Contents list available at CBIORE journal website

Journal of Emerging Science and Engineering

Journal homepage: https://journal.cbiore.id/index.php/jese/index

Research Article

An insight from homogeneity testing of long-term rainfall datasets over East Java, Indonesia

Heri Mulyanti[a,b*], Istadi[a,c], Rahmat Gernowo[a,d]

[a]Doctoral Program of Environmental Science, Universitas Diponegoro, 50241, Semarang, Indonesia

[b]Faculty of Science and Engineering, Universitas Bojonegoro, 62119, Bojonegoro, Indonesia

[c]Department of Chemical Engineering, Universitas Diponegoro, 50275, Semarang, Indonesia

[d]Department of Physics, Faculty of Science and Mathematics, Universitas Diponegoro, 50275, Semarang, Indonesia

Abstract. . Robust, reliable, and trustworthy ground observation datasets are the preliminary requirement for assessing the impact of climate change over regions. Principal testing to assess the quality of ground observation rely on the missing data and homogeneity result. The study used 40 years of monthly rainfall documented from different topographical features in the monsoonal region of East Java, Indonesia. The test included annual rainfall, early rainy season (October-November-December), and primary rain season (January-February-March). The homogeneity of rainfall determined by absolute technique: Pettitt's test, the Standard Normal Homogeneity Test, the Buishand Rank Test, and the von Neumann Ratio. Among the time series, October-November-December observation results in better homogeneity. However, the rainfall datasets during primary rainy season showed the worst homogeneity. By performing annual and seasonal homogeneity test from 67 rainfall stations: 5 stations out of data length required, 5% stations 'rejected', 11% 'suspect', 11% 'doubtful', and 73% were 'trusted'. Therefore, a total of 45 stations can be used as metadata for relative comparison and 7 stations can be considered to be useful for analysis despite 'doubtful'. The remaining 10 stations need careful consideration to be used for future water management. Change point detected particularly between the year of 1997 through 2000. Pettitt's test has outstanding results in the case of extreme climatic anomaly, but less sensitive of continuous abrupt change. The von Neumann test could detect abnormal data, but was not suitable for datasets containing few extreme values. The insights from homogeneity testing were: a) it is important to remove any outliers in the datasets before conducting homogeneity testing, b) both parametric and nonparametric homogeneity tests should be performed, and c) comparisons should be made with surrounding rainfall stations. Comparison with trusted long-term rainfall data is valuable for stations labeled as 'doubtful' or 'suspect' to mitigate false detections in individual homogeneity tests. The identified 'useful' rainfall data can then serve as reference stations for relative homogeneity tests. These findings suggest that reference stations should be assessed within similar rainfall zones.

Keywords: rainfall, homogeneity, East Java, reference stations

Received: 15th April 2024; Revised: 18th May 2024; Accepted: 20th May 2024; Available online: 26th May 2024

1. Introduction

Human influence has affect the warming of the ocean, atmosphere, and land (Intergovernmental Panel On Climate Change (IPCC) 2023). The impacts apparently increasing on regional and local scales. Climate change has altered rainfall amount, distributions, and seasonal patterns (Anripa et al. 2023; Chen et al. 2023; Gu et al. 2020). As global warming continues, every region is expected to undergo more frequent and diverse changes in factors impacting climate. Therefore, knowledge on possible outcomes and assessment of climate-related risks regarding as important path for reducing the climate risk to the community. The first initial step is to ensure valid and reliable climate dataset as climatological analysis rely on data documentation.

Robust, reliable, and trustworthy ground observation dataset are the preliminary requirement for assessing the impact of climate change over regions (Rummukainen 1997, 2016). However, the majority of long-term raw climate time series do not adhere to this principle and are internally heterogeneous. Heterogeneous time series meaning they are not uniform in nature and consequently unsuitable for statistical climate change analysis. Non-climatic factors, such as alterations in measurement location, can significantly influence values and caused distortions in the statistical behavior of the time series. These distortions may lead to erroneous conclusions since the abrupt can be comparable to climate change itself (Aguilar et al. 2003). Therefore, it's crucial to conduct a thorough investigation to identify and rectify errors in the data before studying observational records.

Homogeneity tests can be categorized into absolute and relative tests. Absolute tests are based on the characteristics of individual stations, while relative homogeneity utilizes data from nearby stations with similar characteristics (Aguilar *et al.* 2003; Wijngaard *et al.* 2003). A time series becomes non-homogeneous when the mean of the series does not remain stable over time or when changes appear in the series. Climate series is considered homogeneous when its variations are caused solely by changes in weather and climate, without any influence from non-climatic factors that mislead the true climate variation.

* Corresponding author
Email: herimulyanti@students.undip.ac.id (H. Mulyanti)

https://doi.org/10.61435/jese.2024.e23
ISSN: p-ISSN: 3026-0817; e-ISSN: 3026-0183.The Author(s). Published by CBIORE

The study examines the homogeneity of long-term rainfall datasets in East Java as a preliminary step for climate change mitigation. East Java Province is one of the drought-prone regions on the island of Java (Aldrian and Djamil 2008; D'Arrigo et al. 2006; Mulyanti et al. 2023). East Java is identified as a priority area for drought management in Indonesia. However, rainfall data testing has been relatively limited for East Java Province. Previous studies have tested the spatial trends of rainfall (Aldrian and Djamil 2008; Avia 2019; Wahab et al. 2009).

Therefore, the aim of this study is to spatially map the results of homogeneity tests to evaluate the performance of rainfall datasets. Although rainfall testing has been widely conducted (Akinsanola and Ogunjobi 2017; Arikan and Kahya 2019; Ay 2020; de Gois et al. 2020; Hadi et al. 2019; Kabbilawsh et al. 2023; Yaşa et al. 2023), comprehensive examinations of how data may fail one test but succeed in another have not yet been undertaken. This study emphasizes the importance of meticulously identifying rainfall data by distribution and time series patterns. It aims to enhance understanding of the appropriateness of each homogeneity test..

2. Data and Methods

2.1 Data

This study utilized data from 61 rain gauge stations distributed throughout East Java Province. The gauge stations were distributed based on the north-south topography, with a higher concentration in the western region and fewer in the east. The data were documented by the Department of Public Works and Water Resources of East Java Province. The selected rain gauge stations for analysis should be spanning over 20 years. The data types included annual total rainfall and rainfall during the early and primary rainy season. In East Java, September is typically a dry month. Consequently, the early rainy season is defined as October-November-December (OND), while the primary rainy season is defined as January-February-March (JFM). Rainfall data from the dry season were excluded from homogeneity testing due to their high vulnerability (As-syakur et al. 2016; Tanaka 1994). Missing data accounted for less than 5% of the dataset and were imputed using the time series mean of the corresponding month.

The rain gauge stations are evenly distributed across East Java Province, both meridionally and latitudinally. Observations indicate that the rainfall characteristics on Java Island are influenced by topography. Horizontally, the region can be divided into western and eastern parts. Vertically, it can be categorized into the northern coast, central lowlands, and southern mountains. Fig. 1 shows the distribution of rain gauge stations in East Java Province, covering the entire topography, both vertical and horizontal.

