Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 16(1) 2023: 59-66
http://ejournal.undip.ac.id/index.php/media_statistika

MAX-STABLE PROCESS WITH GEOMETRIC GAUSSIAN MODEL ON RAINFALL DATA IN SEMARANG CITY

Arief Rachman Hakim, Rukun Santoso, Hasbi Yasin, Masithoh Yessi Rochayani
Department of Statistics, Diponegoro University, Indonesia

e-mail: arief.rachman@live.undip.ac.id

DOI: 10.14710/medstat.16.1.59-66

Article Info:
Received: 21 July 2023
Accepted: 20 September 2023
Available Online: 4 October 2023

Keywords:
Rainfall; Geometric Gaussian; Max-stable Process; Spatial.

Abstract: Spatial extreme value (SEV) is a statistical technique for modeling extreme events at multiple locations with spatial dependencies between locations. High intensity rainfall can cause disasters such as floods and landslides. Rainfall modelling is needed as an early detection step. SEV was developed from the univariate Extreme Value Theory (EVT) method to become multivariate. This work uses the SEV approach, namely the Max-stable process, which is an extension of the multivariate EVT into infinite dimensions. There are 4 Max-stable process models, namely Smith, Schlater, Brown Resnik, and Geometric Gaussian, which have the Generalized Extreme Value (GEV) distribution. This study models extreme rainfall, using rainfall data in the city of Semarang. This research was carried out by modeling data using the Geometric Gaussian model. This method is developed from the Smith and Schlater model, so this model can get better modeling results than the previous model. The maximum extreme rainfall prediction results for the next two periods are Semarang climatology station 129.30 mm^3, Ahmad Yani 121.40 mm^3, and Tanjung Mas 111.00 mm^3. The result from this study can be used as an alternative for the government for early detection of the possibility of extreme rainfall.

1. INTRODUCTION

Rainfall is measured based on the height limit of rainwater that collects in a place with a flat cross-section, which does not decrease, infiltrate, and also does not flow (BMKG, 2014). Conditions are said to be extreme if the intensity of rainfall reaches 100 mm^3. High and unpredictable rainfall often results in floods, landslides, and crop failures. Semarang is one of the cities that has had a relatively high negative impact due to the occurrence of too high an intensity of rain. The north coast route which passes through the city of Semarang, has many industries which are one of the main drivers of the economy in the city of Semarang and the surrounding area. If a flood disaster occurs on the route, it will cut off access to the distribution of goods and services. This has a wide and large impact on the city of Semarang, this reason is enough to be the basis for the need to make a model that can study and predict rainfall in the city of Semarang. It is hoped that there are anticipatory steps that can be made, intending to minimize the negative impact. The branch of statistics that studies extreme phenomena and data is Extreme Value Theory (EVT).

Spatial Extreme Value (SEV) is a statistical tool for modeling extreme events across multiple locations with dependencies. Detailed studies on the use of spatial extreme have also been performed by Ribatet (2013), Davison (2015), and Huser (2016). This study uses SEV with a Max-stable process approach, which is an extension of multivariate EVT into an infinite dimension. One of the models utilized in the Max-stable procedure is the Smith model, which was introduced by Smith in 1990. This model was tested in England in cases of extreme rain intensity. Schlather in 2002 put forward the Schlather Model which is a SEV model, using a Max-stable process based on a Gaussian random field. Subsequent developments used the Brown-Resnick process principle proposed by Brown and Resnick in 1977, which became known as the Brown-Resnick Model (Kabluchko, 2009). The combination of the Smith and Schlather models produces a new model called the Geometric Gaussian Model introduced by Davison et al. (2010).

In general, all MSP models have Generalized Extreme Value (GEV) distribution. Gaussian geometric modeling is often used to model extreme events such as extreme heat, extreme snow and extreme rainfall in countries that have four seasons. The application of this method in Indonesia, which only has two seasons, is likely to be more effective because the rainy season period is longer and the pattern of rainy season data can be better captured by the SEV method. This study uses rainfall data in the city of Semarang, then extreme data are chosen using block maxima. The fundamental idea of the block maxima is that the extreme data is chosen from the maximum data in each time periods block. In the next phase, the extreme data are modeled using geometric Gaussian models to determine parameter estimates. Hakim's research (2021) concerning the Geometric Gaussian Model on sea wave height data, found that the performance of this method is quite good. This method was developed from the Smith and Schlater model. Return level in the Geometric Gaussian Model is used to estimate the amount of extreme rainfall.

