Extreme rainfall prediction using spatial extreme value by Max Stable Process (MSP) Smith model approach

H Yasin¹, A R Hakim¹, B Warsito¹ and R Santoso¹
¹Department of Statistics, Diponegoro University
Jl. Prof. Soedharto, SH, Tembalang, Semarang 50275, Indonesia
E-mail: hasbiyasin@live.undip.ac.id

Abstract. This research covers Spatial Extreme Value method application with Max-Stable Process (MSP) approach that will be used to analysis Extreme Rainfall in Semarang city. Extreme value sample are selected by Block Maxima methods, it will be estimated into Spatial Extreme Value form by including location factors. Then it transform to Frechet distribution because it has a Heavy tail pattern. Max Stable Process (MSP) is an extension of the multivariate extreme value distribution into infinite dimension of the Extreme Value Theory. MSP using Generalized Extreme Value (GEV) approach, it has three forms distribution that is Gumbel, Frechet and Weibull. Extreme Value Samples that are selected by Block Maxima will follow GEV distribution. The best model that obtained based on Max-Stable with Smith model can be used to predict Extreme Rainfall by count Return Level. Then can be further analysis of various aspects, so it can provide a suggestion about the potential of natural disasters which are caused by extreme rainfall. Extreme rainfall prediction using the Smith model in the return period of the next two years, at the Semarang City Climatology, Tanjung Mas and Ahmad Yani Station is predicted to be a maximum of 100.7539 mm, 100.1052 mm, and 109.9379 mm.

1. Introduction
Rain intensity is quite high in recent times resulting in several regions in Indonesia under siege with the threat of flooding, even some areas have been flooded due to the resulting high rainfall intensity. According to the climate analysis from the BMKG Class I Climatology Station in Semarang City, current rainfall (as of October 2017) has reached the range of 60-70 ml per hour [1]. The city of Semarang is the capital of the province of Central Java, which is the centre of the economy that connects between regions in Central Java. In the event of a disaster caused by extreme weather will cause the access to economic access to various regions to be interrupted and result in government activities being hampered, the emergence of various health problems, food security, hampered activities related to the economy and industry, and will adversely affect damage to infrastructure both private and government.

Several studies have been conducted to predict extreme rainfall, especially in Indonesia, including research on extreme rainfall in Ngawi District [2] and research on extreme rainfall in Lamongan District using schlather model [3]. In these studies, there are two approaches used to determine extreme values, namely Block Maxima (BM) and Peaks-Over-Threshold (POT). The BM approach produces extreme value distributions in the form of Generalized Extreme Value (GEV) distributions. Estimation methods of GEV distribution parameters include Maximum Likelihood (ML) and Least Square (LS). The study also discusses the existence of dependency cases between extreme data observations which are often called stochastic extreme data.
Extreme Value Theory (EVT) is a statistical method to identify extreme events. EVT is developed from univariate cases with extreme events on one variable. For data on rainfall, snow, river discharge, temperature, heat or temperature including spatial data which is a multivariate data because it was observed in several locations, therefore the Spatial Extreme Value (SEV) method was developed.

The method for analyzing extreme events with SEV, one of them is the max-stable approach to rainfall data [4, 5]. In this study SEV modelling will be carried out using the Max-Stable Process (MSP) approach. The model that will be used is the Smith model. One of the most important in the EVT study is determining the return level, which is the maximum rainfall threshold value that is expected to be exceeded once in a certain return period. The model that will be used is the Smith model. One of the most important in the EVT study is determining the return level, which is the maximum rainfall threshold value that is expected to be exceeded once in a certain return period. Return level can be used as information so that impacts caused by extreme climates can be minimized.

2. Literature Review

2.1. Extreme Value Theory

The tail behavior of a random distribution can be learned using a method called Extreme Value Theory (EVT). The approach used to identify extreme values, namely the Block Maxima Method (BM) by taking maximum values in a period or block[6]. Block maxima scheme as shown in figure 1.