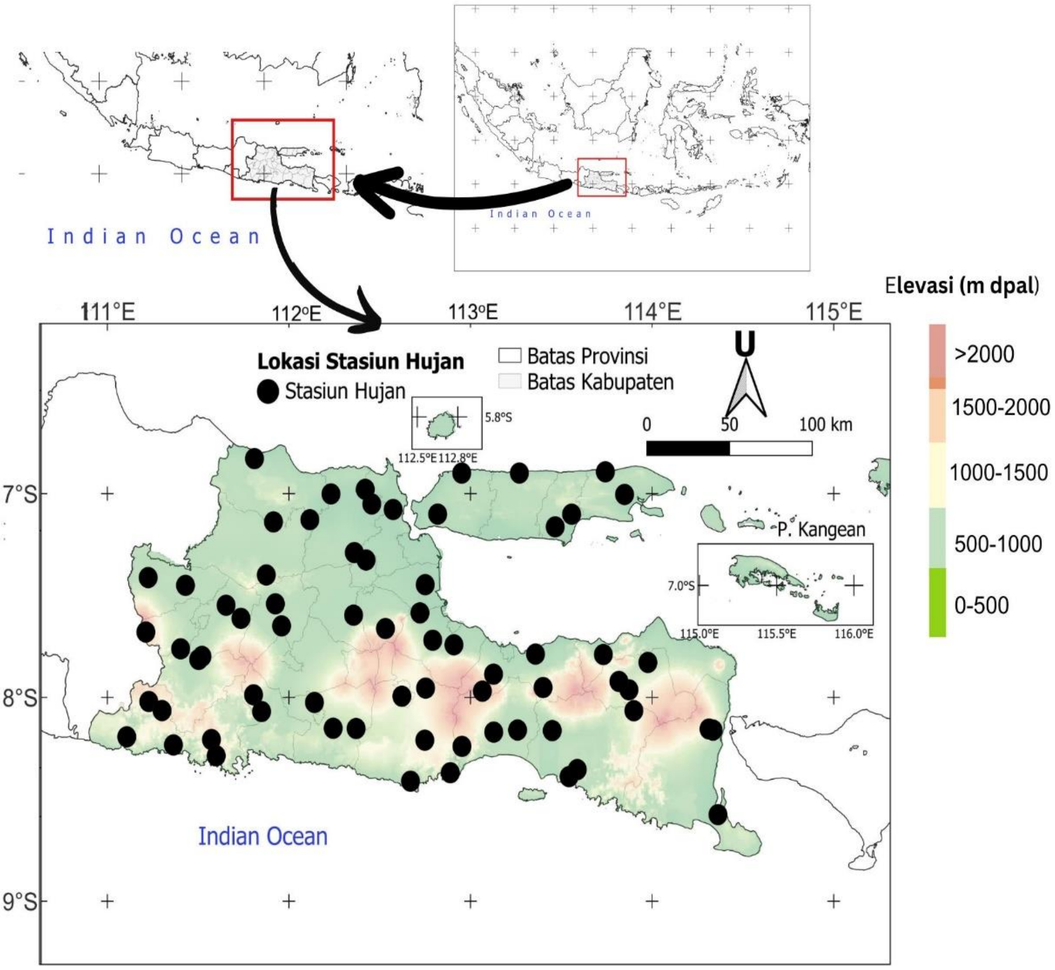

Fig. 1 Location of rain gauge for homogeneity test

2.2 Methods

Non-climatic factors included on the climate documentary can be elude using homogeneity test. The absolute homogeneity test of 61 stations were assessed using 4 methods: Pettitt test, the Standard Normal Homogeneity Test (SNHT), the Buishand Ratio Test (BRT), and the von Neumann ratio test. Hypothesis testing the homogeneity test are: $\boldsymbol{H_0}$ the data homogeneous, while alternate hypothesis $\boldsymbol{H_a}$ that there is found a break inside the time series dataset for significance level of 5%. Pettitt, SNHT, and Buishand methods are categorized as break year identifier for time series.

Homogeneity testing was conducted using four methods: Pettitt's, SNHT, Buishand, and von Neumann. These four tests have been widely employed in various studies (Akinsanola and Ogunjobi 2017; Arikan and Kahya 2019; Ay 2020). The rainy season was chosen as a complement to annual testing because rainfall is concentrated at over 40% in JFM. Meanwhile, OND rainfall accounts for approximately 30% of the total annual rainfall.

Pettitt's Test

Pettitt's test is conducted to observe any significance change of data series by transforming observation values into rankings (Pettitt 1979). The test categorized as nonparametric statistic. The highest value is assigned rank number 1, descending to the lowest value of N (Arikan and Kahya 2019).

a. The Pettitt statistical calculation consider the value V, which is the sum of observations substracted by the ranking of i, computed as follows:

$$V_i = (N+1) - 2R_i \qquad (1)$$

where: R_i represents the ranking of i, and N denotes the total observations.

b. The value U is used to determine the maximum cumulative of V_i formulated as:

$$U_i = U_{i+1} + V_i \qquad (2)$$

Because there is a possibility of negative values, U_i converted into absolute form

$$U_{i,abs} = |U_i| \qquad (3)$$

The largest value of $U_{i,abs}$ then called as P_T. The homogeneity of the time series then determined by comparing the value of Pettitt's p with α = 0.05. The Pettitt's p calculated as:

$$p = 2\ exp\left(\frac{-6{P_T}^2}{T^3+T^2}\right) \qquad (4)$$

where p = Pettitt's value, P_T is the largest value of $U_{i,abs}$ from Eq (3), T is total observations. Once the value of p smaller than the α, the null hypothesis rejected and means that there is a substantial change on the dataset.

Standard Normal Homogeneity Test (SNHT)

This test was conducted to detect significant changes by transforming the series into normal z scores (Alexandersson 1986). In essence, this technique attempts to divide data into two phases, before *T(n)* and after *T(n)*. If there are significant changes between the data before and after T(n), then the data is considered inhomogeneous. This test operates under the assumption that data with non-normal distribution must be transformed into standard normal (z) to adhere the standards. SNHT is defined as (Alexandersson, 1986):

$$T(k) = k\overline{z_\iota}^2 + (n-k)\overline{z_2}^2 \text{ where } k = 1, 2, \ldots\ldots, n \qquad (5)$$

where $\overline{z_1}$ is the mean of z-value mean series before, and $\overline{z_2}$ is the mean of z-value after the break year to n number of observations.

The probable break year of the record is assumed to be the maximum value of $T(k)$ and can be formulated as:

$$T_n = \max\{T(k)\}\ for\ 1 \le k \le n \qquad (6)$$

Table 1
Critical value of T_0 based on level of significance for SNHT

sign.	Total observation (n)					
	20	30	40	50	70	100
1%	9.56	10.45	11.01	11.38	11.89	12.32
5%	6.95	7.65	8.10	8.45	8.80	9.15

Source: Arikan and Kahya (2019)

The relationship between $T(n)$ and T_0 are calculated as (Jarušková 1996):

$$T_0 = \frac{T(n)^2 n}{T(n)^2 + (n-2)} \quad (7)$$

The critical value of T_0 represented as (Arikan and Kahya 2019) in Table 1

Hypothesis testing for T_0 as follow:
H_0 the data homogeneous
H_a the data non homogeneous
H_0 is rejected if T_0 is above certain level of based on Table…

Buishand Range Test (BRT)

The Buishand range test (Buishand 1982) calculate the adjusted partial sums of the statistics to identify any inhomogeneities. The test can actually be used for various types of distributions, but it is more suitable for normally distributed data. The Buishand test is sensitive to breakpoints in the middle of a time series (Hawkins 1977). The adjusted partial sums of the statistics are calculated:

$$S_0^* = 0 \; and \; S_k^* = \sum_{i=1}^{k} (Y_i - \bar{Y}) \text{ where } k = 1, \ldots, n \quad (8)$$

In a homogeneous series, the values will oscillate around zero, as there are no systematic deviations of the Yi values relative to their mean. However, if a break occurs in year K, the value of S_k^* reaches maximum or minimum around the year k = K.
The maximum and minimum value of S_k^* are used to calculate the rescaled adjusted range, R statistics. The R statistic is value to accept or reject null hypothesis.

$$R = \frac{\left(\max_{0 \le k \le n} S_k^* - \min_{0 \le k \le n} S_k^*\right)}{s} \quad (9)$$

Here, s denoted as standard deviation of the dataset. The $R/\sqrt{N}$ is then compared to the critical value (Buishand 1982). H_0 rejected if the value of $R/\sqrt{N}$ is greater than critical value from Table 2. The Pettitt's, Buishand Range Test, and SNHT could detect the break year of the record.

Table 2
Critical value of T_0 based on level of significance of Buishand Range Test

sign.	Total observation (n)					
	10	20	30	40	50	100
5%	0.93	1.43	1.50	1.53	1.55	1.62

Source: *Buishand* (*1982*)

Von Neumann

The von Neumann ration is defined for the expectation of N >2 as homogeneous.

$$N = \frac{\sum_{i=1}^{n-1} (Y_i - Y_{i+1})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \quad (10)$$

where, Y_i is the observed data of i year and $\bar{Y}$ denotes the mean of the time series:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y \quad (11)$$

Four homogeneity criteria are established across the four tests: a) if all tests pass, it is considered 'trusted'; b) if one test fails, it is categorized as 'doubtful'; c) if two tests fail, it is categorized as 'suspect'; and d) if 3-4 tests fail, it is deemed 'reject'. The tests are conducted for both annual and seasonal rainfall. The homogeneity status is determined by the lowest criterion in the annual, OND, and JFM tests. If any test meets the 'doubtful' criterion, the final status is 'doubtful'. If any test meets the 'reject' criterion, the homogeneity status is 'reject'.

3. Results and Discussions

3.1 Annual Testing

Fig. 2 shows that 17% of stations did not pass the von Neumann homogeneity test of annual time series, whereas only 4% - 7% of stations failed the Buishand and Pettitt tests. A total of 7 stations (10%) failed the SNHT test. Both the Pettitt and SNHT tests reveal the similar stations experiencing inhomogeneous conditions. In contrast, stations failing the Buishand test are located on Madura

Island. The Buishand results for Madura Island closely resemble those of the SNHT, but exhibit discrepancies for the main island region. The von Neumann test identifies stations failing the test in a distinct pattern compared to the other three tests.