2. MAX-STABLE PROCESS (MSP)

2.1. Definition

Suppose the set of indexes $\{Y_i(x)\}_{x\in X}, i = 1, 2, \ldots, n$ are n independent replications of a continuous stochastic process. Suppose that given a continuous function where $a_n(x) > 0$ and $b_n(x) \in R$ so that:

$$Y(x) = \lim_{n\to\infty} \frac{\max_{i=1}^{n} Y_i(x) - b_n(x)}{a_n(x)}; n \to \infty, x \in X \qquad (1)$$

the limit process $Y(x)$ is said to be a Max-Stable Process, if in equation 1 the limit value exists (de Haan, 1984). If $a_n(s) = n$, $b_n(s) = 0$ then $Y(x)$ is also a simple MSP. Max-Stable Process has two main properties. First, the marginal distribution in dimension one follows the Generalized Extreme Value distribution $Y \sim GEV(\mu, \sigma, \xi)$, with the probability density function defined as follows:

$$f(y; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]_{+}^{-\frac{1}{\xi}}\right\}, -\infty < \mu, \xi < \infty, \sigma > 0$$

The second characteristic is that for k - dimensions, the marginal distribution follows the Multivariate Extreme Value Distribution (Yasin et al, 2019). The parameters obtained based on the GEV probability density function are used to transform *Y(x)* into the Frechet margin using the function in Equation 2.

$$\{Z(x)\}_{x\in X} = \left\{1 + \frac{\xi(x)(Y(x) - \mu(x))}{\lambda(x)}\right\}_{+}^{\frac{1}{\xi(x)}}{}_{x\in X} \tag{2}$$

where $\{Z(x)\}_{x\in X}$is a Max-Stable Process. Process Z is also a Max-Stable Process, with parameters $\mu(x)$, $\xi(x)$, and, and $\lambda(x) > 0$ is a continuous function (Padoan et al. 2010). The Geometric Gaussian model uses a dependency structure, $W_i(x) = \exp\left(\sigma\varepsilon(x) - \frac{\sigma^2}{2}\right)$, with extremal coefficient $\theta(h)$ following Equation 3 research article from Xu and Genton (2016). Then the Z model equation as follows:

$$Z(x) = \max_i \xi_i \exp\left(\sigma\varepsilon(x) - \frac{\sigma^2}{2}\right), x \in X \tag{3}$$

The standard Gaussian process is denoted ε_i, $\rho(h)$ is a correlation function, the value $\varepsilon(0) = 0$ and the bivariate Cumulative Distribution Function (CDF) of this model refers to the bivariate Smith model, stated by Equation (4)

$$P_r[Z(x_1) \le z_1, Z(x_1) \le z_2] = \exp\left[-\frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_1}{z_2}\right)\right] \tag{4}$$

CDF normal standard is denoted by Φ and $a=\sigma\sqrt{2(1-\rho(h))}$ (Davison et al., 2010). This changes Equation 4 to Equation 5:

$$P_r[Z(x_1) \le z_1, Z(x_1) \le z_2] = \exp\left[-\frac{1}{z_1}\Phi\left(\frac{\sigma\sqrt{2(1-\rho(h))}}{2} + \frac{1}{\sigma\sqrt{2(1-\rho(h))}}\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\Phi\left(\frac{\sigma\sqrt{2(1-\rho(h))}}{2} + \frac{1}{\sigma\sqrt{2(1-\rho(h))}}\log\frac{z_1}{z_2}\right)\right] \tag{5}$$

2.2. Geometric Gaussian Model

The CDF of geometric Gaussian is expressed by Equation (6)

$$F\left(z(x_i), z(x_j)\right) = \exp\left[-\frac{1}{z(x_i)}\Phi\left(\frac{\sigma\sqrt{2(1-\rho(h))}}{2} + \frac{1}{\sigma\sqrt{2(1-\rho(h))}}\log\frac{z(x_j)}{z(x_i)}\right) - \frac{1}{z(x_j)}\Phi\left(\frac{\sigma\sqrt{2(1-\rho(h))}}{2} + \frac{1}{\sigma\sqrt{2(1-\rho(h))}}\log\frac{z(x_i)}{z(x_j)}\right)\right] \tag{6}$$

where $z(x_i)$ and $z(x_j)$ state the *i*-location and *j*-location in the Frechet margin, also included in Max-Stable Process. $\rho(h)$ is the correlation function for Schlather models, including Cauchy, Whittle-Matern, Bessel, and Powered Exponential, and *h* is the Euclidean distance between two locations. Bivariate Geometric Gaussian PDF is obtained by simplifying Equation 6 into Equation 7 as follows:

$$F\big(z(x_i), z(x_j)\big) = \exp\left[-\frac{\Phi\big(w(h)\big)}{z(x_i)} - \frac{\Phi\big(v(h)\big)}{z(x_j)}\right] \tag{7}$$

