Precipitation and flood impact on rice paddies: Statistics in Central Java, Indonesia

E P A Pratiwi1,3, A O Hartono1, H Z Wijdan1, F Nurrochmad1 and C Setyawan2

1 Department of Civil and Environmental Engineering, Universitas Gadjah Mada, Jalan Grafika No. 2, Kampus UGM, Yogyakarta 55281, Indonesia
2 Department of Agricultural and Biosystems Engineering, Universitas Gadjah Mada, Jl. Sosio Justisia, Bulaksumur, Yogyakarta 55281, Indonesia
3Email: endita.prima.ari.pratiwi@ugm.ac.id

Abstract. Rice, the staple food of most Indonesian people, is mostly produced on Java Island. Flood disasters can cause rice damage, especially during the wet season, when torrential rainfall occurs frequently. The current study investigated the risk of rice damage due to flood disasters. To analyze statistical property and behavior from the 23-years data in Central Java, parametric tests, and the equivalent nonparametric tests were performed. Two-sample t-test and two-sample Kolmogorov-Smirnov (KS) tests were used to identify precipitation threshold that leads to damaging flood. Pearson's $r$ and Spearman's $\rho$ coefficients were then computed to measure the strength of association between precipitation and the area of rice paddies affected by flood disasters. The results indicated that there was a clear higher mean ($p < 0.05$) of the precipitation group causing damaging flood than the other group that did not cause damage. The critical precipitation was about 200 mm in a month. The flood-affected area was not a normal distribution, but the log-transformed data appeared a near-normal distribution. Finally, the correlation tests revealed that the log-transformed affected area is linearly and monotonically dependent on precipitation (Pearson's $r = 0.483$, Spearman's $\rho = 0.475$, $p < 0.05$).

Keywords: Flood Disasters, Rainfall Statistics, Parametric Test, Nonparametric Test, Agriculture Impacts

1. Introduction

The majority of Indonesian rely on rice as their staple diet. More than half of rice production in Indonesia is concentrated on Java Island. This island is about 10% of the total area of the country, but more than two-thirds of the Indonesian population lives there [1]. On the other hand, among all provinces in Indonesia, Central Java is the province with the highest number of natural disaster events [2]. Natural disasters cause losses and damages to the agricultural sectors, particularly on the crop subsector. Furthermore, flooding contributed almost 60% of losses and damage in this subsector [3]. A recent study [4] found that in 2014, flood disasters damaged 60,514 hectares of rice paddies in Central Java (3.35% of total rice paddies area). However, the rice production loss that year was almost 10%. It suggested that flood disasters reduce not only harvest area, but also rice yield per harvested area. The study also highlighted that annual precipitation in 2014 was approximately 30% higher than the average yearly precipitation in five years [4].
Excess precipitation is the main cause of floods. Therefore, the assessments of the relation among the total precipitation amount and estimations of the flood-damaged rice paddies are very important in planning crop risk management for major flood events. Much research has proposed the concept of critical rainfall or rainfall threshold for flood forecasting [5–8]. However, these studies were mostly applied to event-based rainfall, small-sized watershed, and landslides mitigation. To the best of our knowledge, works were rarely dedicated to the agriculture sector. Yang et al. [9] investigated precipitation-crop damage relationships and explored meteorological evaluation indicators to prevent rainfall-induced floods and reduce potential crop losses. The results showed that flooding caused rice disasters are more intense in the early mature stage, followed by the jointing-booting and transplanting-tilling stages. Estiningtyas et al. [10] carried out research to determine critical rainfall value that can influence flood and drought events in some of the central food crops in West Java. They found that average of the critical value of rainfall for flood event for paddy field in the coastal area based on median approach is 140 mm/month with probability 0.6, while for other locations is 166 mm/month with probability 0.68.

