Spatial analysis of return period based copula on extreme rainfall data in South Sulawesi

Reski Wahyu Yanti¹, Anwar Fitrianto² and Muhammad Nur Aidi³
¹,²,³Department of Statistics, IPB University, Bogor, Indonesia

Email: ¹ reski_wy@apps.ipb.ac.id

Abstract. The extreme rainfall in an area makes the area vulnerable to various disasters. To reduce the risk of damage caused by floods, it is important to know the characteristics of extreme rainfall. Generally, the characteristics of extreme rainfall are described by one variable. However, most of the extreme rainfall events also need to be explained based on the return period they occur using several variables with copula approach. This study model to the characteristics of extreme rainfall with two variables, they are namely extreme rainfall intensity and extreme rainfall volume. The purpose of this study was to analyze the spatial distribution pattern of the return period from extreme rainfall in South Sulawesi. To determine the characteristics of the return period distribution in South Sulawesi, spatial analysis is carried out using the Moran's index and the LISA index. The results of the spatial autocorrelation analysis with the Moran's index show that there is a relationship between several return period values in South Sulawesi, with the Moran's index value of 0.209. This means that the value of the return period in South Sulawesi has a clustered relationship pattern. Furthermore, the results of the spatial autocorrelation analysis with LISA show that there are seven sub-districts identified as having local spatial autocorrelation. The conclusion obtained from Moran's scatterplot is that 15 sub-districts are the main concern in preventing natural disasters because extreme rainfall in these 15 sub-districts tends to occur more frequently, so that it can lead to various natural disasters.

1. Introduction
Climate is the most important geographical element in influencing human life. The impact of climate change and extreme climate are part of the most serious problems for people's lives in the world [1]. The most influential climate parameter in Indonesia is rainfall. The extreme rainfall in an area makes the area vulnerable to various disasters. Natural disasters caused by extreme rainfall pose a threat to the population and infrastructure in various regions of South Sulawesi Province. Extreme rainfall events have led to dramatic damage to human life and property, most seriously by contributing to urban flooding [2]. Global-scale river flood vulnerability analyses have also been conducted [3]. Generally, the characteristics of extreme rainfall are described by one variable. However, most of the extreme rainfall events also need to be explained based on the return period they occur using several variables to get more accurate results.

This study model to the characteristics of extreme rainfall with two variables, they are namely extreme rainfall intensity and extreme rainfall volume at 53 stations representing sub-districts in South Sulawesi. Extreme rainfall is manifested by a rainfall intensity that is greater than the order of 75% of the rainfall intensity and a rainfall volume that is greater than the order of 75% of the rainfall volume [4]. Both variables are set using the copula approach to determine the best copula type for the data at
The stochastic rainfall generator can take into account the statistical dependence between rainfall event duration and intensity, thanks to a copula approach [5]. Parameter analysis with copula selection is important to model the multivariate probability of natural hazards and combined events [6] so that the type of copula used matches the characteristics of the data at each station. The type of the copula characteristics of the variable extreme rainfall intensity and extreme rainfall volume varies, both variables also have a return period. The return period is the estimated average time between certain events to occur. The return period and risk of extreme hydrological events are critical considerations in water resources management [7]. Water resources design has widely used the average return period as a concept to inform management and communication of the risk of experiencing an exceedance event within a planning horizon [8]. The probabilistic concept of return period is widely used in hydrology as well as in other disciplines of geosciences to indicate critical event rareness [9]. The phenomenon of extreme rainfall does not occur every time, so there is a return period for both variables. The problem discussed in this study is how to formulate a natural hazard mitigation program referring to the distribution pattern of the return period in an area. Knowledge regarding the distribution pattern of the return period can assist policymakers in making decisions regarding extreme rainfall events in South Sulawesi that have the potential to cause various losses.

