Journal of the Civil Engineering Forum, May 2025, 11(2):147-154
DOI 10.22146/jcef.12720
Available Online at https://jurnal.ugm.ac.id/v3/jcef/issue/archive

Identifying Prolonged Zero Value Periods as Part of Quality Control on Daily Rainfall Records in East Java, Indonesia

Dinda Rosyia Wibawanty[1,3], I Putu Santikayasa[1,*], Supari[2]
[1]Department of Meteorology and Geophysics, FMIPA Building, IPB Dramaga Campus, Bogor, INDONESIA
[2]Agency of Meteorology Climatology and Geophysics, Jakarta, INDONESIA
[3]South Sumatra Climatology Station, Palembang,INDONESIA
*Corresponding author: ipsantika@apps.ipb.ac.id

SUBMITTED 27 March 2024 **REVISED** 23 July 2024 **ACCEPTED** 06 August 2024

ABSTRACT Ensuring the quality of surface rainfall records is crucial for obtaining highly representative data and facilitating further comprehensive analysis. Given that surface rainfall observations are predominantly conducted using conventional gauges, they are still susceptible to human errors that can significantly impact data quality. Among various types of errors that may arise, the issue of zero rainfall records is relatively overlooked. Prolonged zero rainfall periods may introduce uncertainty, as mistyped missing data can be erroneously replaced with zero values. The challenge in handling this issue is complicated by the absence of sufficient evidence to conclusively determine the validity or suspicion of consecutive zero rainfall periods. Therefore, we implemented the Affinity Index, altitude difference, and maximum distance approaches to detect and evaluate (validate or reject) any potential invalid sequences of prolonged zero values in the rainfall dataset. The Affinity Index quantifies the agreement of rain and non-rain events between two meteorological stations, functioning as a metric to evaluate the similarity of their rainfall patterns. Utilizing daily data from 682 rain gauge stations in East Java, Indonesia, spanning from January 2010 to December 2019, we identified two major concerns: zero rainfall accumulation during the peak of the rainy season (December/January/February) and extended dry spells lasting more than 180 days. To address the first issue, we flagged the corresponding station and excluded it from the dataset. For the second issue, we established reference stations for each target station to enable meaningful comparisons. The study found that 8.8% of stations detected zero rainfall accumulation during the peak of the rainy season. Regarding prolonged dry spells, we successfully assessed 98% of extended dry spell events in East Java. The majority of these events were considered valid, while around 3% were deemed dubious.

KEYWORDS Quality control; Daily; Rainfall; Dry spell; Reference stations.

1 INTRODUCTION

Serving as ground truth, near-surface precipitation observation plays an essential role in providing the most representative rainfall data. Therefore, evaluations of other precipitation data products (such as remote sensing and reanalysis) and assessments of numerical models rely on ground-based data (Feng et al., 2004). However, near-surface precipitation data is still susceptible to human errors, particularly when obtained from conventional gauges. Common types of issues found in the records include atypical values, outliers, missing data, duplicated records, and inhomogeneities (Durre et al., 2010; Vicente-Serrano, S.M. et al., 2010; Terán-Chaves, 2021). The source of errors can stem from various aspects of the data life cycle, including instrumental issues, observer errors, data transmission problems, key entry mistakes, and errors in the data validation process. Additionally, challenges arise from changing data formats and data summarization (WMO, 2018). All these errors present in the data have a substantial impact on data quality, potentially affecting the precision of subsequent analysis (Scherrer, 2011).

As the importance of data quality gains widespread acknowledgement, there has been a growing number of studies focusing on precipitation data quality control. These studies have contributed to significant advancements in quality control methods (Durre et al., 2010; Vicente-Serrano, S.M. et al., 2010; Hamada et al., 2011; Llabrés-Brustenga, 2019; Terán-Chaves, 2021), there by enhancing the reliability and accuracy of precipitation measurements. Despite these advancements, certain aspects of data errors, particularly those related to zero rainfall records, remain challenging. Consecutive-prolonged zero rainfall periods may introduce uncertainty, as mistyped missing data can be erroneously replaced with zero values, subsequently impacting climate analysis, particularly concerning extreme indices such as consecutive dry days (CDD) and consecutive wet days (CWD) (Hunziker, 2017). Unlike non-zero repetition, identifying zero repetition poses a substantial challenge. The issue arises from insufficient evidence to judge prolonged zero rainfall as incorrect without other independent measurements, such as known rainfall records or historic flood/drought events (Hamada

