Spatial Extreme Models with Copula to Determine Extreme Rainfall Zone

Achi Rinaldi
Department of Mathematics Education, State University of Islamic Studies
Raden Intan Lampung, Indonesia

Anik Djuraidah* and Aji Hamim Wigena
Department of Statistics, Bogor Agricultural University, Indonesia
*Corresponding author

I Wayan Mangku
Department of Mathematics, Bogor Agricultural University, Indonesia

Dodo Gunawan
Meteorological, Climatological, and Geophysical Agency, Indonesia

Abstract

Extreme rainfall zone is needed to determine of vulnerable areas. Unfortunately there is no any study about it until now. The existing zone is only limited to the average rainfall so it is not suitable for extreme rainfall. This study aims to establish the zone of extreme rainfall, especially in West Java. The distribution used is Generalized Extreme Value (GEV) distribution. Zones were built based on the extreme spatial parameter estimation using copula. The zone boundaries were based on contour plot of location parameter estimation using regression splines, thin plate spline and kriging. Spline regression and thin plate spline resulted 19 zones, while kriging resulted 20 zones. Extreme rainfall zones using thin plate spline were the best terms of grouping and image smoothness.
**Keywords**: copula, extremal coefficient, extreme rainfall, F-madogram, GEV, kriging, spline regression, thin plate spline

1 **Introduction**

In general, the analysis of rainfall is presented in the region with the similarity of rainfall measurement. The concept of rainfall zoning in Indonesia aims to simplify the analysis due to various constraints such as the limited number of rain gauge stations in some locations and missing data due to various reasons. To overcome these problems, rainfall zone was created.

Meteorological, Climatological, and Geophysical Agency (BMKG) has its own rainfall zone called Zona Musim (ZOM). The ZOM is an area of average rainfall patterns distinctly different between the dry season and the rainy season [2]. The ZOM were created using cluster method [8]. The same method was also used by Saputro (2012) [11] to create different ZOMs. However, in both studies ZOM boundaries were determined subjectively and the ZOMs are not for extreme rainfall.

The impact of extreme rainfall is very harmful to humans, such as the floods that can destroy a wide range of public facilities even loss of life, so it is important to study on the zoning of extreme rainfall. Unfortunately, until now there has been no study on the zoning of extreme rainfall. This study aims to establish the zone of extreme rainfall that is expected to determine the disaster-prone areas due to the impact of extreme rainfall.

The determination of the extreme rainfall zones based on location parameter estimator of Generalized Extreme Value (GEV) distribution, while we use spatial extreme models which has proven to give good performance [7]. Spatial extreme dependency was measured using F-madogram plot and extremal coefficient. Determination of the territorial boundaries of the zone through a contour plot that was formed through the spline regression methods, kriging, and thin plate spline.

2 **Literature Review**

This chapter describes some theories associated with the determination of extreme rainfall zones. Extreme rainfall identification in this study is using Block Maxima (BM) so the distribution used is the of Generalized Extreme Value (GEV). F-madogram and extremal coefficient are used for spatial extreme dependencies. Copula is used to estimate spatial extreme parameter. The method of contour for determining the boundaries of extreme rainfall zone were generated through spline regression, thin plate spline and kriging.

2.1. Generalized Extreme Value Distribution

Generalized Extreme Value (GEV) distribution can be expressed as [3]:

$$F(x) = \begin{cases} 
1 - \left(1 + \frac{\lambda (x-\mu)}{\sigma}\right)^{-\frac{1}{\lambda}}, & \text{if } \lambda \neq 0, \\
\exp\left(-\frac{x-\mu}{\sigma}\right), & \text{if } \lambda = 0,
\end{cases}$$

where $\mu$ is the location parameter, $\sigma$ is the scale parameter, $\lambda$ is the shape parameter, and $F(x)$ is the cumulative distribution function.
\[ G(y) = \begin{cases} \exp\left\{ -\left(1 + \frac{y - \mu}{\sigma}\right)^{-1/\xi}\right\}, & \xi \neq 0 \\ \exp\left\{ -\exp\left(-\frac{y - \mu}{\sigma}\right)\right\}, & \xi = 0 \end{cases} \] (1)

with \( \mu \) is the location parameter, \( \sigma \) is the scale parameter, and \( \xi \) is the shape parameter.