The maximum value in each block will be used as a sample of extreme values, with $y_1, y_2, y_3$ being the extreme value for each block. According to [6], the Block Maxima method applies Fisher-Tippet Gnedenko's theorem that sample data of extreme values taken from this method will follow the distribution of Generalized Extreme Value (GEV) which has Cumulative Distribution Function (CDF) as equation (1).

$$F(y; \mu, \sigma, \xi) = \begin{cases} \exp \left\{ -\left[1 + \frac{y - \mu}{\sigma} \right]^{-\frac{1}{\xi}} \right\}, & -\infty < y < \infty, \xi \neq 0, 1 + \frac{y - \mu}{\sigma} > 0 \\ \exp \left\{ -\exp \left( -\frac{y - \mu}{\sigma} \right) \right\}, & -\infty < y < \infty, \xi = 0, \end{cases}$$

(1)
Where \( y \) is extreme values obtained by Block Maxima, \( \mu \) is location parameter with \(-\infty < \mu < \infty\), is scale parameter with \( \sigma > 0 \), is shape parameter.

### 2.2. Max Stable process

The Max Stable Process (MSP) is an extension of the multivariate extreme value distribution to the infinite dimensional dimensions of Extreme Value Theory where samples are taken from maximum values (maxima) at each location (spatial process) [7]. An arbitrary distribution function \( G \) is said to be max stable if and only if \( G \) is GEV distributed. Suppose that \( \{Y_i(s)\}_{s \in S}, i = 1, 2, \ldots, n \) where \( n \) is an independent replication of a stochastic process in the set \( S \). Assume there is a continuous function where

\[
(\xi(s))_{s \in S} \quad \xi(s) = \frac{Y_i(s) - b_i(s)}{a_i(s)}, \quad n \to \infty, s \in S
\]

Where \( Y_i, \ldots, Y_n \) are independent replication of \( Y \), if the limit \( Y(s) \) is exist, then the limit process \( Y(s) \) called Max-Stable process. If \( a_n(s) = n \) and \( b_n(s) = 0 \) then \( Y(s) \) is also called Simple Max Stable Process [6]. Assuming that each component in each location has a GEV distribution, then transformed into a marginal Frechet unit:

\[
F(z) = \exp\left(\frac{-1}{z}\right), \quad z > 0
\]

This process can be obtained by standardizing \( \{Y(s)\}_{s \in S} \) so that it is obtained:

\[
\left\{Z(s)\right\}_{s \in S} = \left\{1 + \frac{\xi(s)(Y(s) - \mu(s))}{\lambda(s)}\right\}^{1/\xi(s)}, \quad s \in S
\]

where \( \mu(s), \xi(s) \) and \( \lambda(s) > 0 \) are the continuous functions. This \( Z \) process is still included in the max-stable process [6]. In general, the Max Stable Process with the marginal Frechet unit can be explained by the equation (5):

\[
Z(s) := \max_{j \geq 1} \left\{U_j W_j(s)\right\}, \quad s \in S,
\]

Where \( W_j \) and \( U_j \) are Poisson process at \((0, +\infty) \times \mathbb{R^2}\) with measurement intensity \( v(dx) \times u^2 du \). From the general model, it was developed into a max-stable model, one of which was Smith's model.

### 2.3. Smith model

Smith's Storm Profile or Smith's model is proposed by [8, 9], where \( W_j(s) = f(X_j - s) \) so the equation (5) becomes:

\[
Z(s) = \max_{j \geq 1} \left\{U_j f(X_j - s)\right\}, \quad s \in S
\]

With \( U \) representing the magnitude of the storm, \( X \) is the center, and \( f \) is the shape parameter. The bivariate cumulative distribution function (CDF) of Smith model is:

\[
F(z_1, z_2) = \exp\left[ -\frac{1}{z_1} \Phi\left( \frac{a(h)}{2} + \frac{1}{a(h)} \log \frac{z_1}{z_j} \right) \right]
- \frac{1}{z_j} \Phi\left( \frac{a(h)}{2} + \frac{1}{a(h)} \log \frac{z_i}{z_j} \right)
\]

(7)

Bivariate probability density function (PDF) of Smith Model is:
To estimate the parameters of the Smith model, the Maximum Pairwise Likelihood Estimation (MPLE) method is used, where $\Phi$ is the CDF of standard normal of $s_j$ and $s_{j'}$, $T_j$ is the CDF of standard normal of $s_{j'}$ and $s_j$, and $\alpha(h) = \frac{1}{2} + \frac{\log(z_k/\sqrt{z_j})}{a(h)}$ and $\nu(h) = \alpha(h) + w(h)$.