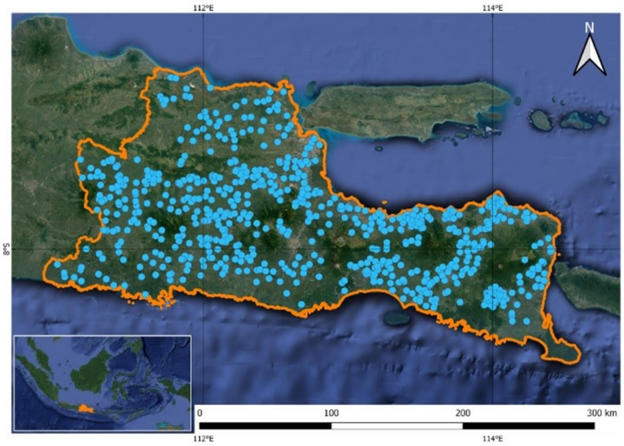


Figure 1 East Java as the study area, blue points indicate the locations of rain gauges

et al., 2011; Lewis, 2021). On the other hand, the availability of auxiliary data may be limited to some extent, adding to the distinct challenges that necessitate further exploration.

Fundamentally, there are two main approaches to identifying potential errors in precipitation data: the first involves data from a single station, while the second employs data from multiple stations, allowing comparison with neighboring stations (Estévez et al., 2022). Nevertheless, the first method is considered insensitive since the intermittent nature and high variability of daily rainfall do not allow reliable confidence intervals to be estimated from the historical data of the station under consideration (Sciuto, 2009). Moreover, the first method requires a long period of data to cover the highest temporal variability of rainfall ever recorded. Therefore, our focus will be on the second method, which involves a comparison between the target station and its corresponding neighboring stations as references. Many studies have demonstrated the criteria for defining neighboring stations based on parameters that gauge the similarity between the target and reference stations. Blenkinsop et al. (2017) employed Percentage Correct Statistics, later known as the Affinity Index (AI) (Lewis, 2021), as a measure of matching statistics, illustrating the concordance of rain and non-rain events between target and reference stations. Besides AI, numerous studies use parameters such as maximum distance range, height difference, and correlation to capture the representative closure of rainfall characteristics between stations. However, the use of correlation may be less useful due to the high variability in daily rainfall data, including the presence of zero values (Sciuto, 2009).

Regarding the state-of-the-art as mentioned above related to precipitation data quality control, this study aims to identify issues related to zero rainfall values in East Java, Indonesia and proposes methods to evaluate (validate or reject) any potentially invalid sequences of prolonged zero rainfall values based on the neighboring stations approach. Through this investigation, we seek to contribute to the refinement of precipitation data quality control methods and enhance our understanding of the challenges associated with zero rainfall records.

2 METHODS

2.1 Study Area

This study focuses on East Java, Indonesia, utilizing daily rainfall data from 682 rain gauges maintained by the National Agency for Meteorology, Climatology and Geophysics (BMKG). The dataset covers 10 years, starting from 1 January 2010 to 31 December 2019, with less than 10% of missing data annually. However, the amount of missing data may increase as a result of the quality control measures proposed in this study. The observation network consists of standard Observatorium rain gauges, which collect rainfall over specific intervals and are meticulously measured by trained observers employing calibrated measuring instruments. The spatial distribution of these gauges, depicted in Figure 1, reveals a relatively sparse network, with an average density of 1 gauge per 62 square kilometers and an average inter-station distance of 129.5 km. East Java's topography is defined by a mountainous terrain extending east-west, while its northern region is predominantly lowlands (Supari, 2012). Being part of Java Island, East Java's climate is principally shaped by the wet northwest monsoon from November to March (NDJFM) and the dry southeast monsoon from May to September (MJJAS) (Aldrian and Susanto, 2003).

2.2 Assessment of Data Quality

Conducting a preliminary data quality analysis offers an initial insight into the condition of the dataset. We utilize the Quality Index (Q) proposed by Llabrés-Brustenga (2019) and select parameters most relevant to our research. The Q index employed here represents the completeness of the data, the distribution of gaps within the dataset, and the proportion of monthly nonzero precipitation accumulation. We opted not to use the outlier and weekly data cycle indicators mentioned in Llabrés-Brustenga (2019), as our research focuses more on identifying dubious prolonged dry spell periods. Calculated using Equations 1 to 3 for each station annually, the Q index produces values ranging from 0 to 100. Values greater than 80 indicate acceptable quality, while those below 50 suggest very low quality.