Hydrological data, including precipitations and floods, often have the following characteristics, for example, non-negative values, the presence of outliers with high values, positive skewness, non-normal distribution, and dependence on other variables [11]. Lai et al. [12] also observed that rice damage does not follow a normal distribution. It is important to understand the basic characteristics of the data to choose the suitable procedures for data analysis. Classical, widely-used, parametric methods include the assumption about the population parameters and the distributions that the data came from, for example, normality, symmetry, and independency. False assumptions may result in incorrect or inconclusive interpretations from the analysis. Every parametric test has equivalent nonparametric tests that do not require any assumption. Regarded as a robust technique, the nonparametric method works reasonably well over a wide range of situations, as opposed to a technique that could be optimal for some particular situation. The current study aimed to statistically analyze the characteristics of monthly precipitation causing damaging-floods, the data characteristics of rice paddies area damaged by flood disasters, and the correlation of total precipitation and damaged-area. This study also compared the results of the analysis using traditional parametric methods and the alternatives nonparametric methods.

2. Methods

2.1. Study site and data collection
The target area comprised Central Java, a province located in the middle part of Java Island, Indonesia. Central Java covers an area of 32,548 km² and consists of 29 regencies and six cities. Located in the tropical monsoon region, Central Java is familiar with flooding, a common disaster as a result of torrential rainfall, especially in the wet season during October–March. Mean annual precipitation of Central Java is 1,500–2,500 mm. Figure 1 portrays the study area from a geographical point of view.

Daily precipitation datasets from 1987 to 2019 (33 years) were obtained for four weather stations, namely Tegal, Maritim Tanjung Emas, Semarang, and Tunggul Wulung (see figure 1). Daily precipitation data were accessed from the Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG) online database (http://dataonline.bmkg.go.id/). After calculating the average value of daily precipitation from four stations, monthly precipitation was determined as the cumulative value of daily precipitation in a month. A dataset consisting of affected rice paddies area due to flood disasters in Central Java over the year 1997–2019 published by the Ministry of Agriculture, Republic of Indonesia was available and accessed on http://prasarana.pertanian.go.id/iklimoptdpimy/.
2.2. Statistical Analysis

2.2.1. Descriptive statistics. The first step in interpreting the data was to define and explain the data in ways that reflect the basic characteristics of the data, i.e., the central tendency, the variability, and the distribution symmetry. The arithmetic mean (\( \bar{x} \)), the sample variance (\( s^2 \)), and the coefficient of skewness (\( g \)) are the classical measures of central tendency, variability, and distribution symmetry. Kurtosis, a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution, is also calculated.

The classical measures above are not robust in the presence of, or to changes in the magnitudes of, a few outliers. The median, the interquartile range (IQR), and the quartile skew (\( k \)) are the measures of central tendency, variability, and distribution symmetry that are more robust to the influence of outliers. The median, or 50th percentile (\( P_{0.50} \)), is the central value of the distribution when the data are sorted by magnitude. The IQR is defined as the 75th percentile minus the 25th percentile. The 75th percentile (\( P_{0.75} \)), or the upper quartile, is a value that exceeds no more than 75 percent of the data and is therefore exceeded by no more than 25 percent of the data. The 25th percentile (\( P_{0.25} \)), or the lower quartile, is a value that exceeds no more than 25 percent of the data and is therefore exceeded by no more than 75 percent. The quartile skew is defined as the difference in distances of the upper and lower quartiles from the median, divided by the IQR, see equation (1).

\[
k = \frac{P_{0.75} - P_{0.50}}{P_{0.75} - P_{0.25}}
\]

