Spatial variability of annual rainfall in East Java Regions: application of histogram and normal QQ-Plot

E B Setiawan¹, Indarto*¹, Askin², Retno Utami A³,
¹ Dept. Agricultural Water Resources Management, Postgraduate Faculty, University of Jember, East Java, Indonesia
² Dept. Agricultural Engineering, Faculty of Agricultural Technology, University of Jember, East Java, Indonesia
³ Dept. Civil Engineering, Faculty of Engineering, University of Jember, East Java, Indonesia

*indarto.ftp@unej.ac.id

Abstract. This research aims to analyse the spatial variability of annual rainfall. Daily rainfall data from 936 rain-gauges in East Java regions are used as the primary input. The average annual rainfall and the maximum annual rainfall are obtained from the daily rainfall data. Research procedures consist of data pre-treatment, ESDA, interpolation, and interpretation. Histogram and QQ-Plot were used to describe the spatial variability of annual rainfall in each sub-region. Next, an IDW interpolation method was used to create a thematic map of the annual rainfall. The results showed that local spatial variability of rainfall could be visualised more detail for each sub-region using histogram and QQ-Plot. The thematic map showed that the distribution of average annual rainfall in the region range from 1000 mm/year up more than 3000 mm/year. Maximum annual rainfall range between 2,100 mm/year up to 4,500 mm/year. The result also shows the positive correlation between the altitude of the rain-gauge and local annual rainfall received.

1. Introduction
The Rainfall has a specific spatial variability in each region. The spatial variability of rainfall is a variation of rainfall based on the spatial region that forming of rainfall distribution. According to Indarto [1], [2] spatial variability of rainfall is a variation of rainfall in the dimensions of space. The rainfall variability can be illustrated by spatial analysis. The rainfall variability used to determine the pattern of rainfall distribution and the cropping patterns for farmers. Also, it can be an early warning system for floods and droughts. The spatial variability shown by the differences in rainfall received for each sub-region annually, monthly, or daily [1], [2].

Exploratory Spatial Data Analysis (ESDA) is commonly used to plot the data distribution, to show the global and local trends, to evaluate the spatial auto-correlation and to understand the covariance between data series [1], [3], [4]. Some statistical tool in the ESDA is suitable for analysis of hydro-meteorological data.

The simple analysis in ESDA is to summarise the statistical value of data series. In other word, ESDA is a simple data model. It is defined associating the element of ‘rough’ and ‘smooth’ with different attribute properties [5]. The graphic in ESDA is generally in the form of histograms, pie charts, box plots and dot-plots. None of the graphs above shows a spatial perspective of data series, However, ESDA aids to link the data with maps and tables. Therefore, data is more meaningful for spatial analysis. The selection of objects through a linking method can be made automatically through
programming, or users can define it through the graphics [1,3]. Some statistical tools are available in ESDA, i.e., Histograms; Normal QQ-Plots; Voronoi-Map; Thiessen Polygons; Box-Plots [1, 3, 4].

Spatial variability analysis can be used for various measurement methods or climate variables or hydrometeorological variables. Haberlandt [6] uses some conventional data or the results of point measurements at the rain-gauges. Furthermore, Ellouze, et al. [7] uses meteorological radar data to study the spatial variability and data interpolation. [4] The uses the climate data as an input for the interpolation process. Moges, et al. [8] uses interpolated hydro-climatology data.

The time interval used to describe the spatial variability of rainfall, can also vary from hourly, daily, monthly, seasonal, or annually. According to Jones, et al. [9] using monthly and annual data to explain the spatial variability of rainfall. The interpolation method to describe the monthly and annual spatial variability of rainfall is also done by Coulibaly, et al. [10], and Shiddiq, et al. [11].

The conventional rainfall measurement based at measuring point. While the map of spatial distribution that contains the value of variability which can describe the phenomenon called as a region-based map. Furthermore, the point data at rain-gauges can be interpolated using i.e: the Thiessen polygon [1], [3], [4], the spline interpolation [12]–[14], the kriging interpolation [18]–[20], and IDW interpolation methods [1], [2], [14]–[18].

This paper presents the results of the spatial variability analysis of annual rainfall. The spatial distribution of data shows the spatial variability through histogram and the normal QQ-plots. The thematic maps produced are utilised in planning, monitoring and evaluating the water resources management in the region.