### 2.2. F-madogram and Extremal Coefficient

Extreme value dependencies can be measured using the extremal coefficient, while the extreme values that contain spatial element can be predicted using madogram. However, madogram needs the first moment finite that does not always happen in extreme cases, so we used F-Madogram. If \( Z(x) \) is max-stable process that is stationary and isotropic with \( G \) distribution function, then F-madogram can be written as [4]:

\[ v_F(h) = \frac{1}{2} E \left| G(Z(y+h)) - G(Z(y)) \right| \] (2)

with \( v_F(h) \) is the value of F-madogram for any distance \( h \) and \( Z(y+h) - Z(y) \) is the difference between the value of the variable, for example precipitation from two locations within \( h \) distance. The relationship between F-madogram and extremal coefficient \( (\theta) \) which is a measure of extreme dependency can be written as follows [4]:

\[ \theta(h) = \left(1 + 2v_F(h)\right)\left(1 - 2v_F(h)\right)^{-1}. \] (3)

### 2.3. Parameter Estimation of Spatial Extreme Model

Spatial extreme models base on the model of Davison et al. [7] which describe the marginal parameters into the form:

\[ \mu_i = \beta_0 + \beta_1 \text{lon}_i + \beta_2 \text{lat}_i \] (4)

\[ \sigma_i = \beta_0 + \beta_1 \text{lon}_i + \beta_2 \text{lat}_i \] (5)

with \( \text{lon}_i \) is a longitude coordinates and \( \text{lat}_i \) is latitude coordinates of the \( i \)-rain gauge station. The estimation for the shape parameter \( \xi \) is difficult [5], so this parameter is assumed to be constant. In this study we limited only to estimate the location parameter \( \mu \).

Parameter estimation is using copula which can be done using the pseudo maximum likelihood estimation (PMLE) [13]. PMLE transform the original data into a pseudo observation followed by maximum likelihood estimators. For example \( X_{ij} \) is a sample of size \( d \), \( i = 1, \ldots, n \) and \( j = 1, \ldots, d \). By using \( G \) distribution function of observation pseudo \( \mathbb{U} = U_{ij} \) shown by [9]:

\[ U_{ij} = \frac{n}{n+1} G_j(X_{ij}). \] (6)

PMLE \( \hat{\theta}_n^{\text{PML}} \) calculated by observing \( \mathbb{U} \) to maximize:

\[ \ell_U(\theta) = \sum_{i=1}^n \log c(U_{i1}, \ldots, U_{id}|\theta) \] (7)
with \( \theta = (\mu, \sigma, \xi) \) is the vector of copula parameters, and \( c \) is a probability function of parameter \( \theta \) which is given by:

\[
c(u_1, \ldots, u_d | \theta) = \frac{\partial C(u_1, \ldots, u_d | \theta)}{\partial u_1, \ldots, u_d}, \ u_1, \ldots, u_d \in [0,1]
\]  

(8)

PMLE is given by:

\[
\hat{\theta}^{\text{PML}}_n(\mathbb{U}) = \arg\max_{\theta} \ell_\mathbb{U}(\theta).
\]  

(9)

2.4. Extreme Rainfall Contour Plot

One way to do zoning or grouping of rainfall is contour plot. We focused on the manufacture of contour plot with GEV location parameter that will become a zone of extreme rainfall. We conducted using three methods to make contour plot i.e. spline regression, kriging, and thin plate spline. We used coordinates of longitude and latitude as explanatory variables with additive and interaction form in spline regression [12], where the equation can be written as:

\[
\hat{\mu}_i = f_1(\text{lon}_i) + f_2(\text{lat}_i) + f_3(\text{lon}_i, \text{lat}_i)
\]  

(10)

\( f_1, f_2 \) and \( f_3 \) are a smoothing spline, while the parameter estimator \( \hat{\mu}_i \) was estimated first with copula method.

