Height wave modelling using spatial extreme value with max stable process (MSP) brown-resnick model

A R Hakim, B Warsito, H Yasin

Department of Statistics, Faculty of Science and Mathematics, Diponegoro University
Jl. Prof. Soedharto, SH, Tembalang, Semarang 50275, Indonesia

Corresponding author: arief.rachman@live.undip.ac.id

Abstract. Analysis of extreme data containing spatial elements can be used the Spatial Extreme Value method, which is applied with a max-stable process approach. The data preprocessing process begins by identifying the extreme data of each location for each time period or block. This method is also called Block maxima. The data obtained from the Block Maxima process will follow the Generalized Extreme Value (GEV) distribution. GEV has three forms of distribution, namely Gumbel, Frechet and Weibull. In this study, the data will be transformed into the frechet distribution margin because the frechet has a heavy tail shape, this is required in MSP modeling. Smith model, Schlater model, and Brown Resnik model are the three main models in MSP. Brown technical model is used in this research because Brown Resnik form is a flexible ta of stationary max stable processes in Gaussian random fields. The best model is selecting based on the smallest TIC value from all combinations of Trend surface models. The selected model is then used to determine the prediction of the extreme wave height for each location with a certain time period. The location used in this study is a point where fishermen are widely used to find fish, namely Semarang, Pekalongan, and Rembang which are in Central Java Province.

1. Introduction

The city of Semarang is one of the cities in Central Java which is directly adjacent to the Java Sea in the north. Of course, the sea sector is one of the supporters of the Semarang city economy. Marine activities are also active in the city of Semarang, starting from fishermen, fish pond farmers, shipping to export and import activities by sea to loading and unloading activities at the port of Tanjung Mas. The city of Semarang is the main supporter of the economy in Central Java, from the fisheries sector itself, according to BPS data in 2016 reached 15,000 tons of products produced from marine products. Water waves or waves of more than 2 meters can be said to be high and less safe for fishing, so this causes fishermen's activities to stop temporarily. The threshold for safe fishing as set by BMKG is 1 to 1.5 meters. The availability of information on daily wave height is considered important as a basic reference for fishermen to carry out their activities. Of course this will also affect shipping activities to loading and unloading at the port, which will then impact on the economy in Central Java, especially Semarang[1].

Information about extreme sea wave height must also be available, so that it can be used as a reference to take action to minimize its impact. The statistical method that can be used is the Spatial Extreme Value (SEV) method. Some studies using SEV include Judges (2016)[2] with case studies to model extreme rainfall in the Ngawi region using Schlater and other studies conducted by Malika (2015)[3] to model the extreme rainfall in the Lamongan area.

There are two ways to identify and select extreme data, namely Block Maxima and Peak Over Threshold. In extreme data research, selected based on the Block Maxima (BM) approach, where the highest data from each block or a certain time period will be selected to be used as extreme data. The
extreme data obtained based on the BM approach will follow the form of the Generalized Extreme Value (GEV) distribution. Parameter estimation for the GEV distribution Maximum Pairwise Likelihood Estimation (MPLE).

SEV is a method developed from the Extreme Value Theory (EVT) and the spatial effect on each multivariate extreme data. EVT is a method used to identify extreme events with one variable (univariate). Modeling SEV using the Max-Stable Process (MSP) approach [4], it has three main models, namely Smith, Schlater, and Brown Resnick.

In this study, we will discuss how to identify data extremes that also have spatial dependency effects. The model formed based on the Brown Resnick model will later be used to determine the predicted wave height based on a certain time period. Predictions are calculated using the rate of return approach. Prediction of wave height is expected to provide initial information so that it can minimize the risk that is caused.

2. Literature review

2.1. Extreme Value Theory (EVT)

EVT is a statistical method used to analyze the distribution pattern of tails and was developed for the univariate case. There are two ways to identify and select extreme data, namely Block Maxima and Peak Over Threshold. In extreme data research, selected based on the Block Maxima (BM) approach, where the highest data from each block or a certain time period will be selected to be used as extreme data. The extreme data obtained based on the BM approach will follow the form of the Generalized Extreme Value (GEV) distribution. Block maxima scheme as shown in Figure 1.

![Block Maxima Illustration](image)

Figure 1. Block Maxima Illustration

Samples of extreme data ($y_1, y_2, y_3$) are selected based on the maximum value of each particular block or period. Block Maxima apply Fisher-Tippet Gnedenko (1928) theorem, that sample extreme value data taken from block maxima will follow the distribution of Generalized Extreme Value (GEV) [4]. The cumulative Probability Function of GEV is

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

Where $\mu (-\infty < \mu < \infty)$, $\sigma > 0$, ($\xi \neq 0$) are respectively location, scale, and shape parameters.