2. Methods
2.1 Location and Time of Research
This research was conducted from February to December 2018. The research was held in East Java Province.

![Figure 1. Location of rain-gauges](image)

There are 936 rain-gauges locations used in this study (Figure 1). The data processing was carried out in the Laboratory of Environmental Control and Conservation Engineering (TPKL) Department of Agricultural Engineering - Agricultural Technology Faculty - University of Jember.
2.2 Data input and tools

The input for this study is the daily rainfall data obtained from 936 rain-gauges (Figure 1). The annual rainfall is calculated as cumulative daily rainfall for one year. The lengths of the recording period of daily rainfall data range from 5 to more than 50 years. This study uses means and maximum annual rainfall from existing recording periods. Each point location is represented using two values (mean and maximum of annual rainfall).

The focal point of this study is the spatial variability. The point distribution over space (East Java regions) is more important than the variation of a variable in time. Therefore, it is assumed that the length of recording periods is less important than the spatial distribution of point location (measurement site). The average and the maximum value in this study is assumed as a variable that represented the value of annual rainfall phenomenon for specifics point location.

The equipment used for the analysis includes (1) PC, (2) Excel and GIS software for the preliminary analysis, ESDA, and thematic maps. The DEM map clipped from ASTER G-DEM2 (https://asterweb.jpl.nasa.gov/gdem.asp), and it is utilised to illustrate the elevation of the place in the region.

2.3 Research procedure

The stages of research cover (1) data inventory, (2) preliminary analysis, (3) ESDA analysis, (4) interpolation of IDW methods, and (7) the thematic maps making.

2.4 Data Inventory

Rain gauge measurement is conducted on a daily basis with a 24 h observation period, beginning and ending each day at 9:00 a.m. After the data tabulation and verification process is complete. These data and their corresponding geographical coordinates were exported into a geographic information system (GIS) in shapefile format[21].

The identification of each column in the table (1) is as follows: column 1 is the rain-gauge name; column 2 and 3 are mT (East meter) and mU (North meter) for the UTM Zone 49S WGS84 projection system. Furthermore, column 4 is El (m) contain the altitude of the rain-gauges location (masl); column 5 contains Pr which is the period of data recording (years); column 6 (Hth_Max) contains the maximum annual rainfall, and column 7 (HThn_rrt) contains the average annual rainfall.

| Rain-gauge name | mT     | mU     | El (m) | Pr (year) | Hthn_Max | HThn_rrt | HRB 1hr max | Hjn 1hr max | Hjn 2hr max | Hjn 3hr max |
|-----------------|--------|--------|--------|------------|-----------|-----------|-------------|-------------|-------------|-------------|
| Adiboyo         | 760007 | 9138450| 11     | 29         | 1995      | 1234.3    | 102.9       | 190         | 230         | 250         |
| Arah Makam      | 774657 | 9132016| 190    | 29         | 4449      | 2568.0    | 214.0       | 258         | 315         | 413         |
| Asemjajar       | 783667 | 9138745| 135    | 30         | 2998      | 1670.4    | 139.2       | 175         | 216         | 267         |
| Bago            | 775218 | 9134509| 140    | 30         | 3102      | 1999.7    | 166.6       | 175         | 226         | 313         |
| Bantaran        | 737688 | 9129928| 87     | 30         | 2691      | 1640.5    | 136.7       | 322         | 364         | 386         |
| Banyuanyar      | 752564 | 9130402| 89     | 29         | 2660      | 1624.8    | 135.4       | 135         | 235         | 271         |

Sources: data processing

2.5 Preliminary Analysis

The preliminary analysis is performed by displaying the histogram of annual rainfall distribution in each sub-region. In this case, the 35-year annual rainfall is displayed in the interval class from 500 to 5000 mm/year.

2.6 ESDA Analysis

ESDA (Exploratory Spatial Data Analysis) is applied to analyse the distribution and trend statistically. Many features can be utilised in ESDA (Histogram, Voronoi, Normal QQ-Plot, Thiessen Polygon, etc). ESDA can be used to analyse the spatial-variability of data through the histogram, distribution
frequency, Voronoi Map and QQ-Plot. In this paper, the ESDA analysis results are histogram and QQ-Plot tools. The analysis follows some procedures as published in Indarto [1], [2].