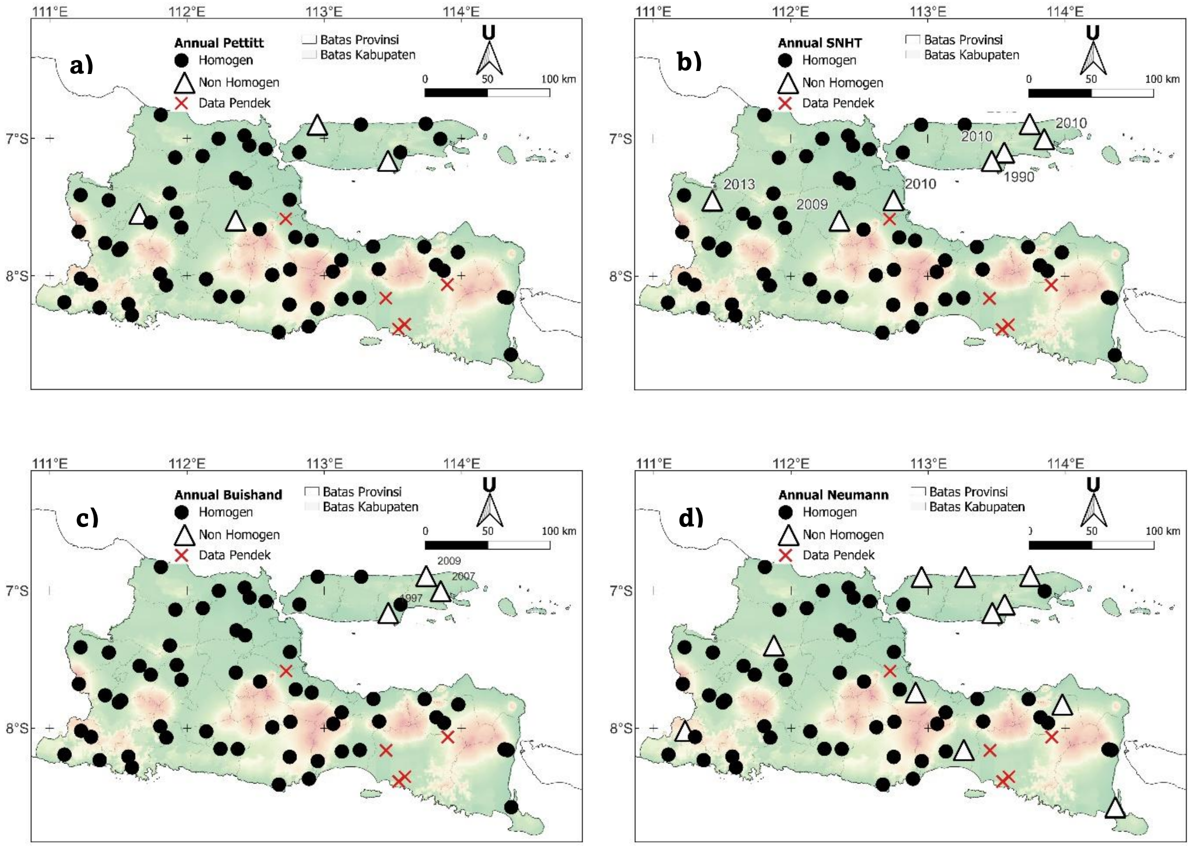

Fig. 2 Homogeneity test for total annual rainfall, a) Pettitt's, b) SNHT, c) Buishand, d) von Neumann.

The break years identified by both the Pettitt and SNHT tests typically align closely, often differing by only one year (Table 3). Similar break years indicate a difference of only one year between tests. Break years are considered quite similar when two out of three series show similarity. Dissimilar break years occur when the three tests indicate different times. From Table 3, it can be inferred that break years primarily occurred after 2000. However, the years causing inhomogeneity in the time series are located in 1990, 2000, and 2010. SNHT demonstrates sensitivity in detecting the break year primarily at the beginning or end of the time series, consistent with prior research findings (Ay 2020; Toreti et al. 2011). The von Neumann criteria for discerning non-homogeneity prioritize the data's mean. The red crosses symbol marked the stations with short length of data.

Table 3
Change point detected from annual dataset. Nonhomogeneous series bold with color.

No	Station	n (year)	Break Year			Similarity Year
			Pettitt	SNHT	Buishand	
1	Tuban Belikanget	39	2001	2002	2001	S
2	Bojonegoro	42	1987	1987	1987	S
3	Baureno	42	1997	1997	1997	S
4	Bojonegoro-Sugihan	42	2005	2007	2006	S
5	Lamongan Ngimbang	40	2009	2009	2008	S
6	Lamongan Jabung	41	2004	1994	1994	QS
7	Lamongan Blawi	41	2013	2015	2014	S
8	Gresik Tambakombo	40	2001	2002	2001	S
9	Gresik Lowayu	40	1997	1997	1997	S
10	Situbondo Wringinanom	30	2013	2014	2013	S
11	Madiun Caruban	41	2001	2002	2001	S

No	Station	n (year)	Break Year			Similarity Year
			Pettitt	SNHT	Buishand	
12	Madiun Gemarang	41	1997	1997	1997	S
13	Madiun Gombal	30	2001	1991	2001	QS
14	Sidoarjo-Budug Bulus	37	1992	1992	1992	S
15	Surabaya Gubeng	42	2004	2010	2009	QS
16	Pasuruan P3GI	29	1999	2009	2009	QS
17	Pasuruan Kasri	31	2009	2009	2009	S
18	Probolinggo Segaran	31	2003	2002	2001	S
19	Probolinggo Ronggo Tali	30	1997	1993	1997	QS
20	Probolinggo Adiboyo	31	2005	1991	2005	DS
21	Problolinggo Lumbang	31	2008	2009	2009	S
22	Nganjuk Matokan	27	2003	2003	2003	S
23	Nganjuk Ngrambek	27	2003	2009	2009	QS
24	Mojokerto Pacet	40	1993	1980	1992	QS
25	Magetan Lambeyan	39	2009	1980	2009	QS
26	Magetan Sarangan	40	1999	1980	1999	QS
27	Jombang Mojoagung	30	2001	2009	2008	QS
28	Ngawi Walikukun	38	2001	2011	2010	QS
29	Sumenep Ambunten	42	2009	2010	2009	S
30	Sumenep Kebonagung	42	2009	2010	2009	S
31	Sampang Ketapang	42	2001	2001	1998	QS
32	Pamekasan Pakong	36	2009	2010	2009	S
33	Pamekasan Samiran	34	1990	1990	2019	QS
34	Bangkalan Sepulu	40	1990	1990	1990	S
35	Bangkalan Tragah	42	1996	2015	2009	DS
36	Banyuwangi Alasbulu	24	2015	2016	2015	S
37	Banyuwangi Dadapan	26	2015	2016	2015	S
38	Malang Dampit	21	2010	2016	2015	QS
39	Trenggalek Dongko	31	1997	1997	1997	S
40	Blitar Gandusari	21	2010	2014	2010	QS
41	Malang Jabung	21	2013	2014	2013	S
42	Blitar Kayumanis	21	2009	2018	2009	QS
43	Lumajang Kedungwringin	16	2015	2016	2015	S
44	Banyuwangi Licin	26	2011	2011	2011	S
45	Blitar Lodoyo	21	2009	2000	2009	S
46	Trenggalek Munjungan	31	2002	2002	2002	S
47	Malang Pagak	21	2009	2000	2009	QS
48	Tulungagung Pagerwojo	21	2015	2009	2009	QS
49	Tulungagung Paingan	21	2002	2017	2003	QS
50	Lumajang Pronojiwo	30	1995	1992	1995	QS
51	Malang Sitiarjo	21	2003	2009	2009	QS
52	Kota Malang Sukun	21	2012	1999	2012	QS
53	Lumajang Wonokerto	40	1990	1990	1990	S
54	Bondowoso Glendengan	37	2006	2007	2006	S
55	Bondowoso Kesemek	24	2012	2013	2012	S
56	Bondowoso Sentral PU	32	2007	1999	1999	QS
57	Ponorogo Babadan	42	2001	2016	2012	DS
58	Ponorogo Balong	40	2012	2013	2012	S

No	Station	n (year)	Break Year			Similarity Year
			Pettitt	SNHT	Buishand	
59	Pacitan Nawangan	41	1989	2010	2009	QS
60	Pacitan Pacitan	41	1997	2018	1997	QS
61	Pacitan Sudimoro	41	1994	1981	1983	DS
62	Pacitan Tegalombo	40	2009	2010	2009	S

Where: S = similar, QS = quite similar, and DS = dissimilar

3.1 Seasonal Testing

The occurrence of inhomogeneous rainfall time series is less frequent at the onset of the rainy season and becomes more prevalent during the main rainy season. Testing suggests that seasonal rainfall is more susceptible to inhomogeneity compared to annual testing. In the JFM period, stations identified as non-homogeneous based on the Pettitt, SNHT, Buishand, and von Neumann tests are 10%, 15%, 3%, and 26%, respectively. Conversely, the number of non-homogeneous stations for OND is lower across all four tests, at 3% (Pettitt), 4% (SNHT), 3% (Buishand), and 12% (von Neumann). Nonhomogeneous data is predominantly concentrated in Madura Island, consistent with the findings of annual data testing. Both annual and seasonal tests suggest that von Neumann is more sensitive to changes in the data series compared to other tests.