Spatial extreme modelling.

However, rainfall data, temperature, height wave, and snow are multivariate series because collected from several locations. Therefore extreme value theory approach is not enough, and then it can be modelled using spatial extreme value.
Spatial Extreme Value modelling requires assumptions of spatial dependence. The concept of spatial dependence following Tobler first law is “everything related to everything else, but near thing is more related than a distant thing”. Visually, spatial data can be seen in Figure 2.

One of approach that can be used for the spatial extreme value modelling is Max-Stable. Max-stable processes generalize the infinite dimension of EVT distribution where samples were taken from the maximum value at each location [5]. The Max-stable process begins with transforming the extreme data into a margin Frechet distribution.

Let $M(s,t)$ is data of extreme events with location $(s)$ and period $(t)$ on spatial domain $D \subset \mathbb{R}^2$. Distribution of $M(s,t)$ is:

$$M(s,t) \sim GEV(\mu(s,t), \sigma(s,t), \xi(s,t)),$$

where $\mu(s,t), \sigma(s,t), \xi(s,t)$ it is respectively locations, Scale and shape parameters from GEV distribution. An arbitrary distribution function $G$ is maxed stable if and only if $G$ following GEV distribution.

Let $X$ be an index set and $\{Y_i(x)\}_{x \in X}, i = 1, 2, \ldots, n$ be $n$ independent replication of a continuous stochastic process. Assume that there is a continuous sequence of functions $a_n(x) > 0$ and $b_n(x) \in \mathbb{R}$ such that:

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

where $Y_1, \ldots, Y_n$ are independent replications of $Y$, if this limit exists is max-stable process de Haan [6]. If $a_n(x) = n$ and $b_n(x) = 0$ $Y(x)$ also is simple Max Stable Processes [4]. Assume that each component on each location has distribution of GEV, and then do transformation into Frechet margins:

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

This process can obtain by standardizing $\{Y(x)\}_{x \in X}$ such that:

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

where $\mu(x), \xi(x)$ and $\lambda(x) > 0$ is a continuous function. $Z$ is still max-stable processes [7]. In general, the Max Stable Process with the marginal Frechet unit can be explained by the equation (3):

$$Z(x) := \max_{j \geq 1} \left\{U_jW_j(x)\right\}, \quad x \in X,$$
where $W_j$ and $U_j$ are Poisson process at $(0, +\infty) \times \mathbb{R}^2$ with measurement intensity $\nu(dx) \times u^{-2} du$. From the general model, it was developed into a max-stable model, one of which was Smith’s model. The Smith model is

$$Z(x) := \max_{j \geq 1} \left( U_j f \left( Y_j - x \right) \right), \quad x \in X$$

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

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

2.2. Brown Resnick model

Brown Resnick’s MSP model was proposed by Brown and Resnick (1977) and generalized by Kabluchko [8], by defining the dependency structure in equation (3) so that it becomes equation (4).

$$Z(x) = \max_{j \geq 1} \left( U_j \exp \left( \varepsilon_j(x) - \gamma(x) \right) \right), \quad x \in X,$$

Where $\varepsilon_j$ normally distributed with the semivariogram $\gamma(h)$ and $\varepsilon(0) = 0$. Model Brown-Resnick has the same bivariate cumulative distribution function as the Smith model with $a(h) = \sqrt{2\gamma(h)}$ so it looks like this:

$$F(z_i, z_j) = \exp \left[ -\frac{1}{z_i} \Phi \left( \frac{\sqrt{2}\gamma(h)}{2} + \frac{1}{\sqrt{2}\gamma(h)} \log \frac{z_j}{z_i} \right) \right] - \frac{1}{z_j} \Phi \left( \frac{\sqrt{2}\gamma(h)}{2} + \frac{1}{\sqrt{2}\gamma(h)} \log \frac{z_i}{z_j} \right),$$

with $\Phi$ is the standard normal cumulative distribution function, and for locations $x_i$ dan $x_j$ the vector $h$ is $(x_j - x_i)^T$ and $a(h) = (h^T \Sigma^{-1} h)^{1/2}$.

Coefficient extremal

In Extreme Value Modelling using Max-Stable process, measurement extremal dependence is indispensable. Extremal coefficient describes characteristics of the matrix of dependencies tail. Extremal coefficient function for Brown Resnick model is following as:

$$\theta \left( x_1 - x_2 \right) = -z \log \left\{ Z(x_1) \leq z, Z(x_2) \leq z \right\}$$

$$\theta(x_j - x_k) = 2\Phi \left( \frac{\sqrt{\gamma(x_j - x_k)}}{2} \right)$$

where $1 \leq \theta(s_1 - s_2) \leq 2$ if $\theta(s_1 - s_2) = 1$ then full dependencies and if $\theta(s_1 - s_2) = 2$ then independence [9].