Thin plate spline is a spline method for the multi dimension which is also known as spline-d, with \( d \) as the dimensions of \( t_i \in \mathbb{R}^d \). In the two-dimensional, the spline basis function is \( C(r) = r^2 \log r \), while the function of spline-2 at the location \( t = (x_1, x_2) \) is defined as:

\[
\hat{\mu}(t) = \beta_0 + \beta_1 t + \sum_{k=1}^K u_k C(||t - \mathbf{k}_k||)
\]  

(11)

with \( \beta = (\beta_0, \beta_{11}, \beta_{12}, u_1, \ldots, u_K) \) is a vector of parameters and \( k_1 < \cdots < k_K \) are knots in \( \mathbb{R}^2 \), \( ||r|| = ||t - \mathbf{k}_k|| \) is the distance between \( t \) and knot \( \mathbf{k}_k \) [10]. In this study \((x_1, x_2)\) is the location of rain station with coordinates \( x_1 = \text{longitude} \) and \( x_2 = \text{latitude} \), while \( \hat{\mu} \) is the location parameter estimator.

Kriging is a method for interpolation of variable value at some locations based on observing certain similar data in other locations. The estimated value obtained by the formula is as follows [6]:

\[
\hat{Z}_p = \sum_{i=1}^n \lambda_i Z_i
\]  

(12)

with \( \hat{Z}_p \) is the estimated value of \( Z \) variable at the \( p \)-point, \( Z_i \) is the value of \( Z \) variable at the \( i \)-point, \( \lambda_i \) is the weight on \( i \)-point. We used \( Z_i = \hat{\mu}_i \) as a location parameter estimator at \( i \)-point and \( \hat{Z}_p = \hat{\mu}_p \) as a location parameter estimator at \( p \)-point.

2.5. Goodness of Fit Test

There are two types goodness of fit test used to know the suitability of the group that has been formed through a contour plot. The first goodness of fit test is goodness of fit measure variance (GVF), while the second one is tabular accuracy index (TAI) which is defined as [1]:
Spatial extreme models with copula

\[ GVF = 1 - \frac{\sum_{j=1}^{k} \sum_{i=1}^{N_j} (z_{ij} - \bar{z}_j)^2}{\sum_{i=1}^{N} (z_i - \bar{z})^2} \]  
(13)

\[ TAI = 1 - \frac{\sum_{j=1}^{k} \sum_{i=1}^{N_j} |z_{ij} - \bar{z}_j|}{\sum_{i=1}^{N} |z_i - \bar{z}|} \]  
(14)

with:
- \( z_{ij} \) = \( i \)-th observation in \( j \)-th class, \( 1 < j < k \)
- \( z_i \) = \( i \)-th observation
- \( \bar{z}_j \) = \( j \)-th class mean, \( 1 < j < k \)
- \( \bar{z} \) = mean of all observations.

The values of GVF and TAI are between 0 and 1. If the values of GVF and TAI are close to 1, then the grouping method is getting better.

3 Data and Methodology

3.1. Data

The data are monthly rainfall data (1981-2012) of 69 stations in West Java. Data from Meteorological, Climatological, and Geophysical Agency (BMKG).

3.2. Methodology

The steps of this study are:
1. Identify extreme rainfall data using Block Maxima (BM) with GEV distribution.
2. Estimate spatial dependencies of extreme rainfall using F-madogram plot and extremal coefficient plot.
3. Estimate the location parameter estimators with copula method.
4. Make a contour plot with spline regression, kriging, and thin plate spline.
5. Calculate goodness of fit test of zoning group.