$$Q = \frac{1}{3}\left(P + Q_{gaps+\mathrm{Q}_0^m}\right) \tag{1}$$

$$Q_{gaps} = 100 - 100\frac{2 * n_{gap} + \mathrm{L}_{gap}^{max}}{n} \tag{2}$$

$$\mathrm{Q}_0^m = 100 - 100\left(\frac{m_0}{m}\right) \tag{3}$$

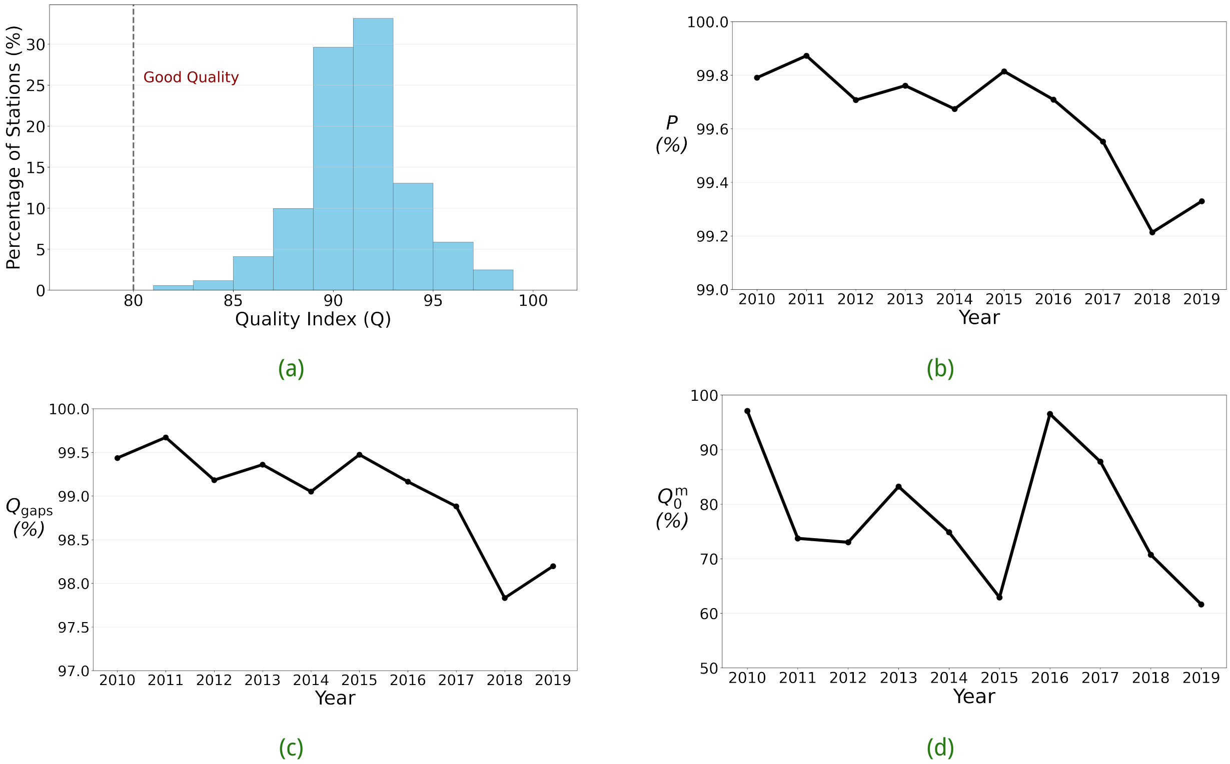


Figure 2 Data Quality Index: a) Q Index of stations in East Java; b) Average data availability (P) across all stations in East Java; c) Average data gaps (Q_{gaps}) across all stations in East Java; d) Average non-zero months (Q_0^m) across all stations in East Java

In Eq.1-3, which focuses on a yearly period despite the data sometimes being identified in days or months, the following terms are defined: Q represents the data quality index assigned to each station for every year of the samples (%); P indicates the percentage of annual data available (%); Q_{gaps} reflects the yearly distribution of data gaps (%); n_{gap} denotes the total number of empty days in each year within the dataset (days), where factor 2 in the equation accounts for the weighted impact of the number of empty days; L_{gap}^{max} represents the maximum consecutive spell of empty days in a yearly period (days); n indicates the number of days within a year (days); Q_0^m represents the percentage of months in a year with non-zero rainfall accumulation (%); m_0 indicates the number of complete months in a year with zero monthly precipitation accumulation and m denotes the number of months in the year for the given series.

2.3 Quality Control of Prolonged Zero Rainfall Records

The methodology presented here aims to assess prolonged, consecutive zero rainfall records by comparing sequences from the target station to those from reference stations. The prolonged dry spells that will be inspected further are those equal to or extending 180 days. This threshold is attributed to the monsoon activity that typically changes direction and affects rainfall characteristics over six months (Mulsandi, 2024). Each target station can have up to 10 reference stations, with a maximum distance threshold of 50 km from the target station (Lewis et al., 2021). However, in the study region, the actual maximum distance is found to be 35 km. The maximum altitude difference is set to ±200 m, taking into account the orography of the study region (Sciuto, 2009). The AI index, set at 0.8 or higher, is used to characterize the daily rainfall patterns between the target and reference stations. It serves as a matching statistic for non-rain (dry) and rain (wet) conditions between these stations (Eq. 4-6) (Blenkinsop et al., 2017; Lewis, 2018, 2021). The use of AI is based on the consideration that other alternative criteria, such as correlation, may not be representative due to the existence of zero values in the data. In order to proceed with the AI, we should first convert daily rainfall data into binary form: zero if daily rainfall is less than 1 mm and one if it is 1 mm or more.

$$AI = I_1 + I_2 \tag{4}$$

$$I_1 = F_r\left[Y_t = 0, \mathrm{Y}_t^r = 0\right] \tag{5}$$

$$I_2 = F_r\left[Y_t = 1, \mathrm{Y}_t^r = 1\right] \tag{6}$$

Referring to Eq.4 through Eq.6, F_r represents the frequency of events; Y_t represents the dry (0) or wet (1) condition of the target station, while Y_t^r represents the same condition in reference stations.