2. Related Work
Climate elements such as rainfall, apart from being a much needed natural resource, can also be a source of disasters. The extreme rainfall in an area makes the area vulnerable to floods or other disasters. Such extreme rain events harm socio-economic development so that cases like this are an important concern to know their characteristics. In this study, a spatial analysis was carried out to see the distribution pattern of the return period at 53 stations in South Sulawesi. Climatic observables are often correlated across long spatial distances and extreme events so that revealing spatial patterns is of great importance for weather forecasting in general and extreme-event prediction in particular [10].

Research with spatial analysis of extreme rainfall data was conducted by [4] for cases of extreme rainfall in Xinjiang, China. The research in [11] also conducted a spatial analysis of daily data for 68 stations using indices of extreme rainfall and temperature extremes. Other research related in [12] analyzed the spatial relationship with the number of variables in space using monthly rainfall. In addition, the journal in [13] analyzed the spatial autocorrelation of extreme rainfall events using grid data on the 95th percentile daily rainfall that was examined with the Moran's index and the LISA index, the results obtained showed a strong spatial autocorrelation in extreme rainfall events.

For the preparation of a natural hazard mitigation program, we need to study the distribution pattern of the return period in an area, so that it can provide knowledge about the distribution of configured points in space. Analysis of point distribution patterns contains several techniques to explain the spatial distribution of these points to see whether a pattern of the distribution of a point tends to be clustered, random, or regular pattern. Therefore, a spatial analysis was conducted to determine the spatial association of the return period in South Sulawesi. Spatial analysis is a set of methods to find and describe the level or pattern of a spatial phenomenon. Through spatial analysis, it is hoped that new information can be used as a basis for decision making. Based on this description, this study aims to analyze the characteristics of the spatial distribution of the return period for extreme rainfall in South Sulawesi.

3. Methodology

3.1. Data
The data used in this study are returned period data obtained from extreme rainfall data in the form of 75th percentile intensity and 75th percentile volume of monthly rainfall. Rainfall data is obtained from the Global Satellite Mapping of Precipitation (GSMaP) which is downloaded via ftp://hokusai.eorc.jaxa.jp. The data consists of rainfall data at 53 stations representing sub-districts in South Sulawesi in the period 2011-2018.
3.2. Data analysis steps

The research carried out was completed through the following data analysis steps:

- Calculates the number of days where rainfall is greater than or equal to the 75th percentile of daily rainfall in a month \( (D_{75}) \).
- Calculates the volume of rainfall with rainfall greater than or equal to the 75th percentile of the daily rainfall for a month \( (P_{75}) \).
- Calculate the intensity value of the 75th percentile of rainfall using the formula \( P_{75} / D_{75} \) is symbolized by \( I_{75} \).
- Determining the best copula model for the dependency of the two variables using the empirical copula method with the following equation [14].

\[
K_t(u, v) = \frac{1}{n} \sum_{t=1}^{n} \left( \left( \frac{r_1(r_{2|u})}{2} \right) \left( \frac{r_2(r_{1|v})}{2} \right) \right)
\]

- Calculating the return period for the intensity and volume of extreme rainfall based on the selected copula type using the following equation [4].

\[
T_{(X>x \text{ or } Y>y)} = \frac{1}{p(X > x \text{ or } Y > y)} = \frac{1}{1 - F(x, y)}
\]

- Creating a spatial weighting matrix based on the coordinates of each sub-district in South Sulawesi province.
- Calculate the statistical value of Moran's index using the following equation [15].

\[
I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2}
\]

Information:

\( n \) = the number of observations

\( x_i \) = variable data i-th location \( (i = 1, 2, ... , n) \)

\( x_j \) = variable data j-th location \( (j = 1, 2, ... , n) \)

\( \bar{x} \) = the average of data

\( W_{ij} \) = spatial weighted matrix element i-th row j-th column

- Testing the Moran's index hypothesis on the return period value to a state that there is a positive or negative spatial autocorrelation. At this stage, it will be seen whether the characteristics of the return period in a sub-district will influence (or be influenced) by the characteristics in the nearest sub-district. The characteristics of the return period between sub-districts in South Sulawesi are analyzed using the Moran's index to see whether the pattern of the distribution of a point tends to be clustered or clustered, random, or forms a regular pattern.