4 Results and Discussion

4.1. Data Exploration

The maximum monthly rainfall values of 69 rain gauge stations shown in Figure 1 shows that the maximum precipitation for each station varies greatly. Although the values used are the maximum value, it turns out that the majority of rainfall stations still have outliers at the top of the diagram grid lines on boxes with longer lines. It shows that the distribution of the data tend to protrude toward the right (extreme right), which means rainfall likely to far exceed the average value.

4.2. Spatial Extreme Estimator and Dependencies

Spatial dependencies of the maximum monthly rainfall can be seen through the plot between the semivarian \( v(h) \) and the \( (h) \) distance, which is known as the
F-madogram plot. While extreme spatial dependencies can be seen from the plot of extremal coefficient \( \theta(h) \). F-madogram (Figure 2a) shows the pattern follows a model that is ideal semivarian. The model shows that increasing the distance of the station also increase the value of the diversity of rainfall.

Figure 1 Box plot Maximum Monthly Rainfall in 69 stations in 1981-2014

Figure 2 Plot F-madogram (a) and extremal coefficient (b)

Extremal coefficient values are approaching 1 which indicates the extreme dependencies (Figure 2b). Extremal coefficient obtained through the plot between the coefficient extremal \( \theta(h) \) and the distance \( (h) \). The F-madogram plot and the extremal coefficient plot indicates an extreme spatial dependencies of the maximum monthly rainfall in the West Java.
The result of the location parameter estimators for the maximum rainfall year period 1981-2012 can be seen in Figure 3 which shows the value of $\hat{\mu}_i$ diverse. The smallest value of location parameter estimator can be seen at UPTD Serpong rain gauge station while the highest at Wanayasa rain gauge station. Because of the diversity of its value, this parameter can be used as the basis of the establishment of the extreme rainfall zones.

![Figure 3 Histogram of Location Parameter Estimator ($\hat{\mu}_i$)](image)

### 4.1 Contour Plot of Spatial Extreme Parameters

Contour formed by methods of spline regression and thin plate spline can be grouped into 19 zones, while kriging grouped into 20 zones, which can be seen in Figure 4, 5 and 6. From the contour plot for methods of spline regression and thin plate spline produce 6 zones with parameter highest extreme locations with rainfall above 500 mm, while kriging has 7 zones. Zone with the highest value of extreme parameter estimators provide relevant results when compared with the data of extreme events, such as floods. These regions do have a higher frequency of flooding compared to other areas, as well as having high contribution to the neighboring regions such as in the case of Jakarta flooding that caused by heavy rainfall which actually occurred in Bogor (zone 1).
Figure 4 Contour Plot using Spline Regression

Figure 5 Contour Plot using Kriging

Figure 6 Contour Plot using Thin Plate Spline
The final stage is calculating the goodness of fit test of a well-formed group from the contours of the spline regression, kriging, and thin plate spline. Goodness of fit test results can be seen in Table 1. From Table 1 it appears that the goodness of fit test of thin plate spline give better results than kriging and spline regression. In general, the thin plate spline gives the best results in terms of grouping and image smoothness.

<table>
<thead>
<tr>
<th>Model</th>
<th>GVF</th>
<th>TAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline Regression</td>
<td>0.746836794</td>
<td>0.56711244</td>
</tr>
<tr>
<td>Kriging</td>
<td>0.789641209</td>
<td>0.59244830</td>
</tr>
<tr>
<td>Thin Plate Spline</td>
<td>0.792902766</td>
<td>0.59729796</td>
</tr>
</tbody>
</table>

5 Conclusion

Extreme rainfall zone formed with spline regression and thin plate spline are 19 zones, while kriging resulted in 20 zones. The zoning of thin plate spline give better results than kriging and spline regression when viewed from the goodness of fit test. We recomend for extreme rainfall zone by thin plate spline because it gives the best results in terms of grouping and image smoothness.

Acknowledgements. The author would like to thank the Meteorological, Climatological, and Geophysical Agency (BMKG) for cooperation in obtaining research data.

References


Received: April 20, 2017; Published: May 15, 2017