2.7 Interpolation and thematic maps
The Inverse Distance Weighting (IDW) method is utilised to create thematic maps of the spatial distribution of annual rainfall. The final result of this analysis is a map of the spatial distribution of annual rainfall. IDW interpolation method assumes that the closer the distance of a point to an unknown point, the higher the influence would be [1], [3], [4]. The evaluation of the IDW methods is done by plotting the graph through the cross-validation feature. Cross-validation is used to assess the accuracy of the interpolation model.

3. Results and Discussion

3.1 Histogram of frequency distribution
Figure 2 displays the histogram of frequency distribution of eight rainfall stations sampled by altitude class. The altitude map is generated from the DEM (digital elevation model). The ASTER G-DEM2 downloaded from the website (https://asterweb.jpl.nasa.gov/gdem.asp) and clipped for East Java region boundary.

![Figure 2. The histogram frequency distribution of annual rainfall](image)

The correlation between the altitude of the rain-gauge and the annual rainfall received are illustrated as follow. The altitude of Socah and Ajung are relatively similar (between 0 to 200 masl) in the coastal area. The annual rainfall recorded at the two sites range from 1500 to 2000 mm. Sambiroto and Saronggi are located at the altitudes range from 201 to 500 meters above sea level (masl). The annual rainfall in Sambiroto occurs ~2000 mm/year, while in Saronggi, the annual rainfall occurs between 1000-2000 mm. Then, Wonosroyo and Sukorejo are located at altitude class range from 501 to 1,000 masl. The recorded annual rainfall at Wonosroyo occurs mostly between 1500-2000 mm. While, in Sukorejo the annual rainfall occurs from 2000 to 2500 mm. Furthermore, Senduro and Ijen Crater represent the rain-gauges where located at an altitude of > 1000 masl. At Senduro the annual rainfall occurs mostly between 2000-2500 mm/year. In Ijen Crater the annual rainfall occurs between 1000-1500 mm.
Based on the description above, it can be seen that even though the position and altitude of the rainfall station are relatively similar, the amount of annual rainfall received is varies. This show the spatial variability of annual rainfall in sub-region.

3.2 Annual rainfall vs. rain-gauge altitude

The analysis of the annual rainfall through histogram show that the higher the location of the rain-gauge causing heavier rainfall. The following description shows a plot of the correlation between the amount of annual rainfall and the rain-gauges altitude. These samples are taken from two sub-regions as they are limited by the rectangle in Figure (3).

Figure 4 shows the correlation between the altitude of rain-gauges and the average annual rainfall received in mm/year for two samples regions (a) and (b). Figure 4a show the value of $r^2 = 0.7$ and Figure 4b, the value of $r^2 = 0.5$. The topography (altitude of location) influences the annual rainfall received, specifically in sub-region west (Figure 4.a). In this case the altitude of annual rainfall are linearly decreased from south to north east direction and it is linear with the rainfall received for each location. In other case, in sub-region east the increase and decrease of altitude is not linear to one direction, therefore the correlation between altitude and annual rainfall received is less important.

![Figure 3. Two sampling areas](image)

(a) sub-region west  
(b) sub-region east

**Figure 4** Scatter-plot of annual rainfall vs altitude of rain-gauges.
3.3 The histogram for spatial variability analysis

3.3.1 The average annual rainfall. Figure 5 shows the average annual rainfall histogram constructed from the 936 rain-gauges. Furthermore, the average annual rainfall is calculated from the average value of all data available.

The average annual rainfall value in every rain-gauge is analyzed using the histogram feature on ESDA. The average annual rainfall histogram is constructed from the 936 rain-gauges. The histogram (Figure 5) shows the statistical summary of average annual rainfall values from all locations (i.e. Min = 201, Max = 4308, Average = 1736).

The histogram (Figure 5) also visualised the statistical distribution values (i.e. standard deviation = Std.Dev = 469.41, the number of stations = 936, Skewness = 0.70, Kurtosis = 4.69, 1st Quartile = 1444 mm/year, median = 1676 mm/year, and 3rd Quartile = 1993 mm/year.