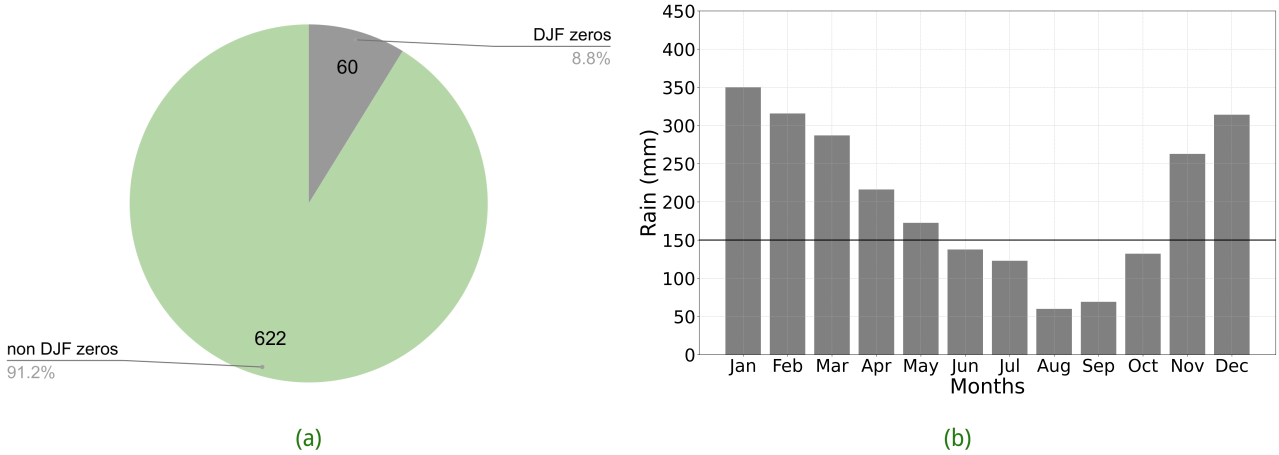


Figure 3 a) Proportion of stations detecting zero monthly precipitation accumulation during DJF; b) General pattern of rainfall in East Java, identified as monsoonal, typically peaks during December to February (DJF) (BMKG, 2022*b*)

The neighboring check is not performed if the final analysis yields fewer than three reference stations. To minimize the effect of the seasonal cycle, we divide our reference station detection into wet (NDJFMA) and dry (MJJASO) season periods. For the flagging criteria, we adopt the procedure outlined by Lewis (2021), employing a 15-day moving window. A sequence of 15 days of consecutive zero periods is then considered dubious if the average occurrence of wet days from all reference stations is equal to or greater than three days. Lewis (2018) identified this threshold as Type E events, where likely false dry events are detected during the validation of a dry spell rule base with a 20% threshold. Thus, a 15-day period of dry days is likely to be categorized as false if at least 20% (three days) of the reference sequence is identified as wet days.

3 RESULTS

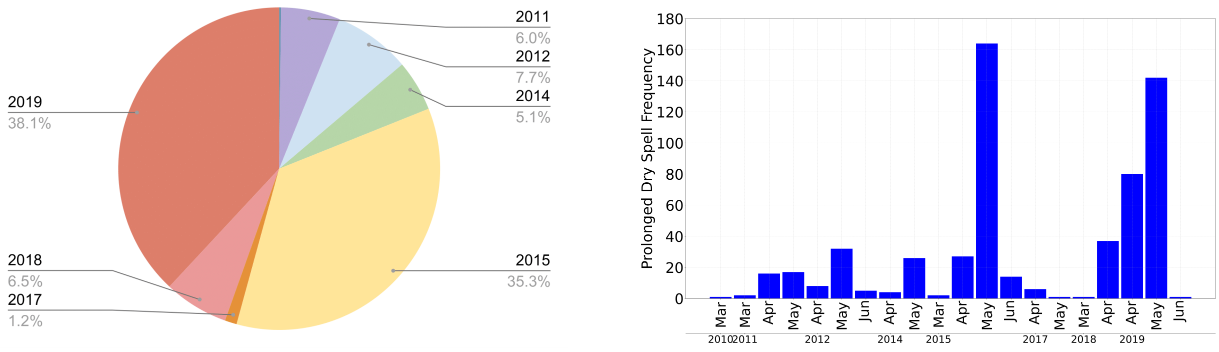
3.1 Data Quality Index

The 10-year average of the data quality index across every station in East Java surpasses the threshold for good quality, as shown in Figure 2a. More than 80% of the stations have a quality index exceeding 90%. The quality index ranges from 83.3% to 99.2%, with a mean across all stations reaching 92.3%. These results suggest that the initial data quality can be considered high in terms of data completeness, gap distribution, and the proportion of non-zero accumulated precipitation.