Fig. 3 illustrates that the break years for OND occurred in 1997 and 2020 based on the Buishand test. The year 1997 is considered an anomaly due to the El Niño event. Break years identified by SNHT during a similar season span between 2000 and 2018 and are found in three locations. However, the von Neumann test reveals a larger number of failed observations, with seven locations regarded as inhomogeneous.

The highest number of failed tests occurred during JFM (Fig. 4), which is the primary rainy season. At least 7 stations failed in Pettitt's test, while 9 stations failed in SNHT. However, the Buishand results are comparatively better across all tests. Von Neumann yielded the poorest result, with 18 stations failing. Stations that failed the test are concentrated in separate topographic areas but are rarely found near Malang. The most frequent occurrences of inhomogeneity are observed in Madura Island, the northern coast of the main island, and the western part of the island. The primary rainy season accounts for approximately 40% of the total rainfall, and the inhomogeneity of this season may affect the overall data quality. Break years for JFM occurred during La Niña events, such as 2016, 2001, and 1990.

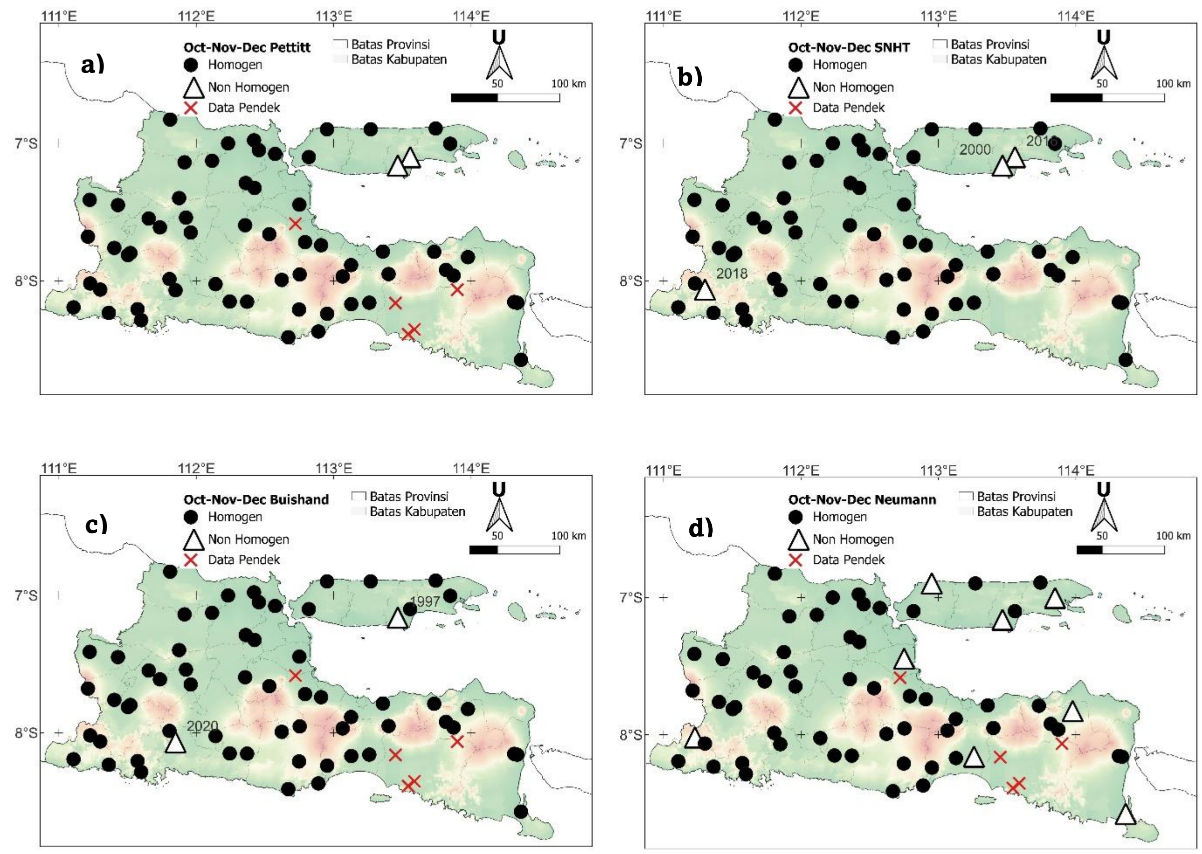

Fig. 3 Homogeneity test for total October-November-December rainfall, a) Pettitt's, b) SNHT, c) Buishand, d) von Neuman

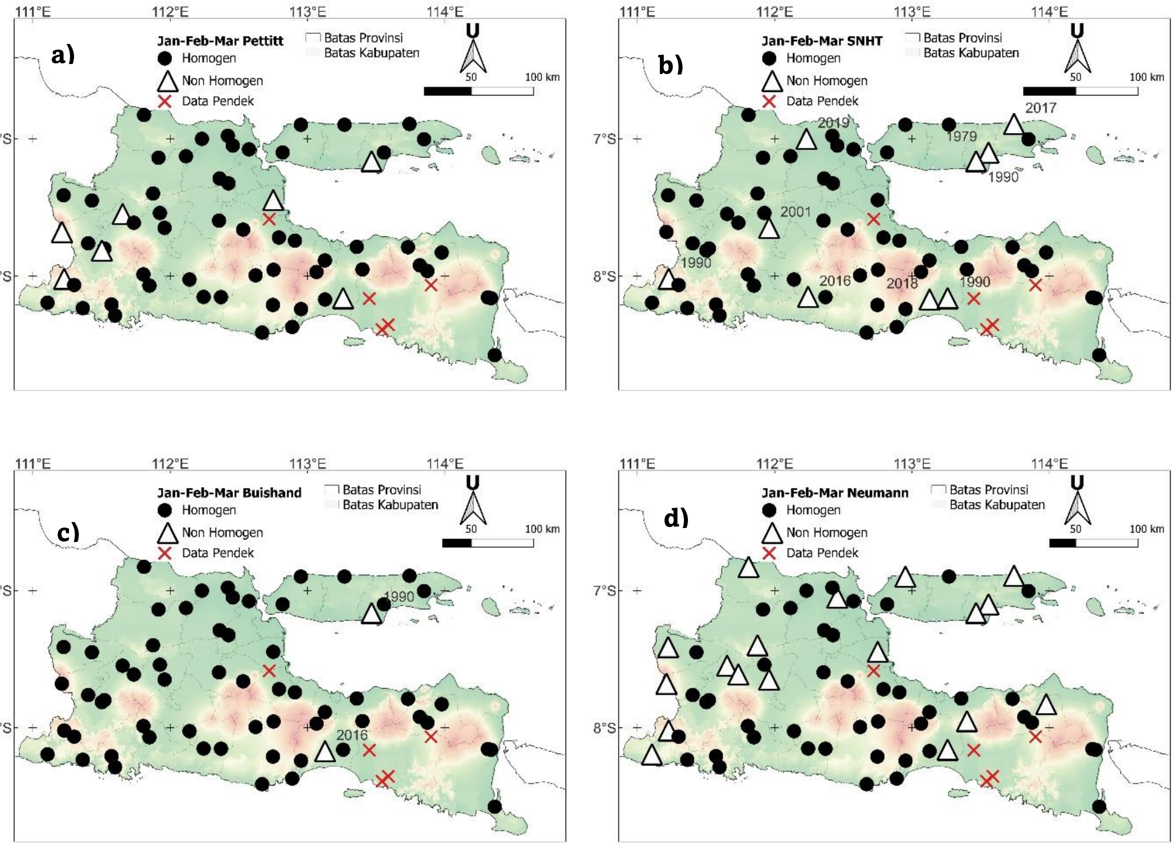

Fig. 4 Homogeneity test for total January-February-March rainfall, a) Pettitt's, b) SNHT, c) Buishand, d) von Neumann

3.3 Insights from Homogeneity Test

3.3.1 Pettitt's Test

Fig. 5 provides a clearer illustration of how the Pettitt test does not consider continuous changes in the same positive/negative direction as an inhomogeneity. An example of this occurs at the Madiun-Caruban station and the Sumenep-Ambunten station. While Sumenep-Ambunten passed Pettitt's test but failed another test, Madiun-Caruban only failed Pettitt's change detection. Comparing the average rainfall before and after the breakpoint year reveals insights into Pettitt's limitation in identifying breakpoints when the dataset experiences uniform changes. This occurs due to the lack of variation in the ranking of data points. Figure ... illustrates significant variation in data ranking after the breakpoint year, with values either markedly high or low. This contrasts somewhat with the observational data at Sumenep-Ambunten, where ranking changes occur gradually.

