicates that there are significant differences between Catchment 05 with Catchments 02 and 08 and Catchment 14 with Catchments 02, 03, 06, 08, and 18 at significant level of 0.05.

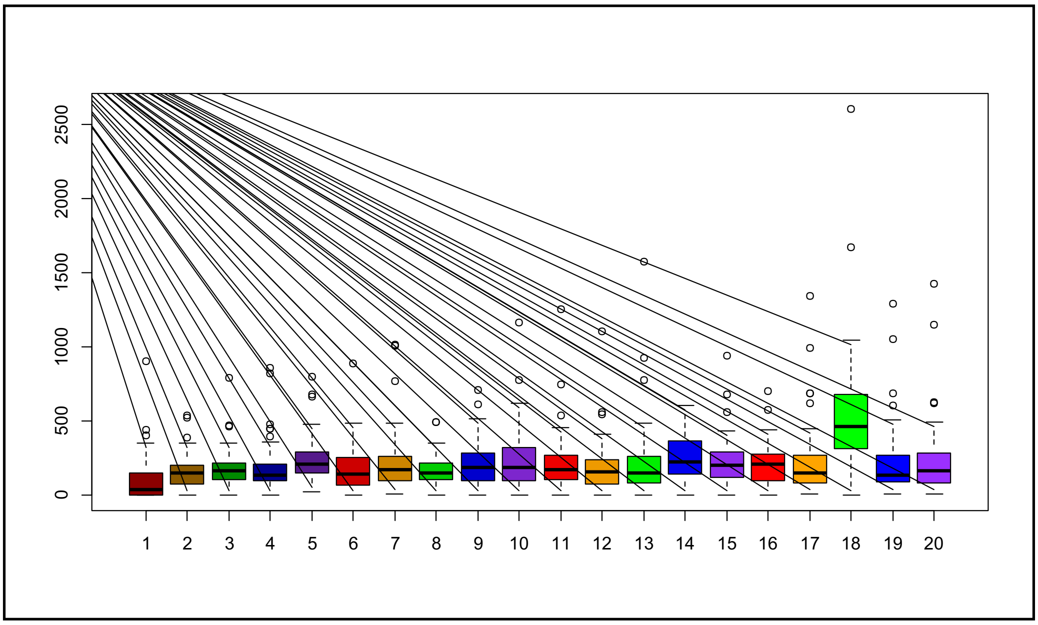
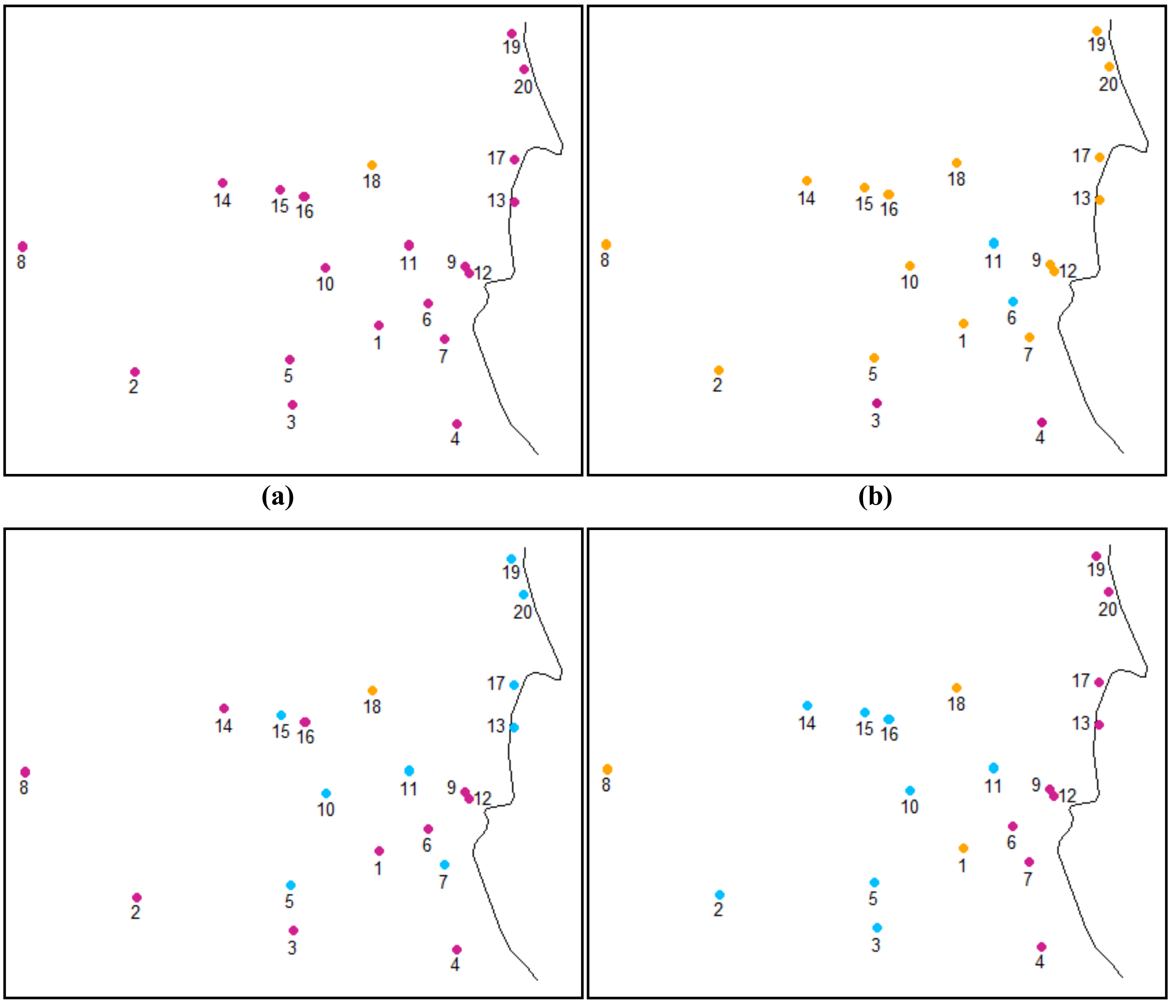