We also analyze each parameter that defines the quality index, including data completeness (P), gap distribution (Q_{gaps}), and non-zero monthly precipitation accumulation (Q_0^m), as depicted in Figure 2b, 2c, and 2d, respectively. The P index shows a tight range between 99.21% and 99.87%, indicating nonsignificant temporal fluctuations in data completeness. This near-complete data availability, as reflected by the P index, is followed by a low gap distribution of missing data. As shown in Figure 2c, Q_{gaps} range from a minimum value of around 97.83% to a maximum value of around 99.67%, suggesting minimal gaps with short durations present in the dataset. However, the Q_0^m parameter exhibits a broader range of values, from around 60% to above 90%. Six out of 10 years oscillate below the Q_0^m value of 80%, with the lowest values occurring in 2015 and 2019. This outcome indicates a significant amount of monthly rainfall accumulation equaling zero exists within the dataset. At least two possible factors may trigger this condition: first, it may be attributed to natural climate phenomena such as El Niño and the Indian Ocean Dipole (IOD) events that occurred in 2015 and 2019, and second, some of the data may contain human error, such as mistyping between missing data and zero values. However, further investigation is required to get to the closest justification.

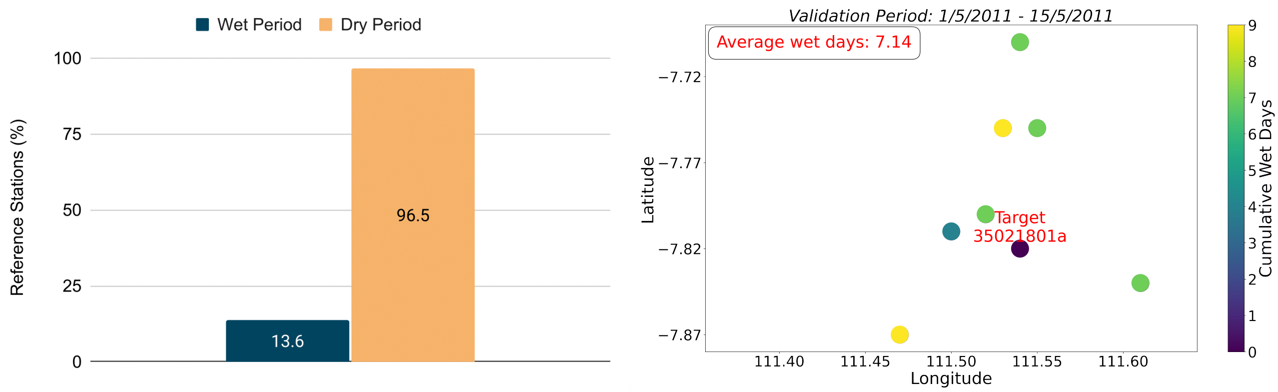
3.2 Zero Rainfall Periods During Rainy Season Peaks

Identifying faulty zeros in the dataset presents a challenging issue, given the high variability of rainfall characteristics across both temporal and spatial scales. Despite this complexity, certain climate types exhibit distinct patterns in rainfall occurrence over specific periods. In parts of Indonesia, for instance, such patterns manifest as distinct peaks and troughs in rainfall accumulation, attributed to the activity of the northwest and southeast monsoons. East Java is one such region experiencing this monsoonal pattern, as illustrated in Figure 3b, with rainfall typically peaking during the DJF period (BMKG, 2022*a*). Based on the expected rainfall pattern, any month of December, January, or February experiencing zero monthly rainfall accumulation would be considered a data error. According to Figure 3a, 8.8% of stations in East Java recorded no precipitation throughout the entire months of December, January, or February. As it appears to contradict the



(a) (b)

Figure 4 a) Proportion of prolonged dry spells period over the years; b) Distribution of prolonged dry spells based on starting month



(a) (b)

Figure 5 a) Proportion of detected reference stations by wet and dry period; b) Flagging criteria illustration for period 1 May 2011 until 15 May 2011

climatic characteristic of the area, we flagged and excluded the corresponding stations with zero rainfall accumulations found in the DJF period.