Pettitt's test involves ranking data which suitable for very skewed data. The test apply naturally to datasets presented in ranks where numerical distinction lack interpretability (Krzywinski and Altman 2014). However, nonparametric testing can handle unexpected, outlying observations (Tsybakov 2009; Whitley and Ball 2002) that may arise due to climate anomalies. For normally distributed data, it is advisable to obtain hypothesis testing by using parametric test (Whitley and Ball 2002). For the Madiun-Caruban case, the reliability of Pettitt's test status could be elevated from 'doubtful' into 'trusted'.

An opposite case which occurred in Sumenep-Ambunten proven how Pettitt's less sensitive to continuous abrupt change (Fig. 6). Despite passing the nonparametric Pettitt test, this station failed three other tests. Upon closer examination, the Pettitt test failed to detect significant data changes post-2009 indicative of inhomogeneities. Here, presented that Inhomogeneities in the Pettitt test may be less accurately identified when nonhomogeneous data coincide with gradual changes.

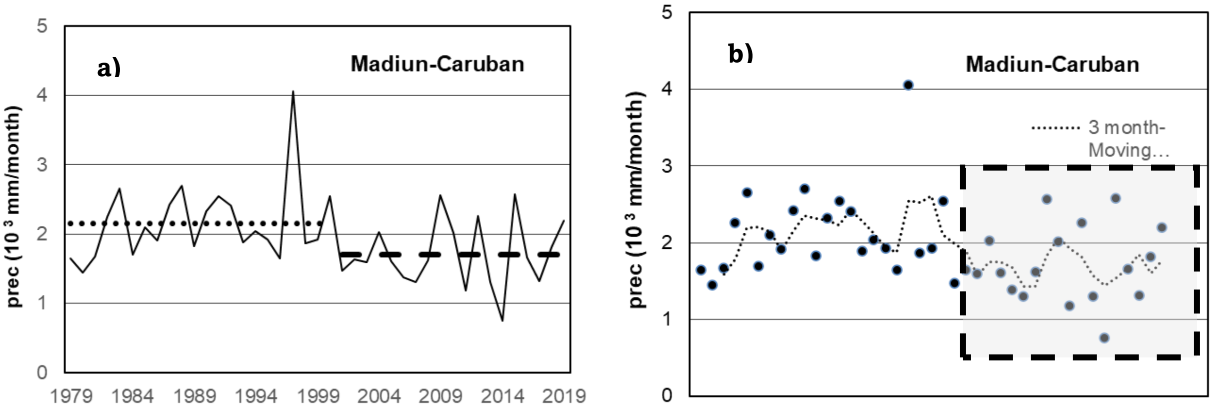

Fig. 5 Station which failed Pettitt's test, but passed 3 other tests. a) Madiun-Caruban station comparison before and after break year, which did not pass the Pettitt test but passed the other three tests; b) Madiun-Caruban 3-month moving average. Dashed box indicate break of the record

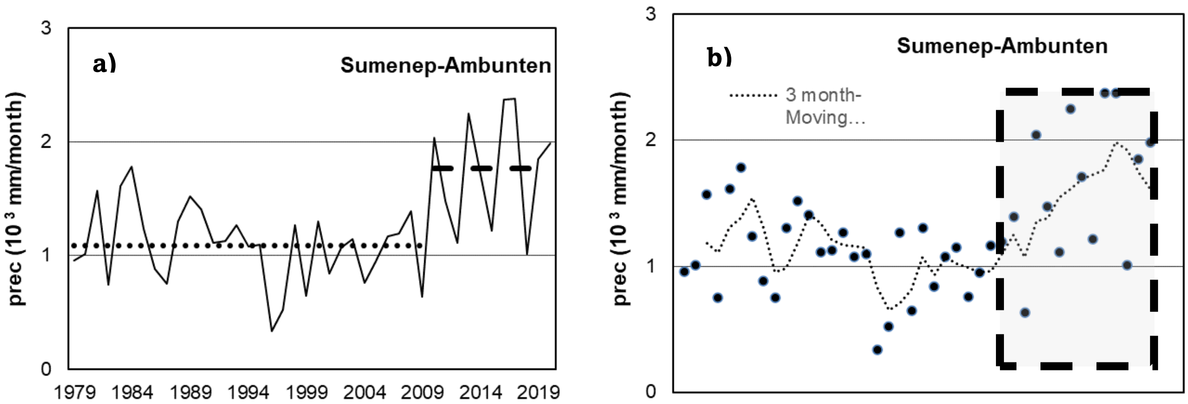

Fig. 6 Station which passed Pettitt's test, but failed 3 other tests. a) Sumenep-Ambunten station comparison before and after break year, which passed the Pettitt's but failed other three tests; b) Sumenep-Ambunten 3-month moving average. Dashed box indicate break of the record

3.3.2 SNHT

The SNHT test yields results similar to Pettitt, detecting the presence of break years. The difference is that SNHT is a parametric test that compares the original data before and after the change. The only station that did not pass the SNHT test is Ponorogo-Balong. Based on Fig. 7, the break year is identified at the end of the time series. Similar results are also mentioned in previous research. (Pandžić et al. 2020; Toreti et al. 2011). Upon closer inspection, the changes in Ponorogo-Balong can be considered relatively small. However, the results of the moving average indicate that significant changes occurred at the end of the time series. Change point may be caused by both climatic and non-climatic event. However, break point or inhomogeneity denotes a change point caused by non-climatic factors.

Hence, it is important to emphasize that a change point signifies the influence of one or more factors, both climatic and non-climatic whereas a break point or inhomogeneity denotes a change point instigated by non-climatic factors. The primary objective of homogenization is to eliminate non-climatic factors while preserving the climatic signal. SNHT, along with other methodologies, should be utilized by comparing series with a reference to ascertain reliable break points (Aguilar et al. 2003; Alexandersson and Moberg 1997; Peterson and Easterling 1994; Toreti et al. 2011). Furthermore, it's imperative to consider the limitations of this test during the homogenization procedure. Incorrect application of homogenization techniques to climate data could result in unreliable climate analyses. Employing multiple tests to compare detected break points is an effective strategy, particularly when metadata is unavailable. This approach yields robust results, preventing overestimation and correction of false inhomogeneities.

The Pettitt test shares similarities with SNHT. Fig. 8 illustrates rainfall time series data that failed both Pettitt and SNHT in Jombang-Mojoagung. As a complementary measure, Pettitt is capable of identifying changes in the time series in the middle of the data. It seems that SNHT is able to identify drastic changes in the data at the beginning and end, as seen in the Ponorogo-Balong station where the break year is 2013. The large difference between the ranks of observations after the break year presents a notable issue. Distribution analysis suggests that fluctuations after the break year cannot be interpreted as typical climatic signals. Climatic signals typically exhibit regular periodicity, which is absent in the dataset. Therefore, a combined approach of parametric and nonparametric rainfall testing may provide a more thorough assessment.

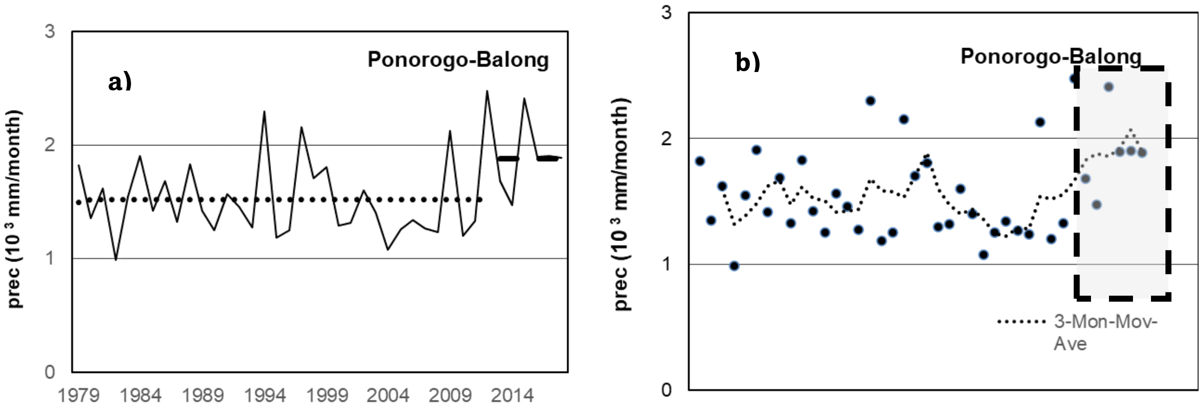

Fig. 7 Station which failed SNHT solely. a) Ponorogo-Balong station comparison before and after break year, which did not pass the SNHT test but passed the other three tests; b) Sumenep-Ambunten 3-month moving average. Dashed box indicate break of the record.