2.2.2. Precipitation threshold determination. The total area of rice paddies affected by floods, or the rice damage, were represented by coupling with historical precipitation data. Rice damage and precipitation data have been coupled to identify the precipitation parameters that might be used as risk indicators and determine the threshold values above which rice damage occurs. Before determining the threshold, we first checked whether precipitations in which damage occur came from different distributions. All precipitation data were classified into two groups, the group of precipitations that caused damage or affecting rice paddies (\( P_{-1} \)) and the group of precipitation that did not cause damage (\( P_0 \)). Classical one-tailed t-test was then performed to see whether the first group (\( P_{-1} \)) have higher mean values than the latter group (\( P_0 \)). Before the t-test performed, F-value for variances was computed to check the variance equality of the two groups. Nonparametric two-sample Kolmogorov-Smirnov (KS) was used as an alternative. The two-sample KS provides an objective ranking of parameter importance. The statistic test of two-sample KS was performed based on the maximum distance (\( D_{\text{max}} \)) between these two empirical distribution functions, as in equation (2).
where \( F_{1,n}(x) \) and \( F_{2,m}(x) \) are the empirical distribution functions of the first and the second sample, respectively, \( n \) and \( m \) are the sizes of each sample and \( \sup \) is the supremum (maximum) function. Note that both \( F_{1,n}(x) \) and \( F_{2,m}(x) \) were computed at each point in each sample. The null hypothesis is the two samples come from similar probability distributions. It is rejected at a given significance level if the test statistic exceeds a critical value \( D_{\text{crit}} \) as in equation (3).

\[
D_{\text{crit}} = c(\alpha) \sqrt{\frac{n + m}{n \times m}}
\]  

The critical threshold of precipitation causing rice damage was defined with two methods. The first was the mean lower limit of 95% confidence interval, and the second was the lower quartiles, or the 25th percentile \( (P_{0.25}) \).

2.2.3. Rainfall-flood damage correlation. Analysis of correlation, consecutively, were conducted for precipitation data in the P_1 group. The correlation between precipitation and the area of rice paddies affected by flood events was analyzed with parametric Pearson's \( r \) and nonparametric Spearman's \( \rho \) correlation. For the large size of the sample \( (n > 20) \), the test for significance using the test statistic \( t \) follows the \( t \)-distribution with \( n - 2 \) degrees of freedom and computed by equation (4).

\[
t = r \sqrt{\frac{n-2}{1-r^2}}
\]  

2.2.4. Plotting empirical probability. In this study, the Weibull plotting method was used to draw an empirical distribution graph. Consider each dataset ordered from smallest to largest: \( x_i, i = 1, 2, 3, ..., n \), the calculated empirical probability of the \( i \)th ranked observation \( (p_i) \) is given using the Weibull plotting method as in equation (5).

\[
p_i = \frac{R_i}{n + 1}
\]  

where \( p_i \) is the probability of the \( i \)th ranked observation, \( R_i \) is the rank of observation, and \( n \) is the sample size. The degree of significance \( (\alpha) \) for any statistical test accepted in this study was 0.05. Before performing a statistical test that needs a normal distribution assumption, normality of the data was tested using a graphical tool and probability plot correlation coefficient (PPCC).

3. Results and discussion

3.1. Characteristics of data

Graphs offer essential detail on data that is impossible to access in some other way. Thus, we first plotted monthly precipitation vs. the area of rice paddies damaged by floods in graphs, as shown in figure 2. It is seen that damaging flood events coincided in months with high rainfall. For example, the monthly precipitation in January 2014 was more than 600 mm and impacted more than 40 thousand hectares of paddy fields. Coupling precipitation data with rice damage data resulted in 162 months in which total precipitation affecting rice paddies and 114 months with non-affecting rainfalls. The annual rainfall in 2014 (2,390 mm) was only slightly higher than the average annual rainfall in 30-years (2,300 mm), but January 2014 precipitation total (657.5 mm) was almost two times than the average January precipitation in 30-years (356.4 mm).

In Table 1, the data of non-affecting precipitations appear to be normal distribution with skewness 0.75 \((\approx 0)\) and kurtosis 2.32 \((\approx 3)\). Likewise, the data of affecting precipitation seems to be distributed
normally with skewness 0.39 (≈ 0) and kurtosis 3.87 ≈ 3). Nevertheless, the affected rice area does not follow a normal distribution. It has skewness 3.31 > 0 and kurtosis 16.45 > 3. Log-transformed rice affected area due to flood could show a normal distribution with skewness -0.35 (≈ 0) and kurtosis 2.40 (≈ 3). That was an arbitrary determination using the values of skewness and kurtosis. Section 3.3 explains a more objective test of normality.