Selecting the best model

Takeuchi Information Criterion (TIC) will be employed to choose the best trend surface model, estimate parameters process $\beta_0$, $\beta_1$, and $\beta_2$ based on PDF of Brown Resnick Max-Stable model using Maximum Pairwise Likelihood Estimation (MPLE). Trend surface model following as:
\[ \hat{\mu}(s) = \beta_{\mu,0} + \beta_{\mu,1} \text{lon}(s) + \beta_{\mu,2} \text{lat}(s) \]

\[ \hat{\sigma}(s) = \beta_{\sigma,0} + \beta_{\sigma,1} \text{lon}(s) + \beta_{\sigma,2} \text{lat}(s) \]

\[ \hat{\xi}(s) = \beta_{\xi,0} \]

(7)

Trend Surface models are several combinations of models with latitude and longitude spatial components, then from the combination of these models one is selected based on the smallest TIC.

Padoan, Ribatet, and Sisson [10] propose selecting the best model using Composite Likelihood Information Criterion which has developed into TIC. TIC function is given by:

\[ TIC = -2 \ell_p (\hat{\psi}) + 2tr \left\{ H (\hat{\psi})^{-1} J (\hat{\psi}) \right\}, \]

where

\[ H (\hat{\psi}) = -\frac{\partial^2 \log f(z_{ik}, z_{jk}; \hat{\psi})}{\partial \psi \partial \psi^T} \]

\[ J (\hat{\psi}) = -\sum_{k=1}^{K} \frac{\partial \log f(z_{ik}, z_{jk}; \hat{\psi})}{\partial \psi} \frac{\partial \log f(z_{ik}, z_{jk}; \hat{\psi})}{\partial \psi^T} \]

and \( i = 1, 2, \ldots, m-1 \), \( j = 1, 2, \ldots, m \), \( \hat{\psi} \) is estimator Maximum pairwise likelihood. Then the extreme rainfall prediction uses the concept of return level.

3. Methodology

The data used is secondary data obtained from the Meteorology, Climatology and Geophysics Agency (BMKG) of Central Java. The variable data used in this study is the wave height measured from several stations, namely Pekalongan Station, Rembang Station, Semarang Station from 2016-2019. The analysis process begins with the identification of Extreme data, then the wave height data is transformed into a frechet margin that has a heavy tail pattern, calculating the univariate GEV parameter estimates, modeling with the max-stable approach to the brown resnick model process, selecting the best model and calculating the predicted wave height with the return level approach.

4. Results and discussion

Description of wave height at three sea observation Stations, can be used as preliminary information to determine the characteristics or general description and rainfall patterns used. Overall based on the sea wave height maximum value is 3.7 m that is in the waters of Rembang Completely presented in Table 1

| Location | Mean  | St.Dev | Min  | Max  |
|----------|-------|--------|------|------|
| Pekalongan Stations | 0,407 | 0,4022 | 0    | 2,5  |
| Rembang Stations      | 0,5497| 0,6054 | 0    | 3,7  |
| Semarang Stations     | 0,3019| 0,3937 | 0    | 2,5  |

Tail behavior identification aims to find out whether there is a heavy tail data pattern that indicates that the data contains extreme values.

The heavy tail pattern of the three monitoring stations can be seen in Figure 3, it can be seen that the distribution of the tail is slowing down, this is an indication that the daily data for wave height contain extreme values. Extreme data obtained from the maxima block approach uses 12 blocks or time periods in one year.
Figure 3 Histogram Of Rainfall Intensity Each Location Probability Plot

From Figure 3 distribution patterns have a high frequency in zero values so that distribution pattern has heavy tailed. Second, selecting extreme value from the data using Block Maxima method begins to make block / period from the data, the block formed adapted to the needs and conditions of data. The extreme value selected from the maximum value of each block.

Figure 4 Probability Plot Extreme Value

Based on Figure 4 can be seen that the samples of extreme height wave with one month period in all monitoring station has followed GEV distribution, it's because all the distribution points follow a linear line which occurs at the other three wave height monitoring stations.
Parameter estimates for $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ univariate each model using MLE and solved with the numeric method using BFGS (Broyden Fletcher Goldfarb Shanno) Quasi-Newton. Count process using R software.

Table 2. Estimates Parameter GEV Univariate

| Station      | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\xi}$ |
|--------------|--------------|----------------|-------------|
| Pekalongan   | 0.84028      | 0.4407         | -0.0088     |
| Rembang      | 0.99443      | 0.8296         | -0.1557     |
| Semarang     | 0.42434      | 0.3412         | 0.53601     |

Then, let transformation extreme Height wave data which obtained from block maxima into Frechet distribution using transformation $Z$:

$$Z(s) = \left(1 + \xi \frac{y - \mu}{\sigma}\right)^{\frac{-\frac{\mu}{\xi}}{\sigma}}$$

where $y$ is extreme value sample, $s_i$ shows the location of the Heightwave observation stations and $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ are GEV parameter.