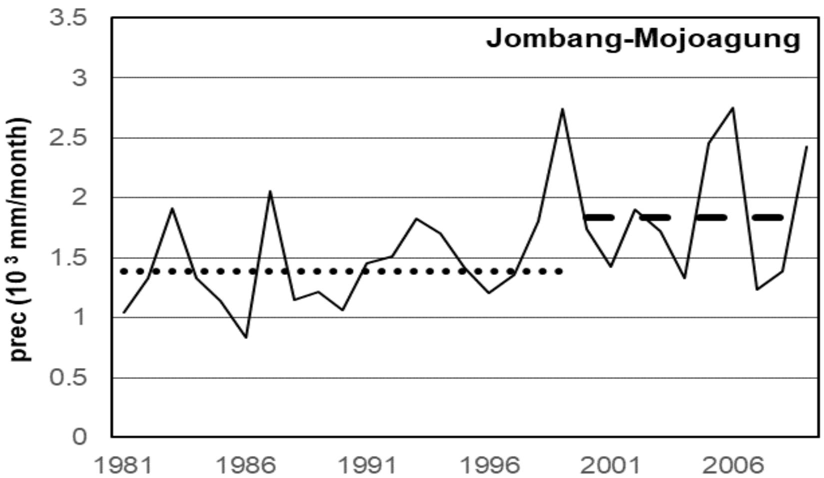

Fig. 8 Station which failed SNHT and Pettitt's, change point detected at 2001.

3.3.3 Buishand Range Test

The Buishand test yields the highest positive results compared to other tests. Only 3 stations display homogeneity issues. Buishand does not detect non-homogeneity on its own. Non-homogeneous data detected by Buishand align with SNHT and von Neumann tests. For instance, the Sumenep-Ambunten station fails Buishand, SNHT, and Pettitt tests. Thus, solely relying on the Buishand test is inadequate for assessing rainfall data homogeneity.

3.3.4 von Neumann Test

Fig. 9 illustrates stations which failing the von Neumann test. This test is sensitive to changes at individual observation points because affected value of mean, $x_i - x_{i-1}$ and $x_i - \overline{X}$ of i^{th} observation. If the value is greater than 2, it indicates homogeneous data. Solely relying on the von Neumann test regarding as insufficient for identifying inhomogeneity. For example, at the Bondowoso-Glendengan, a single extreme point of 6000 mm of annual rainfall record caused the data fall into inhomogeneous. Despite failed into von Neumann test, the station passed Pettitt, SNHT, and Buishand. Previous simulations provide insight that von Neumann is sensitive to detecting slightly abrupt changes along the time series. It appears that von Neumann is capable of detecting anomalies of non-climatic signals. Climatic signals should exhibit repetitiveness, but in cases of von Neumann rejection, the extreme values tend to be individual.

Based on the von Neumann test, it is evident that extreme data points should be excluded from the records. This conclusion arises from the observation that nearly all data rejected H_0 where extreme high or low values found. Regarding the removal of 'outliers', careful consideration is needed. The issue at the Bondowoso-Glendeng station stems from exceptionally high rainfall during the strong La Niña event in 2000, exceeding 6,000 mm/year.

Fig. 9b shows Bondowoso-Glendengan relative to the closest stations. Rainfall amount increasing on the same year, 2000, especially Bondowoso-PU reaching 4,000 mm/year. Somehow, between climatic or nonclimatic anomaly confusing in the case of absolute homogeneity. The matter arises from relative homogeneity appears because different topographic feature. In relation to La Niña event, rainfall in Java Island differ a lot between north and south coast (Qian et al. 2010). However, a careful consideration related to climatic or nonclimatic anomaly need to be performed.

Table 4
Homogeneity status of station based on annual, October-November-December, and January-February-March

Num	Station	N (year)	Annual	OND	JFM	Status of Homogeneity
1	Tuban Belikanget	39	Trusted	Trusted	Doubtful	Trusted
2	Bojonegoro	42	Trusted	Trusted	Trusted	Trusted
3	Baureno	42	Trusted	Trusted	Trusted	Trusted
4	Bojonegoro-Sugihan	42	Doubtful	Trusted	Doubtful	Doubtful
5	Lamongan Ngimbang	40	Trusted	Trusted	Trusted	Trusted
6	Lamongan Jabung	41	Trusted	Trusted	Trusted	Trusted
7	Lamongan Blawi	41	Trusted	Trusted	Doubtful	Trusted
8	Gresik Tambakombo	40	Trusted	Trusted	Trusted	Trusted
9	Gresik Lowayu	40	Trusted	Trusted	Trusted	Trusted
10	Situbondo Wringinanom	30	Trusted	Trusted	Trusted	Trusted
11	Madiun Caruban	41	Doubtful	Trusted	Suspect	Suspect
12	Madiun Gemarang	41	Trusted	Trusted	Doubtful	Trusted
13	Madiun Gombal	30	Trusted	Trusted	Trusted	Trusted
14	Sidoarjo-Budug Bulus	37	Trusted	Trusted	Trusted	Trusted
15	Surabaya Gubeng	42	Doubtful	Doubtful	Suspect	Suspect
16	Pasuruan P3GI	29	Doubtful	Trusted	Trusted	Doubtful
17	Pasuruan Kasri	31	Trusted	Trusted	Trusted	Trusted
18	Probolinggo Segaran	31	Trusted	Trusted	Doubtful	Trusted
19	Probolinggo Ronggo Tali	30	Trusted	Trusted	Trusted	Trusted
20	Probolinggo Adiboyo	31	Trusted	Trusted	Trusted	Trusted
21	Problolinggo Lumbang	31	Trusted	Trusted	Trusted	Trusted
22	Nganjuk Matokan	27	Trusted	Trusted	Trusted	Trusted
23	Nganjuk Ngrambek	27	Trusted	Trusted	Suspect	Suspect
24	Mojokerto Pacet	40	Trusted	Trusted	Trusted	Trusted
25	Magetan Lambeyan	39	Trusted	Trusted	Trusted	Trusted
26	Magetan Sarangan	40	Trusted	Trusted	Suspect	Suspect
27	Jombang Mojoagung	30	Suspect	Trusted	Trusted	Suspect
28	Ngawi Walikukun	38	Trusted	Trusted	Doubtful	Trusted
29	Sumenep Ambunten	42	Reject	Trusted	Suspect	Rejected
30	Sumenep Kebonagung	42	Suspect	Doubtful	Doubtful	Suspect
31	Sampang Ketapang	42	Doubtful	Trusted	Trusted	Doubtful
32	Pamekasan Pakong	36	Suspect	Suspect	Suspect	Suspect
33	Pamekasan Samiran	34	Reject	Reject	reject	Rejected
34	Bangkalan Sepulu	40	Suspect	Doubtful	Doubtful	Suspect
35	Bangkalan Tragah	42	Trusted	Trusted	Trusted	Trusted
36	Banyuwangi Alasbulu	24	Doubtful	Doubtful	Trusted	Doubtful
37	Banyuwangi Dadapan	26	Trusted	Trusted	Trusted	Trusted
38	Malang Dampit	21	Trusted	Trusted	Trusted	Trusted

Num	Station	N (year)	Annual	OND	JFM	Status of Homogeneity
39	Trenggalek Dongko	31	Trusted	Trusted	Trusted	Trusted
40	Blitar Gandusari	21	Trusted	Trusted	Trusted	Trusted
41	Malang Jabung	21	Trusted	Trusted	Trusted	Trusted
42	Blitar Kayumanis	21	Trusted	Trusted	Trusted	Trusted
43	Lumajang Kedungwringin	16	Trusted	Trusted	Trusted	Trusted
44	Banyuwangi Licin	26	Trusted	Trusted	Trusted	Trusted
45	Blitar Lodoyo	21	Trusted	Trusted	Trusted	Trusted
46	Trenggalek Munjungan	31	Trusted	Trusted	Trusted	Trusted
47	Malang Pagak	21	Trusted	Trusted	Trusted	Trusted
48	Tulungagung Pagerwojo	21	Trusted	Trusted	Trusted	Trusted
49	Tulungagung Paingan	21	Trusted	Trusted	Trusted	Trusted
50	Lumajang Pronojiwo	30	Trusted	Trusted	Trusted	Trusted
51	Malang Sitiarjo	21	Trusted	Trusted	Trusted	Trusted
52	Kota Malang Sukun	21	Trusted	Trusted	Trusted	Trusted
53	Lumajang Wonokerto	40	Doubtful	Doubtful	Reject	Rejected
54	Bondowoso Glendengan	37	Doubtful	Doubtful	Doubtful	Doubtful
55	Bondowoso Kesemek	24	Trusted	Trusted	Trusted	Trusted
56	Bondowoso Sentral PU	32	Trusted	Trusted	Trusted	Trusted
57	Ponorogo Babadan	42	Trusted	Trusted	Doubtful	Trusted
58	Ponorogo Balong	40	Doubtful	Trusted	Trusted	Doubtful
59	Pacitan Nawangan	41	Trusted	Trusted	Trusted	Trusted
60	Pacitan Pacitan	41	Trusted	Trusted	Doubtful	Trusted
61	Pacitan Sudimoro	41	Trusted	Trusted	Trusted	Trusted
62	Pacitan Tegalombo	40	Trusted	Doubtful	Trusted	Trusted