- Create Moran's scatterplot

Moran's scatterplot provides an exploratory visual analysis to detect spatial autocorrelation. This diagram can also be used to determine the pattern of relationships that occur between return period values at the sub-district level in South Sulawesi, whether they form a clustering pattern or not.

- Calculate the statistical value of the LISA index using the following equation [15].

\[
I_i = Z_i \sum_{j=1}^{n} W_{ij} z_j
\]

\( z_i \) and \( z_j \) are \( z_i = \frac{(x_i - \bar{x})}{\sigma_x} \) , \( z_j = \frac{(x_j - \bar{x})}{\sigma_x} \)

\( \sigma_x \) is the deviation standard value of the variable \( x \).
Hypothesis test of the LISA index on the return period value.
A special case in spatial dependency is the Local Indicator Spatial Autocorrelation (LISA) to show whether or not there is an effect of the return period value in one sub-district on other nearby sub-districts. From testing the LISA index hypothesis, it will be obtained the significance of the local relationship in each sub-district.

4. Results and Discussion

4.1. Distribution of return period data in South Sulawesi
The following will use the return period data to see the distribution of return period values. Table 1 shows a summary of the return period data as an overview of the return period data.

| Statistics    | Value |
|---------------|-------|
| Lots of data  | 53    |
| Average       | 6.594 |
| Quartile 1    | 5.9   |
| Median        | 6.7   |
| Quartile 3    | 7.1   |
| Minimum       | 5.4   |
| Maximum       | 7.8   |

Based on the description of the return period value in Table 1, it is known that the smallest return period value is 5.4, while the highest return period value is 7.8. The return period value is calculated based on the type of copula selected from the results of the empirical copula method. The types of copula chosen for each station vary, consisting of 16 stations that follow the Clayton copula model, 23 stations that follow the Gumbel copula model, and 14 stations that follow the Frank copula model. Furthermore, the return period data per sub-district will be plotted on a map of South Sulawesi. The results of plotting the return period data can be seen in figure 1 as follows.

Figure 1. Distribution of return period data in South Sulawesi.
The information for the sub-district code of South Sulawesi Province is displayed as follows:

1. Bacukiki
2. Soppeng Riaja
3. Donri
4. Lalabata
5. Barru
6. Taneterangan
7. Pattallassang
8. Labakkang
9. Pangkajene
10. Camba
11. Maros Baru
12. Simbang
13. Panakukang
14. Bonto Marannu
15. Tompuhulu
16. Polomangkeng Selatan
17. Polomangkeng Utara
18. Somba Opu
19. Turatea
20. Tamalatea
21. Arungkeke
22. Bissapu
23. Pajukkukang
24. Ujung Bulu
25. Tompobulu
26. Gangking
27. Bulukumba
28. Sinjai Barat
29. Sinjai Selatan
30. Sinjai Utara
31. Tonra
32. Kahu
33. Mare
34. Bengo
35. Tanete Riattang
36. Awangpone
37. Marioriwawo
38. Sabang Paru
39. Maniang Pajo
40. Dua Pitue
41. Duampanua
42. Pitumpanua
43. Baraka
44. Lembang
45. Alla
46. Palopo
47. Nanggala
48. Telluwanua
49. Lamasi
50. Suka Maju
51. Bone-bone
52. Barau
53. Mangkutana

Figure 1 shows a map of the return period data distribution in South Sulawesi with a dot pattern where on the map it can be seen the high and low return period values in each sub-district based on the size of the circle size. In the sub-districts that have the lowest return period value of 5.4, it can be seen that the red circle representing this value is smaller than the other circles. The low return period value illustrates that this location or region has a more frequent occurrence of extreme rainfall than other locations. The smallest return period value is located in Alla Sub-district, while the largest return period value is located in Camba Sub-district and Maros Baru Sub-district.