![Figure 5. Average annual rainfall histogram](image)

Figure 5 shows the spatial distribution of average annual rainfall values ranges between 200-1430 mm/year. The values of 200-1430 mm/year occur mostly at coastal and lowland area of East Java region. The histogram illustrated the frequency distribution of each class bin and their statistical values as Figure 5.

![Figure 6. Spatial distribution of rainfall values of 200-1430 mm/year](image)

Figure 6

Figure 7 shows the spatial distribution average annual rainfall values between 2670 - 4310 mm in East Java. Figure 7 displays the distribution of average annual rainfall values range from 2670 to 4310 mm/year that occurs mostly at volcanic slopes area in the middle of the regions. The detail of statistical resume and distribution also visualised in the graphical histogram.
Figure 7. Spatial distribution of rainfall values of 2670 - 4310 mm/year

3.3.2 The Maximum Annual Rainfall. The value of annual maximum is determined as the maximum of annual rainfall data series calculated for each station. The histogram and maps (Figure 8) are produced using ESDA from the 936 available rain-gauges.

Figure 8 display the spatial distribution of maximum annual rainfall values ranges between 3100 - 6100 mm/year.

Figure 8. Distribution of spatial rainfall values of 3100 - 6100 mm/year.

The values occurs mostly at plateau area in the middle of the region. The frequency for the three bin class is shown at the complement graphic of Figure 8.

3.4 Normal QQ-Plots
Normal QQ-Plot is utilized to compare the distribution of data series to the standard normal distribution. The distribution of the average and maximum annual rainfall data is compared to the standard normal distribution and presented in Figure 9.

Figure 9. The Normal QQ-plot of: (a) average annual and (b) annual maximum rainfall.
The distribution of data on a standard normal graph is represented in a straight line from the lower left to the upper right of the graphic. The station having un-normally distribution illustrated by the spread of points in the upper right corner.

3.5 Spatial Interpolation

3.5.1 The average annual rainfall map. Figure 10 shows the thematic map of the average annual rainfall. The maps obtained from the interpolation of the average annual rainfall data at each rain-gauges. The interpolation map is done by using the IDW interpolation method. The heaviest rainfall occurs mostly in the middle and south areas of the region. This areas dominated by volcanic and plateau areas in high altitude. The less average annual rainfall (red and orange colours) occurs mostly in the nothern coastal area (from east to west).

3.5.2 The maximum annual rainfall map.

Figure 11 shows the map of spatial distribution of the annual maximum rainfall in East Java. The maximum annual rainfall in the regions ranges from 1850 to 5180 mm/year. The heaviest maximum annual rainfall occurs mostly in the in the middle and the southern area of East Java. It is visualised by the blue and dark green colours areas.
Figure 11. The spatial distribution of maximum annual rainfall map

4. Conclusion
The analysis of the temporal and spatial variability of the average and maximum annual rainfall has been carried out in East Java region. The daily rainfall data from 936 rain-gauges are utilized as the input. The results show that local spatial variability of rainfall can be visualized more detail for each sub-region using histogram and QQ-Plot. The thematic map showed that the distribution of average annual rainfall in the region range from 1000 mm/year up more than 3000 mm/year. The maximum annual rainfall range between 2,100 mm/year up to 4,500 mm/year. The result also shows the positive correlation between the altitude of the rain-gauge and local annual rainfall received.

Acknowledgement
The publication of this research is sponsored by Hibah Keris PSDA (Pengembangan Sumberdaya Air) year 2018. The rainfall data collected from the 9 UPT offices in East Java. The work prepared by undergraduate of Agricultural Engineering Dept (Gazhi, Dhimas) and post grauate student (Erwan Bagus Setaiwan) of Dept. Agricultural Water Resources Management, Postgraduate Faculty, University of Jember.