3.3 Prolonged Zero Rainfall Periods

After detecting more readily identifiable errors, we proceed to address the more complex cases related to prolonged dry spell periods. Employing a threshold of 180 consecutive days of zeros, we filter the periods that require further analysis. Our analysis reveals that half of the stations in the dataset experienced at least one occurrence of prolonged dry spell periods from 2010 to 2019. Figure 4a shows the distribution of 55.47% of stations detecting extended zero periods over the years. Notably, the highest occurrence was observed in 2015 and 2019, while no event was detected in 2013 and 2016. The onset of these prolonged dry spell periods varied between March, April, May, and June, as illustrated in Figure4b, with the majority starting in March and April. This result aligns with the previously analyzed Q_0^m that showed the lowest values in 2015 and 2019, indicating the highest occurrence of consecutive zero rainfall.

We then assessed the dubious data containing extended dry spells. Initially, we identified up to ten nearest stations for each target station within a 50 km radius, followed by filtering the reference station candidates using the Affinity Index and altitude difference. This process was conducted separately for wet (NDJFMA) and dry (MJJASO) periods. The final results yielded pairs of target-reference stations for up to 96.5% of all dubious stations. As depicted in Figure 5a, the dry period produced more pairs of target-reference stations than the wet period. Subsequently, after obtaining the reference stations, we proceeded to identify prolonged consecutive zero periods and applied the flagging criteria. The flagging criteria are illustrated in Figure 5b, where 15 days of zero rainfall is considered dubious if the average occurrence of rainy events across all reference stations reaches a minimum of three days.

Overall, the methodology presented here successfully evaluates 98% of events involving dubious prolonged consecutive zero periods, while 2% of the events do not meet the criteria of the reference stations and, therefore, cannot be evaluated. The majority of the evaluated events are considered valid based on the reference stations' approach, while 3.9% were deemed dubious.

Table 1. Statistics of annual station flags

Years	Dubious Stations	Flagged Stations	Percentage (%)
2010	1	1	100
2011	35	5	14.29
2012	45	5	11.11
2014	30	5	16.67
2015	207	7	3.38
2017	7	3	42.86
2018	38	5	13.16
2019	223	3	1.35

The yearly statistics regarding station flags are outlined in Table 1. Remarkably, the highest flag occurrences were recorded in 2010 and 2017, reaching 100% and 42%, respectively. These peaks were primarily driven by a small initial number of stations experiencing prolonged dry spells, amounting to just one in 2010 and seven in 2017. Across the remaining stations, flag percentages ranged from 1% to 16%. Interestingly, in 2015 and 2019, we observed the highest number of stations detecting prolonged consecutive zero periods, yet the flag percentages remained comparable to other years. This result suggests the potential influence of climate phenomena, such as El Niño and IOD, in amplifying dry spell durations.

4 DISCUSSION

This research presents an analysis concerning zero rainfall records that may raise doubts regarding their occurrence and duration. This assessment is crucial because zero values are part of natural rainfall fluctuations, often less conspicuous compared to upper precipitation limits. Such oversight could erroneously suggest a gauge's suitability for analysis over a given period when, in fact, this might not be the case (Blenkinsop et al., 2017). The initial detection of zero rainfall issues stems from temporal variations in the average Q_0^m, which exhibit a higher range compared to other indices. Yearly variations in Q_0^m also consist of stations with values less than or equal to 50%, indicating low data quality. This finding related to Q_0^m is not revealed in other similar studies implementing *Q* index; others are more focused on identifying the global *Q* index and the rest are focused on different parameters, such as *P*, Q_{gaps} , and outliers (Llabrés-Brustenga, 2019; Estévez et al., 2022).

Building upon the insights obtained from the data quality index, our approach begins by addressing the more pronounced error: zero rainfall occurrences during the DJF period, which aligns with the peak of the rainy season. This step of quality control is based on known climatological characteristics of rainfall in the study area. Eliminating this issue first from the dataset helps to reduce the uncertainty in evaluating the more dubious cases involving prolonged dry spell periods. This hierarchical strategy, moving from simpler checks to more intricate analyses, mirrors methodologies employed in prior research to systematically enhance data quality (Einfalt and Michaelides, 2008). To evaluate these extended dry spell periods, we adopt an approach that certain forms of data issues, such as data input errors or instrument failures, are unlikely to be replicated across gauge networks and possibly to be identified through comparison with neighboring gauges (Lewis, 2018). Additionally, we operate under the assumption that events of such prolonged duration will manifest on a broader spatial scale, enabling neighboring station comparison. This stands in contrast to extreme rainfall, which typically occurs over shorter durations and tends to be more localized.