Figure 2. Monthly precipitation (line) and the area of flood-affected rice paddies (column) in Central Java from January 1997 to December 2019.

Table 1. Descriptive statistics of precipitation (P_0, which do not lead to flood, P_1, which lead to flood affecting rice paddies) and affected area in original value and log-transformed value.

| Statistics Parameters | P_0       | P_1       | Affected area | Log affected area |
|-----------------------|-----------|-----------|---------------|-------------------|
| n-samples             | 114       | 162       | 162           | 162               |
| Mean                  | 87.56     | 259.10    | 3779          | 2.97              |
| Standard deviation    | 83.04     | 110.69    | 6590          | 0.85              |
| Median                | 61.05     | 245.60    | 1301          | 3.11              |
| Lower quantile        | 14.10     | 195.38    | 199           | 2.29              |
| Upper quantile        | 148.75    | 327.35    | 4466          | 3.64              |
| Minimum               | 0.00      | 3.30      | 5             | 0.70              |
| Maximum               | 280.00    | 657.5     | 46072         | 4.66              |
| Interquartile range   | 134.65    | 131.98    | 4266          | 1.35              |
| Skewness              | 0.75      | 0.39      | 3.31          | -0.35             |
| Quartile skew         | 0.30      | 0.24      | 0.48          | -0.21             |
| Kurtosis              | 2.32      | 3.87      | 16.45         | 2.40              |

3.2. Precipitation critical threshold
The boxplots in figure 3a visually provide information that there is a clear gap between dataset of precipitations affecting rice paddies and precipitation that have no impact on rice paddies. The upper quartile of non-affecting precipitation (148.75 mm) is less than the lower quartile of affecting precipitation (195.38 mm). Furthermore, as presented in figure 3b, analysis using two-sample KS also show that the group of precipitations causing flood affecting rice paddies and the precipitations group that have no impact came from different populations ($D_{\text{max}} = 0.624$, $D_{\text{crit}} = 0.167$, $p < 0.05$). The maximum difference in the probability was found at some point between 150 – 200 mm.
The two-sample $F$-test is used to test whether the two samples are from normal populations with equal variances. The null hypothesis is that the variances for the two samples are equal. Two-sample $F$-test applied to the two precipitation groups ($P_0$ and $P_1$) proved that $F_{\text{test}} = 1.78 > F_{\text{crit}} = 1.38$ ($\alpha = 0.05$), so the null hypothesis is rejected. In other words, the variances of the two samples are not equal. Hence, a two-sample $t$-test assuming unequal variances (heteroscedastic) should be used to compare the means of two variables. Upper tailed two-sample $t$-test assuming unequal variance validated that the mean value of affecting precipitation is significantly greater than that of non-affecting precipitation ($t_{\text{test}} = 14.7$, $t_{\text{crit}} = 1.65$, $p < 0.05$). The mean value of precipitation causing rice damage at a 95% confidence interval is 241.92 mm – 276.27 mm, while the mean value of precipitation not causing rice damage is 72.15 mm – 102.97 mm. There is a large gap between the mean upper limit of non-affecting precipitation and the mean lower limit of precipitation that affected the rice area.

![Boxplots and empirical probability distribution of precipitations](image)

**Figure 3.** (a) Boxplots and (b) empirical probability distribution of precipitations that did not lead to the flood-affected area (Open circle, $P_0$) and precipitations that lead to the flood-affected area (Solid circle, $P_1$). Dashed line shows the gap between two precipitation groups.

The following steps were used in determining the threshold values above which rice damage occurs. For convenient application, evaluation indicators were an integer, with 0 as the tails. First, we tried the mean lower limit of 95% confidence interval, i.e., 240 mm as the threshold. When monthly precipitation is more than 240 mm, the probability of rice damage is 92.4%. However, there is high risk since up to 47.5% of rice damage caused by monthly precipitation of less than 240 mm. Next, the rainfall threshold was determined using the lower quartile of precipitation affecting the rice area, i.e., 200 mm. If the total precipitation in a month is more than 200 mm, the risk of rice damage is 88.9%. Only 25% of rice damage had been triggered by cumulative precipitation less than 200 mm in a month. Thus, with those considerations, we suggested cumulative precipitation of 200 mm in a month as the critical threshold level.