2.4. Selection The Best Model
After the trend surface model was obtained, the best model was chosen by Takeuchi Information Criterion (TIC). Trend surface models are:

For $s_{j}$, $s_{j'}$, $n = (s_j - s_j)^T$ and $\alpha(h) = (h^T \Sigma^{-1} h)^{1/2}$.

2.4. Selection The Best Model
After the trend surface model was obtained, the best model was chosen by Takeuchi Information Criterion (TIC). Trend surface models are:

$$
\hat{\mu}(s) = \beta_{\mu,0} + \beta_{\mu,1} \ln(s) + \beta_{\mu,2} \text{lat}(s)
$$

$$
\hat{\sigma}(s) = \beta_{\sigma,0} + \beta_{\sigma,1} \ln(s) + \beta_{\sigma,2} \text{lat}(s)
$$

According to [10] proposed using the composite likelihood information criteria developed into (TIC). Then the extreme rainfall prediction uses the concept of return level.

3. Research Methodology
3.1 Research Variables
The data used in this study is secondary data obtained from the Meteorological, Climatological and Geophysical Agency (BMKG). The data used is daily rainfall data in several locations of the Rainfall Observation Station in the City of Semarang, Central Java in 1990 to 2017 period. The variables used in the study were daily rainfall taken from all rain monitoring stations in Semarang City, namely: Measurement Station Semarang Climatology, Rain Monitoring Station Ahmad Yani and Tanjung Mas Monitoring Station.

3.2. Data Analysis Method
1. Perform descriptive analysis and create bar charts
2. Identify the distribution of rainfall data to find out the distribution of heavy tail data and extreme values with histograms
3. The spatial extreme value modelling procedure uses the Max Stable Process against extreme rainfall data in Semarang City based on each rainfall measurement post location.
4. Predict extreme rainfall for each location by determining the return level value from rainfall data in Semarang City.

4. Results and Discussion
In this section, the steps in working on MSP are explained, starting from the stages of taking extreme samples to getting return levels.
a. Extracting extreme samples by the Block Maxima method

INPUT : Data $x_1, x_2, ..., x_y$

OUTPUT : Extreme Value $y_1, y_2, ..., y_n$

Step 1 Form a block based on the time period (T)
Step 2 Take extreme samples (y) in each block with maximum value criteria
b. Obtain parameter estimates for $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ univariate on each rain monitoring post (s) with MLE and solve numerically by the Nelder-Mead iteration method.

**INPUT** : Extreme Value $y_1, y_2, \ldots, y_n$

**OUTPUT** : Parameter $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$

Step 1: Forms the likelihood function from pdf with the GEV distribution

Step 2: Form ln likelihood function from step 1

Step 3: Obtain the first derivative of the ln likelihood function from each parameter $\mu, \sigma, \xi$

Step 4: If the result is not closed form, then it is solved using numerical method with Broyden-Fletcher-Goldfarb-Shanno (BFGS) iteration method

c. Test the suitability of extreme sample distribution to the distribution of Generalized Extreme Value (GEV) with Anderson Darling test.

d. Transformation of extreme rainfall data to the Frechet distribution. $s_i$ shows the location of the i rainfall observation station.