4.2. Analysis of Moran's index spatial autocorrelation

Hypothesis testing of Moran's global index is carried out to state that there is a positive or negative spatial autocorrelation. Generally, spatial autocorrelation is a measure of the similarity of objects in space (distance, time, and region). If there is a systematic pattern in the spread of an object, then there is spatial autocorrelation. Spatial autocorrelation shows that the observed attribute value in a certain area is related to the observed attribute value in other nearby areas.

The result of the spatial autocorrelation analysis with the Moran's index test shown in figure 2 proves that there is a significant spatial autocorrelation, this indicates that there is a correlation between the return period value of sub-districts in South Sulawesi with the Moran's index value is 0.209. This means that the value of the return period between sub-districts in South Sulawesi is spatially correlated, this can be seen from the Moran's index value which is greater than 0. The positive Moran's index value shows the spatial autocorrelation of the return period values between sub-districts in South Sulawesi which has a clustered relationship pattern and has a tendency to have similar characteristics in adjacent locations.
Furthermore, a test was conducted to see the spatial correlation in the return period data between sub-districts in South Sulawesi. From the results of hypothesis testing, the p-value is 0.006596. The value is smaller than $\alpha = 0.05$, so that the decision is rejected $H_0$. Thus, it can be concluded that there is a correlation (spatial autocorrelation) in the return period value between sub-districts in South Sulawesi at the 5% significant level.

Research with Moran's index analysis in the field of hydrometeorology was also carried out by [16] on 6 stations of rainfall variables. The results of the analysis with the Moran's index in this study show positive spatial autocorrelation, which has spatial connectivity between the observed areas.

4.3. Moran’s scatterplot

Moran's scatterplot provides an exploratory visual analysis to detect spatial autocorrelation. In addition, this diagram can also be used to determine the pattern of relationships that occur between regions, whether they form a clustering pattern or not.

The pattern of the return period values relationship at the sub-district level in South Sulawesi based on Moran's scatterplot is shown in figure 3 above. From this figure, it can be seen that most of the return period relationships between sub-districts in South Sulawesi are in quadrant I (High-High). This means that the sub-districts in this quadrant have high return period values and are surrounded by sub-districts that have high return period values which are also more dominant than the return period relationships in other quadrants. There are 21 sub-districts in quadrant I. Meanwhile, there are 5 sub-districts in quadrant II (Low-High), there are 15 sub-districts in quadrant III (Low-Low), and there are 12 sub-districts in quadrant IV (High-Low).
Furthermore, what concerns us are the sub-districts that are in quadrant III, because the sub-districts that are in this quadrant have small return period values and are surrounded by sub-districts that have small return period values too, meaning that the extreme rainfall that occurs tends to be more frequent in these areas and it can lead to natural disasters such as floods and landslides. Therefore, these areas require more attention than other areas in handling natural disasters. The results of Moran's scatterplot analysis show that 15 sub-districts are in quadrant III as in table 2 below.

**Table 2. List of sub-districts that are located in quadrant III**

| Name of sub-districts | RP |
|-----------------------|----|
| Bacukiki              | 5.6 |
| Soppeng Riaja         | 5.9 |
| Donri                 | 5.6 |
| Lalabata              | 5.7 |
| Barru                 | 5.6 |
| Taneteriaja           | 5.6 |
| Tompobulu             | 6.1 |
| Polombangkeng Utara   | 5.8 |
| Turatea               | 5.8 |
| Tamalatea             | 5.6 |
| Arungkeke             | 5.9 |
| Bissapu               | 5.9 |
| Sinjai Barat          | 5.6 |
| Sinjai Selatan        | 5.6 |
| Sinjai Utara          | 6.4 |

**4.4. Local Indicator Spatial Autocorrelation (LISA)**

The research in [17] conducted a local spatial autocorrelation analysis with the LISA index on DHF incidence data and weather variability such as rainfall, air temperature, and humidity. The results of this study indicate that the distribution of DHF events in the Tegal Regency in 2012-2018 has a spatial pattern that tends to be clustered. In addition, changes in rainfall, air temperature, and humidity can increase the chance of dengue hemorrhagic fever.