### 3.3. Correlation between precipitation and flood-affected area

Coefficients of correlation determine the degree of interaction between precipitation and rice damage or the damage induced by heavy precipitation in the region. Pearson's $r$ correlation coefficient measures the strength of the linear association between variables. The Pearson correlation coefficient is used for normally distributed data. Thus, before calculating the correlation coefficient, the normality of the datasets must be checked. The normality test using a graphical tool showed that the precipitation causing flood datasets follows a normal distribution with $r = 0.988$ ($p < 0.05$, figure 4a). According to earlier analysis, the affected area has a log-normal distribution. The normality test using graphs, therefore, was performed towards the log-value of the affected area. Figure 4b shows that the log-
value of affected area is normal distribution with $r = 0.991$ ($p < 0.05$). The computed Pearson's $r$ correlation coefficient is 0.483 ($p < 0.05$), so precipitation has moderate linearity ($r \approx 0.5$) with log-value of rice affected area (see figure 4c). However, using data transformations can be quite problematic [13]. Although the log transformation can decrease the variability of data and make data conform more closely to the normal distribution, data transformations must be applied very cautiously. We recommended that, in most circumstances, researchers abandon these traditional methods of dealing with skewed data.

![Figure 4](https://example.com/figure4.png)

**Figure 4.** Probability plots for (a) monthly precipitation causing affected area and (b) log value of the flood-affected area, with probability plot correlation coefficient (PPCC) correlation coefficient ($r$), as well as (c) correlation analysis of precipitation and flood-affected area. *$p < 0.05$.

Spearman's $\rho$ correlation measures the strength and direction of monotonic associations between two variables, which is "less restrictive" than that of a linear relationship. Unlike Pearson's $r$ correlation, the normality assumption is not necessary to perform Spearman's $\rho$ correlation analysis. However, the data for precipitation and affected area are separately ranked from smallest to largest. The higher rainfall ranks are associated with the higher ranks of the affected area when a positive association exists. The results of Spearman's $\rho$ correlation analysis demonstrated that precipitation has a monotonic relation with rice affected area ($\rho = 0.475$, $p < 0.05$), as in figure 4c.

### 3.4. Implementation and future predictions

Flood events devastate the rice field, reduce the yield of rice, and trigger financial losses. In the first quarter of 2014, when the most destructive flood occurred, the harvested area was 825 thousand hectares, about 40 thousand hectares lower than the harvested area in the first quarter of 2013. Rice yield declines were also observed, from 5.81 tons per hectare in the first quarter of 2013 to 5.15 tons per hectare in the subsequent year. A total loss of approximately 0.76 million tons of rice grain was then projected. Assuming that the unmilled grain price was Rp 3,500 ($\approx 0.25$ USD) per kg, the economic loss was equivalent to 2.7 trillion rupiahs ($\approx 190.9$ million USD).

Throughout the coming decades, flood risk is projected to continue increasing as a consequence of climate change. The Ministry of National Development Planning Republic of Indonesia [14] reported that from 1901 to 1990, in the Java-Bali area, the chance of monthly precipitation of 500 mm (above
critical rainfall threshold defined in this study, 200 mm) increased by 10% – 20% during rainy months, December–January–February. The study [14] also performed climate change projection for the 2030s and 2080s period against the baseline period (1961–1990). The pattern of rainfall shifts indicates a general trend that the increase of rainfall happens during rainy months, especially in the 2080s period for the A1 and A2 scenarios. More climate uncertainties and variabilities are expected in the future.