**INPUT** : Extreme Value $y_1, y_2, \ldots, y_n$, $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$

**OUTPUT** : $Z(s_i)$

Step 1: Calculate $Z(s_i)$ by equation (4)

Step 2: Get the $Z(s_i)$

e. Calculates spatial dependencies using extreme coefficients

**INPUT** : $Z(s_i)$

**OUTPUT** : Extreme Coefficients $\theta(s_j - s_k)$ of Smith Model

Step 1: Calculate the $\theta(s_j - s_k)$

$$\theta(s_j - s_k) = 2\Phi \left( \frac{\sqrt{(s_j - s_k)^2 \Sigma^{-1}(s_j - s_k)}}{2} \right)$$

Step 2: If $1 \leq \theta(h) \leq 2$ then there are dependencies. If $\theta(h) = 1$ then full dependency occurs and if $\theta(h) = 2$ then there are independent.

f. Parameter Estimates of Spatial GEV: $GEV \left( \hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s) \right)$

**INPUT** : $Z(s_i)$, longitude (u) and latitude (v)

**OUTPUT** : The best combination of $\hat{\mu}(s)$, $\hat{\sigma}(s)$ and $\hat{\xi}(s)$

Step 1: Making a trend surface combination that follows a regression model like the example in equation (9). The combination of models starts from the parameter model with only one explanatory variable to the parameter model with the quadratic model. Parameter $\hat{\xi}(s)$ assumed constant and parameter $\beta$ was estimated by MPLE of pdf GEV.

Step 2: Get the model of trend surface $l$

g. Max-Stable parameter estimation with Smith model. The best combination in step f is used to estimate the following model parameters:

**INPUT** : $Z(s_i)$, longitude (u) and latitude (v) of each location, the best combination of $\hat{\mu}(s), \hat{\sigma}(s)$ and $\hat{\xi}(s)$

**OUTPUT** : $\hat{\mu}(s), \hat{\sigma}(s)$ and $\hat{\xi}(s)$

Step 1: Smith model parameter estimates were obtained from Smith’s PDF model estimated by the MPLE method. The first derivative of each parameter estimated by MPLE can
be ascertained to have an un-closed form, so to solve these equations using the Broyden-Fletcher-Goldfarb-Shanno iteration method (BFGS).

Step 2 Get the model of trend surface

h. Calculate the Return prediction

INPUT : P years of prediction, \( \hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s) \) of Smith Model

OUTPUT : Return level \( z_p(s) \) Smith model

Step 1 Determine the return period (T)
\[ T = P \text{ years} \times 4 \text{ (number of block)} \]

Step 2 Determine the Return Level.

As a preliminary information, the characteristics of rainfall patterns from the data used show the range of daily rainfall in Semarang City ranging from 5.665 to 6.197 mm/day. The highest average daily rainfall is 6.197 mm/day in the Ahmad Yani rain monitoring station, while the lowest average rainfall is at Tanjung Emas rain monitoring station which is 5.665 mm/day. When viewed in general, the highest rainfall is 276 mm which is monitored at Semarang climatology station. Completely presented in table 1.

| Location                              | Mean  | St.Dev | Minimum | Maximum |
|---------------------------------------|-------|--------|---------|---------|
| Semarang Climatology Station          | 6.129 | 15.225 | 0       | 276.0   |
| Ahmad Yani Rain Monitoring Station   | 6.197 | 15.204 | 0       | 255.3   |
| Tanjung Mas Rain Monitoring Station  | 5.665 | 14.217 | 0       | 246.6   |

The heavy tail pattern can be seen in the three Semarang city rain monitoring stations in Figure 2 based on the daily data, where the tail of distribution is slow down, which means that the rainfall data at that location contains extreme values. Extreme rainfall data used, formed from block maxima. Classification of block determination using three-month blocks December, January, February (DJF) - March, April, May (MAM) - June, July, August (JJA) - September, October, November (SON). This uses the BMKG reference which classifies monsoon rainfall patterns in most regions of Java. The trend surface model obtained from the max-stable smith model is as follows:

\[ \hat{\mu}(s) = -3475.87 + 28.66 \text{ longitude}(s) - 44.96 \text{ latitude}(s) \]

\[ \hat{\sigma}(s) = -4795.26 + 39.58 \text{ longitude}(s) - 61.32 \text{ latitude}(s) \]

\[ \hat{\xi}(s) = 0.9476 \]