The total count of data flagged as dubious due to prolonged dry spell periods amounts to 802 days across 31 different stations. The distribution of flagged days at each station ranges from 0.38% to 2.03%, with the average across all stations being 0.71%. The majority of the stations, around 80.65%, had flagging rates of less than 1%. This finding is consistent with Lewis (2021), who conducted quality control measures to address various gauge malfunctions and recording errors, including those associated with consecutive zero rainfall periods. Their study revealed that the majority of stations removed a small proportion of data (less than 5%), where zero rainfall QC contributed the highest to the data removal—a study by Vicente-Serrano, S.M. et al. (2010) resulted in a similar flagging rate, with the maximum reaching 1.04%. From that flagged data, a significant proportion was also related to zero rainfall values. The concern regarding extended dry spell periods also emerged as the primary issue in the study focusing on the UK, despite the presence of other issues such as large rainfall values, accumulated values, and consecutive large values (Lewis, 2018). Another study conducted in Brazil also identified issues related to consecutive rainless periods exceeding 200 days within a yearly cycle (Meira, 2022). The findings indicated that the number of affected stations ranged from 1 to 85 during the 2014-2020 period, with the highest incidence recorded in 2019. However, they did not investigate these cases further and directly labeled the stations as poor quality.

Another noteworthy finding from our research is the temporal variability observed in prolonged dry spell periods, which increased in accordance with the active phase of El Niño. The El Niño effect is prominent in Java, showing a tendency towards drier conditions than normal. Previous research has found that 79 out of 97 stations during SON in El Niño years experienced significantly dry anomalies with magnitudes of CDD greater than 40% relative to neutral years (Supari, 2017). The effect of El Niño is also evident in the longest dry spell duration ever recorded between 1991-2020, reaching 298 days in East Java due to the 2009 El Niño event (BMKG, 2022*b*). Additionally, the simultaneous occurrence of El Niño and positive IOD (pIOD) has the main contribution to drought severity in Java (Siswanto, 2022). In this study, the occurrence of prolonged dry spells is dominated by 2015 and 2019, during which a strong El Niño and the combination of El Niño and IOD were present. This finding suggests that, despite the errors detected within the prolonged period, natural variability also plays a role in these extended dry spell events.

In general, the quality control procedure introduced here aims to shed light on issues related to extended zero rainfall values, particularly in the study area where this concern has seemingly been overlooked. In proposing this method, we evaluate parameters and thresholds defined by previous studies to determine terms that can be effectively implemented in the study area. The challenge here is particularly related to the sparse distribution of rain gauge data. Thus, implementing stricter, more robust thresholds will result in more unvalidated stations. For instance, while other studies have utilized Pearson and Spearman correlations, these parameters seem unsuitable when considering the widely used threshold of 0.7. Therefore, we opt for the Affinity Index instead. The parameter used here represents a good level of applicability without excessively flagging data, addressing the challenge related to sparse station distributions (De Vos et al., 2019).

Regarding the performance of the quality control method, it is imperative to detect errors accurately while minimizing false alarms. Consequently, many studies integrate auxiliary data to validate their methods. In the case where this data is not available, some research deliberately introduces known errors into the rainfall dataset, such as using multiplicative errors, seeding random numbers, and replacing observed values with zero values (You et al., 2007; Sciuto, 2009; Terán-Chaves, 2021). Nonetheless, these approaches introduce uncertainties regarding the appropriate amplitude for seeding errors among realistic precipitation patterns and the suitable level at which errors should be detected (Scherrer, 2011). In our research, we have not conducted comparisons to other datasets, which represents a potential area for further research. However, it's important to note that datasets can never be entirely free from errors (Llabrés-Brustenga, 2019), such as systematic errors due to the acceleration effects of winds and evaporation-induced precipitation loss (Sevruk, 1996; Fankhauser, 1998; Lewis, 2018). Furthermore, rainfall variability itself plays a crucial role in the complexity of quality control procedures, especially when dealing with daily data.

5 CONCLUSION

In conclusion, our study identified two major issues related to zero rainfall values based on their timing and duration: zero rainfall accumulation during the peak of the wet season (DJF) and prolonged dry spell periods extending 180 days. Approximately 8.8% of stations were flagged from the dataset due to zero monthly rainfall accumulation during the DJF period, while a further 55.47% of stations underwent inspection using neighboring stations to assess spatial consistency in detecting prolonged dry spell periods. This involved the utilization of three parameters, including the Affinity Index, maximum distance, and maximum height difference, to select reference stations for each target station. Ultimately, the final assessment validated 96.1% of extended dry spell periods and rejected 3.9% of the remaining cases. The approach employed in our study represents a significant step toward improving the quality of precipitation data, particularly regarding zero rainfall issues. Identifying such issues is crucial for daily datasets to provide a better understanding of climate phenomena and extreme conditions such as Consecutive Dry Days (CDD).

DISCLAIMER

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors thank the National Agency for Meteorology Climatology and Geophysics (BMKG) for providing the rainfall dataset. This research was supported by the Indonesia Endowment Fund for Education Agency (LPDP) as part of the given Master's Degree Scholarship Program.