A special case in spatial dependency is the Local Indicator Spatial Autocorrelation (LISA) which can show local observations of the observed variables. This means that observations at one location depend on observations at other locations that are located nearby. From the LISA test, it was obtained the significance of the relationship locally in each sub-district. The results of the LISA test conducted showed that there was local spatial autocorrelation in several sub-districts as shown in table 3 below.

**Table 3. The results of LISA test for each sub-district in South Sulawesi**

| Name of sub-districts | Ii   | E.ii  | Var.ii | Z.ii  | Pr(z > 0) |
|-----------------------|------|-------|--------|-------|-----------|
| Soppeng Riaja         | 6.77E-01 | -0.01923 | 0.08184 | 2.43212 | 0.00751 |
| Barru                 | 7.18E-01 | -0.01923 | 0.07453 | 2.70020 | 0.00346 |
| Camba                 | 1.68E+00 | -0.01923 | 0.28754 | 3.16768 | 0.00077 |
| Maros Baru            | 1.67E+00 | -0.01923 | 0.27574 | 3.22536 | 0.00063 |
| Turatea               | 6.53E-01 | -0.01923 | 0.13807 | 1.80860 | 0.03526 |
| Tamalatea             | 6.47E-01 | -0.01923 | 0.07989 | 2.35766 | 0.00920 |
| Arungkeke             | 5.80E-01 | -0.01923 | 0.09942 | 1.89892 | 0.02879 |
The results of the Local Indicator Spatial Autocorrelation (LISA) analysis of the return period values in each sub-district in Table 3 show that 7 sub-districts were identified as having spatial autocorrelation with a significance level of 0.05, and 46 sub-districts were not identified as having spatial autocorrelation with a significance level of 0.05. With the following details, there are 7 sub-districts that have a spatial autocorrelation of return period values with a significance level of 0.05, including Soppeng Riaja, Barru, Camba, Maros Baru, Turatea, Tamalatea, and Arunkeke sub-districts. Meanwhile, 46 other sub-districts do not have a spatial autocorrelation of return period values with a significance level of 0.05. Sub-districts that have significant spatial autocorrelation are following the results of Moran's scatterplot, which is located in quadrant I (High-High) and located in quadrant III (Low-Low). Quadrant I (High-High) indicates a location that has a high observation value surrounded by a location that has a high observation value. Quadrant III (Low-Low) indicates a location that has a low observation value surrounded by a location that has a low observation value. This means that sub-districts located in quadrant I and quadrant III have positive spatial autocorrelation. Therefore, the seven sub-districts that have significant spatial autocorrelation obtained from the LISA analysis are included in quadrant I and quadrant III.

5. Conclusion
The results of selecting the best copula model with the Empirical copula show that the characteristics of the distribution model along with the intensity and volume of extreme rainfall vary at each station. Of the 53 stations in South Sulawesi, 16 stations follow the Clayton copula model, 23 stations follow the Gumbel copula model, and 14 stations follow the Frank copula model. This shows the importance of selecting the copula model so that the model used is following the characteristics of the data. Based on the results of the spatial autocorrelation analysis with the Moran index, it shows that there is a relationship between several return period values between districts in South Sulawesi, with the Moran index value of 0.209. This means that the value of the return period between sub-districts in South Sulawesi is spatially correlated with a clustered pattern. Furthermore, what concerns us are sub-districts that are in quadrant III from the results of Moran's scatterplot, because the sub-districts that are in this quadrant have a small return period value and are surrounded by sub-districts that have a small return period value too, meaning that the extreme rainfall that occurs tends to be more often in these areas and it can lead to natural disasters such as floods and landslides. Therefore, these areas require more attention than other areas in handling natural disasters.