Where $w(h) = \frac{a(h)}{2} + \frac{1}{a(h)}\log\frac{z(x_j)}{z(x_i)}$ $\quad w(h) = \frac{a(h)}{2} + \frac{1}{a(h)}\log\frac{z(x_i)}{z(x_j)}$

$$a(h) = \sigma\sqrt{2\big(1-\rho(h)\big)}$$

Therefore, the bivariate PDF form for the Geometric Gaussian model is

$$\begin{aligned} F\big(z(x_i), z(x_j)\big) = \exp\left[-\frac{\Phi\big(w(h)\big)}{z(x_i)} - \frac{\Phi\big(v(h)\big)}{z(x_j)}\right] \\ \times\left[\frac{\Phi\big(w(h)\big)}{z^2(x_i)} + \frac{\varphi\big(w(h)\big)}{a(h)z^2(x_i)} - \frac{\varphi\big(v(h)\big)}{a(h)z(x_i)z(x_j)}\right. \\ \times\frac{\Phi\big(w(h)\big)}{z^2(x_j)} + \frac{\varphi\big(w(h)\big)}{a(h)z^2(x_j)} - \frac{\varphi\big(v(h)\big)}{a(h)z(x_i)z(x_j)} \\ \left.+\frac{v\varphi\big(w(h)\big)}{a^2(h)z^2(x_i)z(x_j)} + \frac{v\varphi\big(w(h)\big)}{a^2(h)z(x_i)z^2(x_j)}\right] \end{aligned} \tag{8}$$

The process of estimating the parameters of Equation 8 is carried out using Maximum Pairwise Likelihood Estimation (MPLE), then the function is made into the form of the likelihood function, then makes the first derivative based on each parameter and equates to zero. Equations resulting from these derivatives are not closed form. The process of calculating the parameter estimates is continued with the optimization of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Spatial GEV model, written in Equation 9.

$$GEV\big(\mu(s), \sigma(s), \xi(s)\big) \tag{9}$$

Each parameter, namely the location parameter $\mu(s)$, the scale parameter $\sigma(s)$, and the shape parameter $\xi(s)$, is formed following the multiple regression model, then adding spatial elements, namely latitude and longitude as location coordinates. This form is known as the Trend Surface model, which is defined in Equation 10.

$$\begin{aligned} \hat{\mu}(s) &= \beta_{\mu,0} + \beta_{\mu,1}\,\text{longitude}(s) + \beta_{\mu,2}\,\text{latitude}(s) \\ \hat{\sigma}(s) &= \beta_{\sigma,0} + \beta_{\sigma,1}\,\text{longitude}(s) + \beta_{\sigma,2}\,\text{latitude}(s) \\ \hat{\xi}(s) &= \beta_{\xi,0} \end{aligned} \tag{10}$$

The parameters β_μ, β_σ, and β_ξ were estimated using MPLE based on the Probability Distribution Function (PDC) of the Gaussian Geometric model. Selection of the best model uses the Takeuchi Information Criterion (TIC) based on a combination of trend surface models.

TIC is defined by the following equation:

$$TIC = -2\ell_p\big(\widehat{\boldsymbol{\beta}}\big) + 2tr\left\{H\big(\widehat{\boldsymbol{\beta}}\big)^{-1}J\big(\widehat{\boldsymbol{\beta}}\big)\right\} \tag{11}$$

where, $\ell_p(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \ln\left(f\left(u_{ji}, u_{ki}; \hat{\boldsymbol{\beta}}\right)\right)$ is the ln pseudo-likelihood function and is the number of parameters to be estimated. The smallest TIC value indicates the best model.