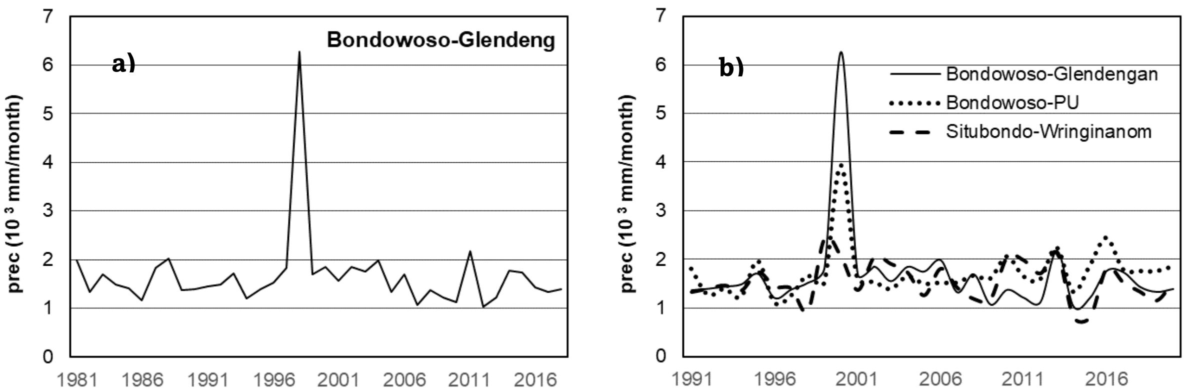

Fig. 9 a) Rainfall pattern of Bondowoso-Glendengan which passed overall testing except von Neumann; **b)** Relative comparison Bondowoso-Glendengan with closest station.

3.4 Homogeneity Status

A summary of the annual and seasonal statistical testing results is presented in Table 4. The analysis covers stations with data spanning over 20 years. Three stations (5%) are rejected due to at least one of the three tests falling into the reject category, indicating failure to meet homogeneity criteria. Pamekasan-Samiran Station stands out as purely rejected, where all three tests indicate rejection. Among the total stations, 11% (7 stations) are marked as 'suspect', requiring further investigation. Another 11% (7 stations) are classified as 'doubtful', which means their data needs a closer look for possible improvements. Most of the stations, 73% or 45 stations, are deemed as 'trusted'.

Fig. 10 compare typical homogeneous and inhomogeneous series rainfall. Fig. 10a) depicts the homogeneous rainfall of Lamongan-Ngimbang, characterized by regular oscillations with no extreme values or abrupt changes. This station ia an excellent example of homogeneous data, making it ideal for assessing the homogeneity of nearby stations. However, in Fig. 10b, the rainfall series from Pamekasan-Samiran appears to be inhomogeneous without doubt. This series displays irregular oscillations with sharp shifts between maximum and minimum values, which are more noticeable in annual data compared to monthly data. Break years, identified in 1997 by Pettitt and 2009 by SNHT and Buishand, exhibit significant differences before and after the change points. The variation between the year of these break points underscores the non-homogeneous nature of the data.

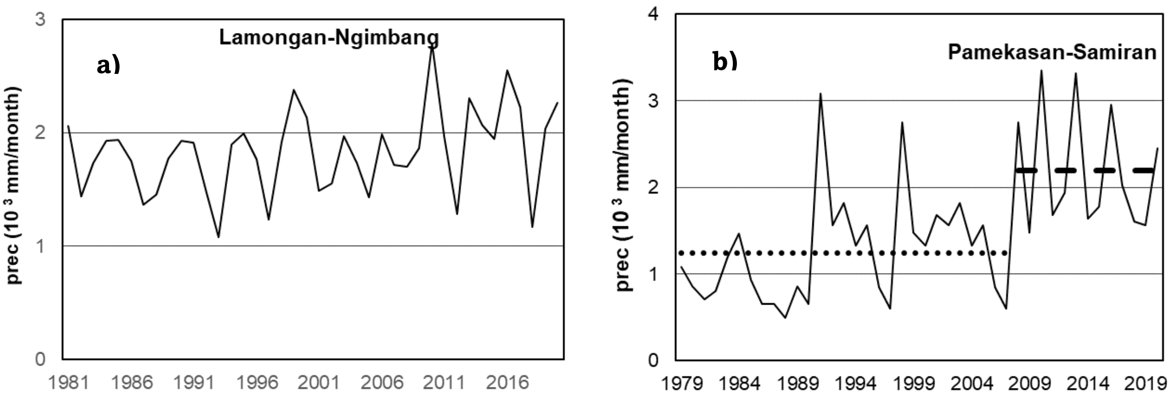

Fig. 10 a) Typical 'trusted' station which homogeneous both annual and seasonal; **b)** typical of 'rejected' statios which inhomogeneous both *annual and seasonal*

3.5 Towards Comprehensive Homogeneity Test

The von Neumann and Buishand tests should be supplemented with Pettitt and SNHT analyses to enhance data assessment. Both parametric and nonparametric should be conducted together to avoid false break detection. In absolute test, SNHT and Pettitt analyses complement each other, mutually confirming breakpoint occurrences. For datasets with high variability, including daily rainfall or monthly rainfall, it may be adequate to use rank based data, like Pettitt's (Aguilar et al. 2003).

Regarding to SNHT and similar methodologies, the test should exclusively be employed on a standardized difference series (Aguilar et al. 2003). Detection of inhomogeneities in beginning or the end should be confirmed by metadata (Toreti et al. 2011). This approach can be implemented using either multiple reference series with high correlation (relative homogeneity) (Aguilar et al. 2003) or singular reference series (Alexandersson and Moberg 1997; Peterson and Easterling 1994; Toreti et al. 2011). One strategy to address this challenge is to utilize series from diverse meteorological networks with similar rainfall characteristics to enhance the likelihood of identifying breaks that affect the entire network.

Both absolute and relative homogeneity tests have disadvantages. Relying solely on data from individual weather stations presents challenges in distinguishing between climatic and non-climatic influences. Peterson et al. (1998) emphasize the necessity of comprehensive metadata support from station history records to accurately assess detected breakpoints (Alexandersson and Moberg 1997; Toreti et al. 2011). Relative methodologies are employed to isolate non-climatic influences, assuming uniform climatic patterns within a geographic area. However, complications arise when climate data series show inhomogeneities due to concurrent changes in the observational network, potentially reducing the sensitivity of relative tests.

As mentioned by scholar, homogeneity testing should consider metadata comparison (Aguilar et al. 2003) as demonstrated in Turkey (Arikan and Kahya 2019). Moreover, homogeneity testing needs to evaluate data trends of reference stations (Ay 2020). An autocorrelation of single series could be useful for breakpoint detection, like Pettitt's, SNHT, Buishand, and von Neumann (Beaulieu et al. 2012). However, with limited information about individual stations, drawing conclusions can be challenging. For results categorized as 'doubtful,' it's essential to identify outliers and anomalous data distributions. The study underscores the significance of similar rainfall zones to streamline the selection of reference stations for homogeneity testing, particularly for stations classified as 'doubtful' or 'suspected'.

6. Conclusion

Among the time series, observations from October to December exhibit better homogeneity compared to the primary rainy season. However, the rainfall datasets during the primary rainy season demonstrate the poorest homogeneity. Through annual and seasonal homogeneity tests conducted on 67 rainfall stations, 5 stations are removed due to insufficient data length, 5% stations are 'rejected', while 11% are categorized as 'suspect' and another 11% as 'doubtful'. The majority, 73%, are deemed 'trusted', totaling 45 stations suitable for metadata comparison. Additionally, 7 stations classified as 'doubtful' can still be considered useful for analysis. However, the remaining 10 stations require careful consideration for future water management. Typical break year of climatic record occurring during extreme El Niño in 1997 or extreme La Niña in 2000. Pettitt's test yields better results for anticipating climatic anomalies. It is crucial to first remove any outliers from the datasets, conduct both parametric and nonparametric tests, and compare with surrounding rainfall stations before homogeneity testing. Comparing with trusted reference stations is beneficial for stations categorized as 'doubtful' or 'suspect'. These findings underscore the importance of assessing reference stations within similar rainfall zones.