The best model selected based on Takeuchi Information Criterion (TIC). The smallest TIC value is 1018.26 from the combination Trend Surface model that is

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

The trend surface model obtained from the max-stable smith model is as follows:

$$\hat{\mu}(s) = -7.096 - 1.179 \text{latitude}(s)$$
$$\hat{\sigma}(s) = -17.531 - 2.724 \text{latitude}(s)$$
$$\hat{\xi}(s) = 1.419$$

Then obtained parameter estimates that differ for each location, where parameter $\hat{\xi}(s)$ are constant in each location, shown in Table 3:

Table 3. Brown resnick Model Parameter Estimates

| Rain Monitoring Post | $\hat{\mu}(s)$ | $\hat{\sigma}(s)$ | $\hat{\xi}(s)$ |
|----------------------|----------------|-------------------|---------------|
| Pekalongan Station   | 0.987          | 1.147             | 1.419         |
| Rembang Station      | 0.805          | 0.728             | 1.419         |
| Semarang Station     | 1.097          | 1.401             | 1.419         |

Prediction of extreme wave height is calculated using the return level approach over a certain time period. The result of Estimation of parameter $\hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s)$ will be used to calculate the predicted extreme wave height for each location. Return level predictions, can only be done starting from the second year period. The last Step is computed to predict Return Level of extreme height wave for each observation station, it will predict for 1, 2 and 3 years. General equation for return level following as:
\[
    z_p(s) = \hat{\mu}(s) - \frac{\hat{\sigma}(s)}{\hat{\xi}(s)} \left( 1 - \ln \left( 1 - \frac{1}{T} \right) \right)^{-\frac{1}{\hat{\xi}(s)}}.
\]

where \( T = 1 \text{ year} \times 12 \) (number of block) = 12.

Table 4 will contain prediction value of Return Level

| Rain Monitoring Post | Return level (mm) |
|----------------------|-------------------|
|                      | 2 years | 3 years | 4 years |
| Pekalongo Station    | 2.89     | 3.14    | 3.31    |
| Rembang Station      | 2.31     | 2.55    | 2.72    |
| Semarang Station     | 3.05     | 3.29    | 3.47    |

Table 4 presents the return level predictions with the return periods of 1, 2, and 3 years ahead and the probability exceeded are 0.042, 0.028 and 0.021 respectively, it is the probability of reaching the maximum threshold for \( z_p \), which is equal to \( p = 1/T \).

5. Conclusion

Estimates parameter of univariate indicates that extreme height wave data in three station observation on central java region already has GEV distribution. Parameter estimate for GEV distribution using Maximum Pairwise Likelihood Estimation (MPLE) method and solved with the numeric method using BFGS Quasi-Newton. Parameter Shape of GEV spatial always has a constant value with estimation negative value. The result of wave height prediction based on the return level using the brown resnick model shows that the location of Semarang occupies the highest position for the prediction of the maximum sea wave height. In this study, the predicted value of wave height in the next 2 years return period shows the maximum number of waves as high as 3.05 meters will occur.

Acknowledgments

Our gratitude goes to Diponegoro University for its support. This research was funded by the 2019 "PTUPT" Research Grant.

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 (diakses pada tanggal 8 maret 2018).
[2] Arief R H, Sutikno and Dedy D P 2016 *The 3rd International Conference on Research, Implementation and Education of Mathematics and Sciences (ICRIEMS)* (Yogyakarta: Indonesia)
[3] Rosna M, Sutikno, Wahyu W 2015 *Prosiding Simposium Nasional Inovasi dan Pembelajaran Sains* (Bandung: Institut Teknologi Bandung)
[4] Gilli M and Kellezi E 2006 *An Application of Extreme Value Theory for Measuring Risk* (Elsevier Science)
[5] Cooley D, Naveau P and Poncet P *Variograms for Spatial Max-Stable Random Fields: Lecture Notes in Statistics* (New York: Springer)
[6] de Haan L 1984 *Ann. Probab.* **12** 4 1194-1200
[7] Smith R L 1990 *Max-stable Processes and spatial extremes* (The university of North Carolina)
[8] Kabluchko Z, Schlather M and de Haan L 2009 *Ann. Probab.* **37** 2042–2065
[9] Schlather M and Tawn J 2003 *Biometrika* **90** 1 139–156
[10] Padoan S A, Ribatet M and Sisson S A 2010 *J. Am. Stat. Assoc.* **105** 489 263-277