The return period investigated in this study was only limited to bivariate cases, while in various cases it could be influenced by many variables so that further research could be developed in multivariate cases. This research is also limited to high extreme cases, while there are still many types of return period formulas that can be used, such as in the case of low extreme rainfall which can also cause various problems in the form of drought and so on.

References
[1] Klein Tank AMG, Zwiers FW and Zhang X 2009 Guidelines on analysis of extremes in a changing climate in support of informed decisions Climate Data and Monitoring WCDMP-No. 72 p 52
[2] Duan W, He B, Nover D, Fan J, Yang G, Chen W, Meng H and Liu C 2016 Floods and associated socioeconomic damages in China over the last century Natural Hazards 82(1) pp 401–13
[3] Tanoue M, Hirabayashi Y and Ikeuchi H 2016 Global-scale river flood vulnerability in the last 50 years Scientific Reports 6 36021
[4] Zhang Q, Singh VP, Li J, Jiang F and Bai Y 2012 Spatio-temporal variations of precipitation extremes in Xinjiang, China Journal of Hydrology pp 7–18
[5] Peres DJ and Cancelliere A 2018 Modeling impacts of climate change on return period of landslide triggering Journal of Hydrology 567 pp 420-34
[6] Sadegh M, Moftakhari H, Gupta H V, Rago E, Mazdiyasni O, Sanders B, Matthew R and
AghaKouchak A 2018 Multihazard scenarios for analysis of compound extreme events. *Geophysical Research Letters* **45** pp 5470–80

[7] Du T, Xiong L, Xu C, Gippel CJ, Guo S and Liu P 2015 Return period and risk analysis of nonstationary low-flow series under climate change *Journal of Hydrology* **527** pp 234–50

[8] Read LK and Vogel RM 2015 Reliability, return periods, and risk under nonstationarity *Water Resources Research* **51** 6381–98

[9] Volpi E, Fiori A, Grimaldi S, Lombardo F and Koutsoyiannis D 2015 One hundred years of return period: Strengths and limitations *Water Resources Research* **51** 8570–85

[10] Boers N, Goswami B, Rheinwalt A, Bookhagen B, Hoskins B and Kurths J 2019 Complex networks reveal global pattern of extreme-rainfall teleconnections *Nature Partner Journal* **566** pp 373–77

[11] Deng H, Chen Y, Shi X, Li W, Wang H, Zhang S and Fang G 2014 Dynamics of temperature and precipitation extremes and their spatial variation in the arid region of northwest China *Atmospheric Research* pp 346–55

[12] Wagesho N, Goel NK and Jain MK 2013 Variabilité temporelle et spatiale des précipitations annuelles et saisonnières sur l’Éthiopie *Hydrological Sciences Journal* **58** pp 354–73

[13] Darand M, Dostkamyan M and Rehmanic MIA 2017 Spatial autocorrelation analysis of extreme precipitation in Iran *Russian Meteorology and Hydrology* **42** pp 415–24

[14] Li X and Babovic V 2018 A new scheme for multivariate, multisite weather generator with inter-variable, inter-site dependence and inter-annual variability based on empirical copula approach *Climate Dynamics* **52** pp 2247–67

[15] Lee J and Wong DWS 2001 *Statistical Analysis with Arcview GIS* (New York: John Wiley And Sons)

[16] Inarossy N and Yulianto SJP 2019 Klasifikasi Wilayah Risiko Bencana Kekeringan Berbasis Citra Satelit Landsat 8 Oli Dengan Kombinasi Metode Moran’s I dan Getis Ord G * (Studi Kasus : Kabupaten Boyolali dan Klaten ) *Indonesian Journal of Computing and Modeling* **2** pp 37–54

[17] Arieskha FTA, Rahardjo M and Joko T 2019 The Association between Weather Variability and Dengue Hemorrhagic Fever in Tegal Regency *Jurnal Kesehatan Lingkungan* **11** p 339