Acknowledgments

The author expresses gratitude to Mr. Hizbullah from the Public Works Office for Water Resources in East Java Province, Mr. Beny from the Bengawan Solo Hilir River Basin Authority, and Mr. Harjono from the Faculty of Science and Engineering at Universitas Bojonegoro for their assistance in data collection and provision.

Author Contributions: H.M.: Conceptualization, methodology, formal analysis, writing—original draft, I.I.; supervision, resources, project administration, R.G.; supervision, project administration, validation. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest: The authors declare no conflict of interest.

References (min 40 references)

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., & Wieringa, J. (2003). *Guidance on Metadata and Homogenization*. Geneva, Switzerland: World Meteorological Organization.

Akinsanola, A. A., & Ogunjobi, K. O. (2017). Recent homogeneity analysis and long-term spatio-temporal rainfall trends in Nigeria. *Theoretical and Applied Climatology*, *128*(1), 275–289. https://doi.org/10.1007/s00704-015-1701-x

Aldrian, E., & Djamil, Y. S. (2008). Spatio-temporal climatic change of rainfall in East Java Indonesia. *International Journal of Climatology*, *28*, 435–448. https://doi.org/10.1002/joc.1543

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, *6*(6), 661–675. https://doi.org/10.1002/joc.3370060607

Alexandersson, H., & Moberg, A. (1997). Homogenization of Swedish Temperature Data. Part I: Homogeneity Test for Linear Trends. *International Journal of Climatology*, *17*(1), 25–34. https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J

Anripa, N., Kumar, A., Maharana, P., & Dimri, A. P. (2023). Climate change over Indonesia and its impact on nutmeg production: An analysis under high-resolution CORDEX-CORE regional simulation framework. *International Journal of Climatology*, *43*(10), 4472–4490. https://doi.org/10.1002/joc.8098

Arikan, B. B., & Kahya, E. (2019). Homogeneity revisited: analysis of updated precipitation series in Turkey. *Theoretical and Applied Climatology*, *135*(1), 211–220. https://doi.org/10.1007/s00704-018-2368-x

As-syakur, Abd. R., Osawa, T., Miura, F., Nuarsa, I. W., Ekayanti, N. W., Dharma, I. G. B. S., et al. (2016). Maritime Continent Rainfall Variability during the TRMM Era: The Role of Monsoon, Topography and El Nino Modoki. *Dynamics of Atmospheres and Oceans*. https://doi.org/10.1016/j.dynatmoce.2016.05.004

Ay, M. (2020). Trend and homogeneity analysis in temperature and rainfall series in western Black Sea region, Turkey. *Theoretical and Applied Climatology*, *139*(3), 837–848. https://doi.org/10.1007/s00704-019-03066-6

Beaulieu, C., Chen, J., & Sarmiento, J. L. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *370*(1962), 1228–1249. https://doi.org/10.1098/rsta.2011.0383

Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, *58*(1), 11–27. https://doi.org/10.1016/0022-1694(82)90066-X

Chen, C., Sahany, S., Moise, A. F., Chua, X. R., Hassim, M. E., Lim, G., & Prasanna, V. (2023). ENSO–Rainfall Teleconnection over the Maritime Continent Enhances and Shifts Eastward under Warming. *Journal of Climate*, *36*(14), 4635–4663. https://doi.org/10.1175/JCLI-D-23-0036.1

D'Arrigo, R., Wilson, R., Palmer, J., Krusic, P., Curtis, A., Sakulich, J., et al. (2006). Monsoon drought over Java, Indonesia, during the past two centuries. *Geophysical Research Letters*, *33*(4), L04709. https://doi.org/10.1029/2005GL025465

Gu, L., Chen, J., Yin, J., Xu, C., & Zhou, J. (2020). Responses of Precipitation and Runoff to Climate Warming and Implications for Future Drought Changes in China. *Earth's Future*, *8*(10), e2020EF001718. https://doi.org/10.1029/2020EF001718

Hadi, S. J., Hadi, A. J., Ismail, K. S., & Tombul, M. (2019). Homogeneity and Trend Analysis of Rainfall and Streamflow of Seyhan Basin (Turkey). In H. I. Chaminé, M. Barbieri, O. Kisi, M. Chen, & B. J. Merkel (Eds.), *Advances in Sustainable and Environmental Hydrology, Hydrogeology, Hydrochemistry and Water Resources* (pp. 17–19). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-01572-5_4

Hawkins, D. M. (1977). Testing a Sequence of Observations for a Shift in Location. *Journal of the American Statistical Association*. https://www.tandfonline.com/doi/abs/10.1080/01621459.1977.10479935. Accessed 6 May 2024

Intergovernmental Panel On Climate Change (IPCC). (2023). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Jarušková, D. (1996). Change-Point Detection in Meteorological Measurement. *Monthly Weather Review*, *124*(7), 1535–1543. https://doi.org/10.1175/1520-0493(1996)124<1535:CPDIMM>2.0.CO;2

Krzywinski, M., & Altman, N. (2014). Nonparametric tests. *Nature Methods*, *11*(5), 467–468. https://doi.org/10.1038/nmeth.2937

Mulyanti, H., Istadi, I., & Gernowo. (2023). Assessing Vulnerability of Agriculture to Drought in East Java, Indonesia: Application of GIS and AHP | Mulyanti |. *Geoplanning: Journal of Geomatics and Planning*, *10*(1), 55–72.

Pandžić, K., Kobold, M., Oskoruš, D., Biondić, B., Biondić, R., Bonacci, O., et al. (2020). Standard normal homogeneity test as a tool to detect change points in climate-related river discharge variation: case study of the Kupa River Basin. *Hydrological Sciences Journal*, *65*(2), 227–241. https://doi.org/10.1080/02626667.2019.1686507

Peterson, T. C., & Easterling, D. R. (1994). Creation of homogeneous composite climatological reference series. *International Journal of Climatology*, *14*(6), 671–679. https://doi.org/10.1002/joc.3370140606

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., et al. (1998). Homogeneity adjustments ofin situ atmospheric climate data: a review. *International Journal of Climatology*, *18*(13), 1493–1517. https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T

Pettitt, A. N. (1979). A non-parametric approach to the change point problem. *Journal of Applied Statistics*, *28*, 126–135.

Qian, J.-H., Robertson, A. W., & Moron, V. (2010). Interactions among ENSO, the Monsoon, and Diurnal Cycle in Rainfall Variability over Java, Indonesia. *Journal of the Atmospheric Sciences*, *67*(11), 3509–3524. https://doi.org/10.1175/2010JAS3348.1

Rummukainen, M. (1997). *Methods for statistical downscaling of GCM simulations*. SMHI. https://urn.kb.se/resolve?urn=urn:nbn:se:smhi:diva-2602. Accessed 7 May 2024

Rummukainen, M. (2016). Added value in regional climate modeling. *WIREs Climate Change*, *7*(1), 145–159. https://doi.org/10.1002/wcc.378

Tanaka, M. (1994). The Onset and Retreat Dates of the Austral Summer Monsoon over Indonesia, Australia and New Guinea. *Journal of the Meteorological Society of Japan*, *72*(2), 255–267.

Toreti, A., Kuglitsch, F. G., Xoplaki, E., Della-Marta, P. M., 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. *International Journal of Climatology*, *31*(4), 630–632. https://doi.org/10.1002/joc.2088

Tsybakov, A. B. (2009). Nonparametric estimators. In *Introduction to Nonparametric Estimation* (pp. 1–76). Springer. https://link.springer.com/book/10.1007/b13794. Accessed 9 May 2024

Wahab, M. I., Sudibyakto, S., Gunadi, S., & Suratman, W. S. (2009). Change in Spatial and Temporal Characteristics of Rainfall in East Java Province in Relation to Global Climate Change Perubahan Karakteristik Curah Hujan Menurut Ruang Dan Waktu Di Provinsi Jawa Timur Dalam Kaitannya Dengan Perubahan Iklim ... *Agromet*, *23*(1), 245931.

Whitley, E., & Ball, J. (2002). Statistics review 6: Nonparametric methods. *Critical Care*, *6*(6), 509. https://doi.org/10.1186/cc1820

Wijngaard, J. B., Klein Tank, A. M. G., & Können, G. P. (2003). Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology*, *23*(6), 679–692. https://doi.org/10.1002/joc.906