Figure 3. The five-number summary of precipitation amount for 20 precipitation catchments located Kuantan River Basin, Pahang.

In order to authenticate the efficiency of the proposed algorithm, Ward^{LR} and Complete^{LR}, algorithms respectively proposed by Ahmad et al. [5] Zakaria et al. [7] are applied as comparisons. Figure 4 illustrated the regionalized homogeneous precipitation catchments based on Ward^{LR}, Complete^{LR}, Complete^{MLR} and the proposed algorithm (Average^{P}), while table 2 illustrated the resulted p-value of K-sample Anderson Darling non-parametric test for various sample sizes. It should be highlighted that this study replaced Euclidean distance as similarity measure for Complete^{MLR} algorithm instead of correlation coefficient (Complete^{LR}). This is due to regionalize the precipitation catchments using Complete^{LR} (figure 4(b)) yielded an unreasonable result, which the 16 precipitation catchments classify as outlier catchments (orange color points).

In principle, the distribution of spatial and temporal is varied between the coastal and inland region [22,23]. Therefore, this study concluded the proposed regionalized algorithm performed more better compared to Ward^{LR}, Complete^{LR}, and Complete^{MLR} although K-sample Anderson Darling non-parametric test shows that all of the regionalized regions using Ward^{LR}, Complete^{LR}, Complete^{MLR} and Average^{LR} algorithms are significantly homogeneous. The proposed algorithm in this study isolated the precipitation catchments located Kuantan River Basin, Pahang into 2 homogeneous regions, namely inland and coastal regions, where 3 out of 20 precipitation catchments classify as outlier precipitation catchments. This study also relocated the misplaced of Catchment 20 as the proposed regionalized algorithm misplaced Catchment 20 in the inland region.



(a) (b) (c) (d)

Figure 4. (a), (b), (c) and (d) are the regionalized homogeneous precipitation catchments based on Ward^{LR}, Complete^{LR}, Complete^{MLR} and Average^{P}, respectively.

Table 2. The p-value of K-sample Anderson Darling nonparametric test based on $Ward^{LR}$,$Complete^{LR}$, $Complete^{MLR}$, and $Average^{P}$ algorithms, respectively.

Algorithm	Cluster	Catchment	P-value					
			100	500	1000	5000	10000	50000
$Ward^{LR}$	1	01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Complete^{LR}$	1	03, 04	0.6500	0.5600	0.5390	0.5178	0.5170	0.5155
	2	06, 11	0.5800	0.5740	0.5750	0.5708	0.5668	0.5661
$Complete^{MLR}$	1	01, 02, 03, 04, 06, 08, 09, 12, 14, 16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2	05, 07, 10, 11, 13, 15, 17, 19, 20	0.6100	0.5440	0.5500	0.5836	0.5804	0.5839
$Average^{P}$	1	04, 06, 07, 09, 12, 13, 17, 19, 20	0.6200	0.5980	0.5700	0.5558	0.5648	0.5699
	2	02, 03, 05, 10, 11, 14, 15, 16	0.5000	0.3840	0.3930	0.4256	0.4246	0.4266

5. Conclusion

This study proposed an efficient regionalized algorithm in identifying homogeneous precipitation catchments for non-stationary time series. The analysis result shows that the proposed regionalized algorithm successfully isolated 20 precipitation catchments located Kuantan River Basin, Pahang into 2 homogenous regions, namely inland and coastal regions. In addition, the analysis result shows that the proposed regionalized algorithm performed more better compared to the proposed regionalized algorithms in previous studies. For future works, this study suggested to further investigation of efficiency for other hierarchical clustering algorithms associated multi-scale bootstrap resampling, which correlation coefficient and absolute correlation coefficient as similarity measures.