Then obtained parameter estimates that differ for each location, where parameter \( \hat{\xi}(s) \) are constant in each location, shown in table 2:
To calculate estimates of extreme rainfall predictions based on a certain time period, the return level is used. The estimation results of parameter $\hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s)$ on each model are then used to calculate the predicted return level of extreme rainfall at the three locations of Semarang city rain monitoring stations. In this study, the time periods are 2, 4, 8 and 10 years. The return level calculation requires that the return periods are inversely proportional to probability value, $T = \frac{1}{p}$ in other word that $p = \frac{1}{T}$, where the maximum probability is $p = 1$. Return level predictions, can only be done starting from the second year period. The results of the return level calculation are shown in table 3.

**Table 3. Return level prediction**

| Rain Monitoring Post                   | Return level (mm) |
|---------------------------------------|-------------------|
|                                       | 2 years | 4 years | 8 years | 10 years |
| Semarang Climatology Station          | 100.7539 | 116.4993 | 131.3712 | 136.0251 |
| Ahmad Yani Rain Monitoring Station    | 109.9379 | 132.6133 | 155.4454 | 162.8795 |
Extreme rainfall prediction using the Smith model in the return period of the next two years, at the Semarang City Climatology Station is predicted to be a maximum of 100.7539 mm. At the Tanjung Mas rain monitoring station is predicted to be a maximum of 100.1052 mm, while at the Ahmad Yani rain monitoring station is predicted to be a maximum of 109.9379 mm. Table 3 presents the return level predictions with the return periods of 2, 4, 8 and 10 years ahead and the probability exceeded are 0.125, 0.0625, 0.031 and 0.025 respectively.

5. Conclusion
Based on the procedure for calculating extreme rainfall with the max-stable smith model, the estimated parameter values are different for each location. The results of extreme rainfall prediction using the return period of the next two years, at the Semarang City Climatology Station is predicted to be a maximum of 100.7539 mm. At Tanjung Mas Rain Monitoring Station, it is predicted that the maximum is 100.102 mm, and at the Ahmad Yani rain monitoring station is predicted to be a maximum of 109.9379 mm. Prediction of extreme rainfall return levels can be used to calculate the return period until T time to come. Future research in spatial extrem value can also use other max-stable models namely Schlater and Brown Resnick models, but can also be developed using other approaches, namely copula and hierarchical Bayesian.

Acknowledgments
We would like to give thanks to Diponegoro University for their support. This research was funded by “RPP” Research Grant 2018.

References
[1] BMKG 2017 Jumpa pers perkembangan cuaca dan musim hujan 2017 http://www.bmkg.go.id/press-release/?p=jumpa-pers-perkembangan-cuaca-dan-musim-hujan-2017-2018&tag=press-release&lang=ID
[2] Hakim A R, Sutikno and Dedy D P 2016 The 3rd International Conference on Research, Implementation and Education of Mathematics and Sciences (ICRIEMS) (Yogyakarta: Indonesia).
[3] Malika R, Sutikno and Wahyu W 2015 Prosiding Simposium Nasional Inovasi dan Pembelajaran Sains (Institut Teknologi Bandung: Indonesia).
[4] Buishand T A, de Haan L and Zhou C 2008 The Annals of Applied Statistics 2 624.
[5] Davison A C, Padoan AS and Ribatet M 2012 Statistical Science 27 161.
[6] Gilli M and Kellezi E 2006 Computational Economics 27 1.
[7] Cooley D, Naveau P and Poncet P. 2006. Variograms for Spatial Max-Stable Random Fields. Lecture Notes in Statistics (New York: Springer) p 373
[8] De Haan L 1984 The Annals of Probability 12 1194.
[9] Smith R A 1990 Max-stable processes and spatial extremes (University of Surrey: England).
[10] Padoan S A, Ribatet M and Sisson S A 2010 Journal of the American Statistical Association 105 263.