REFERENCES

Aldrian, E. and Susanto, R. (2003), 'Identification of three dominant rainfall regions within indonesia and their relationship to sea surface temperature', *International Journal of Climatology* **23**(12), 1435–1452.

Blenkinsop, S., Lewis, E., Chan, S. and Fowler, H. (2017), 'Quality control of an hourly rainfall dataset and climatology of extremes for the uk', *International Journal of Climatology* **37**(2), 722–740.

BMKG (2022*a*), *Pemutakhiran Zona Musim Indonesia*, BMKG Press, Jakarta, Indonesia.

BMKG (2022*b*), *Peta curah hujan ekstrem Indonesia periode 1991-2020*, BMKG Press, Jakarta, Indonesia.

De Vos, L., Leijnse, H., Overeem, A. and Uijlenhoet, R. (2019), 'Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring', *Geophysical Research Letters* **46**(15), 8820–8829.

Durre, I., Menne, M., Gleason, B., Houston, T. and Vose, R. (2010), 'Comprehensive automated quality assurance of daily surface observations', *Journal of Applied Meteorology and Climatology* **49**(8), 1615–1633.

Einfalt, T. and Michaelides, S. (2008), Quality control of precipitation data, *in* S. Michaelides, ed., 'Precipitation: Advances in Measurement, Estimation and Prediction', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 101–126.

Estévez, J., Llabrés-Brustenga, A., Casas-Castillo, M., García-Marín, A., Kirchner, R. and Rodríguez-Solà, R. (2022), 'A quality control procedure for long-term series of daily precipitation data in a semi-arid environment', *Theoretical and Applied Climatology* **149**(3–4), 1029–1041.

Fankhauser, R. (1998), 'Influence of systematic errors from tipping bucket rain gauges on recorded rainfall data', *Water Science and Technology* **37**(11).

Feng, S., Hu, Q. and Qian, W. (2004), 'Quality control of daily meteorological data in china, 1951–2000: a new dataset', *International Journal of Climatology* **24**, 853–870.

Hamada, A., Arakawa, O. and Yatagai, A. (2011), 'An automated quality control method for daily rain-gauge data', *Global Environmental Research* **15**, 183–192.

Hunziker, S. e. a. (2017), 'Identifying, attributing, and overcoming common data quality issues of manned station observations', *International Journal of Climatology* **37**(11), 4131–4145.

Lewis, E. e. a. (2018), 'A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for great britain: Cehgear1hr', *Journal of Hydrology* **564**, 930–943.

Lewis, E. e. a. (2021), 'Quality control of a global hourly rainfall dataset', *Environmental Modelling Software* **144**, 105169.

Llabrés-Brustenga, A. e. a. (2019), 'Quality control process of the daily rainfall series available in catalonia from 1855 to the present', *Theoretical and Applied Climatology* **137**(3–4), 2715–2729.

Meira, M. e. a. (2022), 'Quality control procedures for sub-hourly rainfall data: An investigation in different spatio-temporal scales in brazil', *Journal of Hydrology* **613**, 128358.

Mulsandi, A. e. a. (2024), 'Detecting indonesian monsoon signals and related features using space–time singular value decomposition (svd)', *Atmosphere* **15**(2), 187.

Scherrer, S. e. a. (2011), 'Operational quality control of daily precipitation using spatio-climatological plausibility testing', *Meteorologische Zeitschrift* **20**(4), 397–407.

Sciuto, G. e. a. (2009), 'Quality control of daily rainfall data with neural networks', *Journal of Hydrology* **364**(1–2), 13–22.

Sevruk, B. (1996), 'Adjustment of tipping-bucket precipitation gauge measurements', *Atmospheric Research* **42**(1–4), 237–246.

Siswanto, S. e. a. (2022), 'Satellite-based meteorological drought indicator to support food security in java island', *PLOS ONE* **17**(6), e0260982.

Supari, S. e. a. (2012), 'Spatiotemporal characteristics of extreme rainfall events over java island, indonesia', *The Indonesian Journal of Geography* **44**, 62–86.

Supari, S. e. a. (2017), 'Enso modulation of seasonal rainfall and extremes in indonesia', *Climate Dynamics* **51**(7–8), 2559–2580.

Terán-Chaves, C. e. a. (2021), 'Quality control and filling of daily temperature and precipitation time series in colombia', *Meteorologische Zeitschrift* **30**(6), 489–501.

Vicente-Serrano, S.M. et al. (2010), 'A complete daily precipitation database for northeast spain: reconstruction, quality control, and homogeneity', *International Journal of Climatology* **30**(8), 1146–1163.

WMO (2018), *Guide to Climatological Practices*, 2018th edn, World Meteorological Organization, Geneva, Switzerland.

You, J., Hubbard, K., Nadarajah, S. and Kunkel, K. (2007), 'Performance of quality assurance procedures on daily precipitation', *Journal of Atmospheric and Oceanic Technology* **24**(5), 821–834.