3. MATERIAL AND METHOD

The rainfall data used comes from the Meteorology, Climatology, and Geophysics Agency of Semarang City, with several measurement posts in Semarang City. These points are the Semarang Climatology Station, Tanjung Mas, and Ahmad Yani. The data period starts from September 1991 to August 2022. The calculation process is carried out using algorithm from Dombry *et al.* (2013) and Ribatet *et al*, (2015), with the following steps:

1. Using Block Maxima to determine data with extreme categories. If there is no rain at all during a certain time, the data will be represented with a value of 0
2. GEV distribution check
3. Calculate univariate parameter estimates $\hat{\mu}(s)$, $\hat{\sigma}(s)$, and $\hat{\xi}(s)$ for each post/location.
4. Transform to Frechet margin units.
5. Make several combinations of trend surface models and select the best one with minimal TIC criteria.
6. Calculate the value of $\hat{\mu}(s)$, $\hat{\sigma}(s)$, and $\hat{\xi}(s)$ for each location
7. Calculating return level values for extreme rainfall predictions

4. RESULTS AND DISCUSSION

As initial information, extreme values have been taken from 10197 data for each rain post in Semarang City, 100 training data and 20 testing data have been obtained. Semarang City experiences rain with the highest daily average intensity of 6.3631 mm/day, which is at the point of the Ahmad Yani Station. The Summary of Semarang City Rainfall starts from September 1991 to August 2022 is presented in Table 1. During this time interval, the city of Semarang experienced the highest rainfall with 276 mm^3 and was included in an extreme event.

Table 1. Summary of Semarang City Rainfall

Measurement Post	Average (mm^3/day)	Standard Deviation (mm^3/day)	Minimum (mm^3/day)	Maximum (mm^3/day)
Semarang Climatology Station	6.3103	15.6117	0	276.0000
Ahmad Yani	6.3631	15.5508	0	255.3000
Tanjung Mas	5.8502	14.6602	0	246.6000

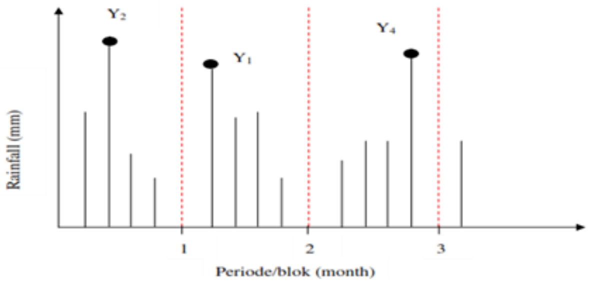

Figure 1. Maxima Block Illustration

The process of univariate estimation of the parameters $\mu(s)$, $\sigma(s)$, and $\xi(s)$, starts from the extreme data from the Maxima Block process and then calculates them based on the GEV PDF model using the MLE estimation method. The Maxima Block process takes the highest data from each block in a certain time as an extreme value, as illustrated in Figure 1. If there is no rain at all during a certain time, the data will be represented with a value of 0.

Table 2 presents the results of univariate parameter estimates $\mu(s)$, $\sigma(s)$, and $\xi(s)$.

Table 2. Univariate GEV Parameter Estimation

Measurement Post	$\hat{\mu}(s)$	$\hat{\sigma}(s)$	$\hat{\xi}(s)$
Semarang Climatology Station	60.27	32.30	0.03
Ahmad Yani	59.65	31.13	0.05
Tanjung Mas	56.32	28.52	-0.03

The parameters obtained in Table 2 are used to transform rainfall data, obtained from the block maxima process into Frechet margins using the Z transformation with the formula Equation 2. The best trend surface model refers to equation 10, selected with the criterion of the smallest TIC value, namely 2537.104. The following best trend surface models:

$$\hat{\mu}(s) = 4.52 + 0.5v(s)$$
$$\hat{\sigma}(s) = 2.48 + 0.21v(s)$$
$$\hat{\xi}(s) = 0.98$$

Then the calculation of the parameter estimation of the geometric Gaussian model in Equation 8 is obtained using the MPLE method. The parameter estimates obtained are $\hat{\mu}(s)$, $\hat{\sigma}(s)$, and $\hat{\xi}(s)$ for each measurement post location. Parameter estimates for each location are listed in Table 3.