Potential climate change and inter-annual climate variability can threaten rice production. Some research [15–17] indicates that the threats of significant changes in drought characteristics are more immediate and frequent than floods. Indeed, rice has a critical requirement for high and constant water supply rather than any other food crop. Globally and nationwide, the impacts of extreme floods on grain yield are not significant, as floods may be highly localized [16]. The current study is focused on sub-national agricultural statistics. The outcome of this analysis suggests a notable flood impact on rice production. Our findings offer insight to develop effective climate change adaptation plans; local policymakers should consider more frequent heavy rainfall in the future that may result in rice production loss.

4. Conclusions
Flood disasters damage rice paddies in Central Java every year, especially when excess precipitations occur. The results showed that monthly rainfall more than certain levels (about 200 mm/month) likely cause damage to rice paddies. The results also showed that the rice affected area does not follow a normal distribution, but data transformation into log-value follows a normal distribution. Nonparametric Spearman's correlation test revealed that the affected area is monotonically dependent on precipitation, and parametric Pearson's correlation test confirmed that the log-value of the affected area is linearly dependent on rainfall. We recommend using the nonparametric method because it is more resistant to high outliers and skewness, which are usually found in the data.

This study provided basic information for understanding the amount of total precipitation and the rice affected area relationship in Central Java. Still, the relationship between floods and rice production is a complex issue. Other factors, such as the rice development stage, irrigation infrastructure damage, as well as pests and diseases after floods, may affect productivity, which have not been considered in this study. Further research should be conducted to account for those factors so that effective crop risk management can be designed. Comprehensive analyses of frameworks for interaction between natural disasters and climate change, economic effects of natural disasters as well as policy changes to mitigate food security risks in Central Java should be pursued to further study. Finally, the results of this study might be of use in risk estimates about rice damaged due to floods in Central Java, which could be applied to other areas in Indonesia.

References
[1] Statistics Indonesia 2019 Statistical Yearbook of Indonesia 2019 (Jakarta: Statistics Indonesia)
[2] Indonesia Disaster Data Information 2019 Disaster Spatial Distribution by Provinces (National Disaster Management Authority, BNPB) https://dibi.bnpb.go.id/
[3] FAO 2015 The Impact of Natural Hazards and Disasters on Agriculture and Food Security and Nutrition: A Call for Action to Build Resilient Livelihoods (Rome: FAO)
[4] Pratiwi E P A, Ramadhani E L, Nurrochmad F and Legono D 2020 J. Civ. Eng. Forum 6 69
[5] Diakakis M 2012 Nat Hazards 60 789–800
[6] Knighton J O and Walter M F 2018 Hydrol. Process. 30 3788–803
[7] Golian S, Saghafian B, and Maknoon R 2010 Water. Resour. Manag. 24 3547–59
[8] Montesarchio V, Napolitano F, Rianna M, Ridolfi E, Russo F and Sebastianelli S 2015 Nat. Hazards 75 909–34
[9] Yang J, Huo Z, Wu L, Wang T and Zhang G 2016 Agr. Ecosyst. Environ. 230 221–30.
[10] Estimingtyas W, Boer R and Buono A 2009 J. Agromet 23 946–52
[11] Helsel D R, Hirsch R M, Ryberg K R, Archfield S A and Gilroy E J 2020 Statistical methods in water resources: US Geological Survey Techniques and Methods book 4 chapter A3 458
[12] Lai L H 2010 Nonlinear Anal. Real World Appl. 11 1243–8
[13] Wang H, Lu N, Chen T, He H, Lu Y and Tu X M 2014 Shanghai Arch. Psychiatry 26 105–9
[14] Indonesia Climate Change Sectoral Roadmap (ICCSR) 2010 Scientific Basis: Analysis and Projection of Temperature and Rainfall (Jakarta: The Ministry of National Development Planning Republic of Indonesia, BAPPENAS)
[15] Kang Y, Khan S and Ma X 2009 Prog. Nat. Sci. 19 1665–74
[16] Lesk C, Rowhani P and Ramankutty N 2016 Nature 529 84–7
[17] Amien I, Redjekiningrum P, Kartiwa B and Estiningtyas W 1999 Clim. Res. 12 145–52