Acknowledgement

The authors would like to express appreciation to Department of Irrigation and Drainage for providing the precipitation data used in this study. A word of appreciation also goes to the Universiti Pahang Malaysia for providing the flagship research grant RDU150393 and the internal research grant RDU1703184. The authors also extend appreciation to all reviewers who providing valuable comments.

References

[1] Guttman NB 1993 The use of L-moments in the determination of regional precipitation climates *Journal of Climate* **6** (12) 2309-2325.
[2] Burn DH, Zrinji Z and Kowalchuk M 1997 Regionalization of catchments for regional flood frequency analysis *Journal Hydrologic Engineering* **2** (2) 76-82.
[3] Venkatesh B and Jose MK 2007 Identification of homogeneous rainfall regimes in parts of Western Ghats region of Karnataka *Journal of Earth System Science* **116** (4) 321-329.
[4] Ngongondo CS, Xu C-Y, Tallaksen LM, Alemaw B and Chirwa T 2001 Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches *Stochastic Environmental Research and Risk Assessment* **25** (7) 939-955.
[5] Ahmad NH, Othman IR and Deni SM 2013 Hierarchical cluster approach for regionalization of Peninsular Malaysia based on the precipitation amount *Journal of Physics: Conference Series* **423** 012018.
[6] Pansera WA, Gomes BM, Boas MAV and Mello ELd 2013 Clustering rainfall stations aiming regional frequency analysis *Journal of Food, Agriculture & Environment* **11** (2) 877-885.

[7] Zakaria R, Radi NFA, Ngongondo CS and Azman MA 2014 A regional frequency analysis of annual maximum rainfall in Terengganu using the index rainfall and L-moments approaches *Proc. of the 10th Indonesia-Malaysia-Thailand Growth Triangle Int. Conf. on Mathematics, Statistics and Its Applications, 14th-16th October 2014, Kuala Terengganu* pp 667-674.
[8] Cavadias GS, Ouarda TBMJ, Bobée B and Girard C 2001 A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins *Hydrological Sciences Journal* **46** (4) 499-512.
[9] Hall MJ, Minns AW and Ashrafuzzaman AKM 2002 The application of data mining techniques for the regionalisation of hydrological variables *Hydrology and Earth System Sciences* **6**(4) 685-694.
[10] Lin G-F and Chen L-H 2006 Identification of homogeneous regions for regional frequency analysis using the self-organizing map *Journal of Hydrology* **324** (1-4) 1-9.
[11] Baeriswyl P-A and Rebetez M 1997 Regionalisation of precipitation in Switzerland by means of principal component analysis *Theoretical and Applied Climatology* **58** (1-2) 31-41.
[12] Singh KK and Singh SV 1996 Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall on sub-Himalayan region and Gangetic plains of India *Climate Research* **6**(3) 251-262.
[13] Tangang FT, Juneng L, Salimun E, Sei KM, Le LJ and Muhamad H 2012 Climate change and variability over Malaysia: Gaps in sciences and research information *Sains Malaysiana* **41** (11) 1355-1366.
[14] Agilan V and Umamahesh NV 2016 Is the covariate based non-stationary rainfall IDF curve capable of encompassing future rainfall changes? *Journal of Hydrology* **541** (Part B) 1441-1455.
[15] Viglione A, Laio F and Claps P 2007 A comparison of homogeneity tests for regional frequency analysis *Water Resources Research* **43** W03428.
[16] Alam MM, Morshed MG, Siwar C and Murad MW 2012 Initiatives and challenges of agricultural crop sector in East Coast Economic Region (ECER) development projects in Malaysia *American-Eurasian Journal Agriculture & Environmental Sciences* **12** (7) 922-931.
[17] Jackson R 2012 *Occupy World Street: A Global Roadmap for Radical Economic and Political Reform* (Hartford: Chelsea green).
[18] Win NL and Win KN 2014 The probability distributions of daily rainfall for Kuantan Basin in Malaysia *Int. Journal of Science and Research* **3**(8) 977-983.
[19] Saeed GAA, Chuan ZL, Zakaria R, Yusoff WNSW and Salleh MZ 2016 Determination of the best single imputation algorithm for missing rainfall data treatment *Journal of Quality Measurement and Analysis* **12** (1-2) 79-87.
[20] Mann HB 1945 Nonparametric tests against trend *Econometrica* **13**(3) 245-259.
[21] Lance GN and Williams WT 1967 A general theory of classificatory sorting strategies 1. hierarchical systems *The Computer Journal* **9**(4) 373-380.
[22] Hamada J-I, Yamanaka MD and Mori S 2008 Differences of rainfall characteristics between coastal and interior areas of central Western Sumatera, Indonesia *Journal of the Meteorological Society of Japan* **86** (5) 593-611.
[23] Azman MA, Zakaria R, Satari SZ and Radi NFA 2017 Entropy of stable seasonal rainfall distribution in Kelantan *AIP Conf. Proc.* **1842** 020002.