Table 3. Multivariate GEV Parameter Estimation

Measurement Post	$\hat{\mu}(s)$	$\hat{\sigma}(s)$	$\hat{\xi}(s)$
Semarang Climatology Station	1.0021	1.0018	0.9703
Ahmad Yani	1.0040	1.0034	0.9703
Tanjung Mas	1.0213	1.0090	0.9703

Return level is a calculation of rainfall threshold predictions based on a period. Rainfall prediction in this study for the next 1-year and 2-year periods with p = 5%. The process of calculating the return level requires a period of $T = \frac{1}{p}$ or $p = \frac{1}{T}$. Results Predicted rainfall based on the value of the return level with a probability of exceeding 1. Table 6 displays the magnitude of the estimated maximum rainfall in the next 1-year and 2-year periods.

Table 6. Return Level Prediction

Period	*Return level*		
	Semarang Climatology Station	Ahmad Yani	Tanjung Mas
1-year	102.42	98.30	89.80
2-year	129.30	121.40	111.00

The SMAPE value is used to measure the performance of the model. Rainfall data divided into training (80%) and testing (20%), obtained a SMAPE value of 16.31% which is included in the good criteria.

5. CONCLUSION

The maximum extreme rainfall prediction results for the next two periods at the Semarang climatology station are 129.30 mm^3, at Ahmad Yani Station are 121.40 mm^3, and at Tanjung Mas are 111.00 mm^3. This model is quite good with the SMAPE value of 16.31. The Max-stable process model in this study gives quite good results for predicting the short period. However, when it is used to forecast the long period, the prediction tends to get bigger and resulting in a decrease in the level of accuracy.

ACKNOWLEDGMENT

This research is funded through the Research Development and Application (RPP) scheme from Universitas Diponegoro in 2022.

REFERENCES

BMKG (2014). *Daftar Istilah Klimatologi*. http://balai3.denpasar.bmkg.go.id/ daftar-istilah-musim#sthash.eC4BlOVG.dpuf (acces on 10 October 2021)

Davison, A.C., Padoan S.A., & Ribatet, M. (2012) Statistical Modelling of Spatial Extremes. *Statistical Science*, 27(2), 161–186.

Brown, B.M. & Resnick, S.I. (1977). Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability*, 14(4), 732–739.

Davison, A.C. & Huser, R., 2015. Statistics of Extremes. *Annu. Rev. Stat. Appl.*, Vol. 2, 203–235

De Haan, L. (1984). A Spectral Representation for Max-Stable Processes. *The Annals of Probability*, 12(4), 1194–1204

Dombry, C., Eyi-Minko, F., & Ribatet, M., 2013. Conditional Simulation of Max-Stable Processes. *Biometrika*, Vol. 100, 111–124.

Hakim, A.R., Yasin, H., & Warsito, B. (2021). Max-Stable Processes With Geometric Gaussian Model On Ocean Wave Height Data. *Journal of Mathematical and Computational Science*, Vol. 11, No. 1, 577–584.

Huser, R. & Genton, M.G. (2016). Non-stationary Dependence Structures for Spatial Extremes. *J. Agric. Biol. Environ. Stat*., 21, 470–491

Kabluchko, Z., Schalater, M., & de Haan, L. (2009). Stationary Max-Stable Fields Associated to Negative Definite Function. *Ann. Probab*., 37, 2042–2065

Padoan, S.A., Ribatet, M., & Sisson, S. A. (2010). Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, Vol. 105, No. 489, *Theory and Methods*, 263–277.

Ribatet, M. (2013). Spatial Extremes: Max-Stable Processes at Work. *J. Soc. Fr. Stat*. 154, 156–177.

Ribatet, M., Singleton, R., & R Core Team. (2015). *SpatialExtremes: modelling spatial extremes*. URL https://CRAN.R-project.org/package=SpatialExtremes R package version 2.0-2.

Smith, R. L. (1990). *Max-Stable Processes and Spatial Extremes*. England: University of Surrey.

Schlather, M. (2002). Models for Stationary Max-Stable Random Fields. *Extremes*. 5(1), 33–44.

Yasin. H, Warsito, B., & Hakim, A. R. (2019). Prediksi Curah Hujan Ekstrem di Kota Semarang Menggunakan Spatial Extreme Value dengan Pendekatan Max Stable Process (MSP). *Media Statistika*, Vol. 12, No. 1, 39–49. https://doi.org/10.14710/medstat.12.1.39-49.

Xu, G. & Genton, M.G. (2016). Tukey Max-Stable Processes for Spatial Extremes. *Spatial Stat.*, 18, 431–443.