References
[1] I Indarto 2013 Analisis Geostatistik (Yogyakarta: Graha Ilmu)
[2] I Indarto 2013 Variabilitas Spasial Hujan Harian di Jawa Timur J Tek Sipil vol 20 no 2 pp 107–120
[3] M J de Smith, M F Goodchild, P A Longley and Associates 2018 Geospatial Analysis A Comprehensive Guide to principles Techniques and Software Tools Sixth Edit Matador Leiceister UK: Retrieved from [www.spatialanalysisonlinecom]
[4] K Johnston, J M Ver Hoef, K Krivoruchko and N Lucas 2003 ArcGIS 9 Using ArcGIS Geostatistical Analyst New York: ESRI
[5] R Haining, S Wise and J Ma 1998 Exploratory spatial data analysis in a geographic information system environment Stat vol 47 no 3 pp 457–469
[6] U Haberlandt 2007 Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event J Hydrol vol 332 no 1–2 pp 144–157
[7] M Ellouze, C Azri and H Abida 2009 Spatial variability of monthly and annual rainfall data
over Southern Tunisia Atmos Res vol 93 no 4 pp 832–839
[8] S A Moges, B F Alemaw, T R Chaoka and R K Kachroo 2007 Rainfall interpolation using a remote sensing CCD data in a tropical basil - A GIS and geostatistical application Phys Chem Earth vol 32 pp 976–983
[9] P G Jones and P K Thornton 1999 Fitting a third-order Markov rainfall model to interpolated climate surfaces Agric For Meteorol vol 97 no 3 pp 213–231
[10] M Coulibaly and S Becker 2007 Spatial Interpolation of Annual Precipitation in South Africa-Comparison and Evaluation of Methods Water Int vol 32 no 3 pp 494–502
[11] D G A Shiddiq, Indarto, S Wahyuningsih and Askin 2018 aplikasi Histogram untuk analisis variabilities temporal dan spasial hujan bulanan : studi di wilayah UPT PSDA di Pasuruan Jawa Timur Teknol Pertan Andalas vol 22 no 1 pp 1–12
[12] M F Hutchinson 1995 Interpolating mean rainfall using thin plate smoothing splines Int J Geogr Inf Syst vol 9 no 4 pp 385–403
[13] S Naoum and I Tsanis 2004 A hydro informatic approach to assess interpolation techniques in high spatial and temporal resolution Can Water Resour J vol 29 no August 2003 pp 23–45
[14] J Pasaribu and N Haryani 2012 Perbandingan Teknik Interpolasi DEM SRTM dengan Metode Inversr Distance Weighted (IDW) Natural Neighbor and Spline (Comparison Of DEM SRTM Interpolation Techniques Using Inverse Distance Weighted (IDW) Natural Neighbor and Spline Method) J Penginderaan Jauh vol 9 no 2 pp 126–139
[15] I Indarto 2011 Tutorial: Exploratory Spatial Data Analysis (ESDA) Menggunakan ArcGis Geostatistical Analyst (Jember: Universitas Jember)
[16] L Li and P Revesz 2004 Interpolation methods for spatio-temporal geographic data Comput Environ Urban Syst vol 28 no 3 pp 201–227
[17] M L Segond, N Neokleous, C Makropoulos, C Onof and C Maksimovic 2007 Simulation and spatio-temporal disaggregation of multi-site rainfall data for urban drainage applications Hydrol Sci J vol 52 no 5 pp 917–935
[18] E Respatti, R Goejantoro and S Wahyuningsih 2014 Perbandingan Metode Ordinary Kriging dan Inverse Distance Weighted untuk Estimasi Elevasi Pada Data Topografi ( Studi Kasus : Topografi Wilayah FMIPA Universitas Mulawarman ) Comparison of Ordinary Kriging and Inverse Distance Weighted Methods for Estimation geographical data Springerplus vol 5 no 2 pp 163–170
[19] Z Arétouyap, P Njandjock Nouck, R Nouayou, F E Ghomsi Kemgang, A D Piépi Toko and J Asfahani 2016 Lessening the adverse effect of the semivariogram model selection on an interpolative survey using kriging technique Springerplus vol 5 no 1 pp 1–11
[20] J Arfaini and H H Handayani 2016 Analisa Data Foto Udara untuk DEM dengan Metode TIN IDW dan Krigging J Tek ITS vol 5 no 2 pp 2–7
[21] M R Mahmud, M Hashim, H Matsuyama, S Numata and T Hosaka 2018 Spatial downscaling of satellite precipitation data in humid tropics using a site-specific seasonal coefficient Water (Switzerland) vol 10 no 4 pp 1–19