Identification of Extreme Rainfall Pattern
Using Extremogram in West Java

Achi Rinaldi, Anik Djuraidah, Aji Hamim Wigena,
I Wayan Mangku, Dodo Gunawan.

1Department of Mathematics Education, UIN Raden Intan, B. Lampung, Indonesia
2Department of Statistics, Bogor Agricultural University, Bogor, Indonesia
3Department of Mathematics, Bogor Agricultural University, Bogor, Indonesia
4Meteorological, Climatological, and Geophysical Agency, Indonesia

ABSTRACT
The extremogram is a method for measuring extreme events dependence in stationary time series, which can be viewed analog as the autocorrelation function (ACF) of extreme value. This study aims to measure the time dependency of extreme rainfall in West Java using extremogram, and to observe the pattern of time dependencies of extreme rainfall between stations and see the relationship based on distance using cross-extremogram. The data used is monthly rainfall from 30 BMKG rain observation stations with period from 1981 to 2012. In general, the extremogram plot shows that extreme rainfall in West Java has periodic pattern except for zones 14 and 18. As seen from the peaks of maximum values for monthly rainfall in extremogram, extreme rainfall occurring over a certain period of time, especially in November, December, and January. The periodic pattern is illustrated by the recurrent extremogram peak values for lag-11, lag-12, and lag-13. The plot between the cross-extremogram and the distance shows the further distance between the rain stations, the value of extreme rainfall dependency gives a smaller correlation value.

Keywords: ACF, cross-extremogram, extremogram.

1. Introduction
Extreme events are events with very rare frequencies by its definition. Usually extreme events are associated with disasters or events that can cause damage, such as: floods, hurricanes, earthquakes, and tsunami. Through the concept of statistics, researchers are trying to model and predict extreme events in disaster prevention and mitigation efforts resulting from the impacts. One of the factors causing the disaster that is interesting to be studied by the researchers is the rainfall, especially rainfall whose intensity far exceeds the average value or referred to as extreme rainfall. There are some research on extreme rainfall in Indonesia such as Sadik [1], Prang [2], Irfan [3], Sari [4], and Rinaldi et al. [5]. Sadik [1], Prang [2] and Irfan [3] only modeled extreme rainfall using univariate data, while Sari [4] added the location factor to the extreme rainfall model based on the model developed by Davison et al. [6]. Rinaldi et al. [5] modeled extreme rainfall with spatial model and resulted in the form of extreme rainfall zone in West Java, unfortunately all of these studies have not included the element of time in it. In addition, modelling extreme events in climate data was addressed in Davison and Smith [7] and Smith [8].

One of the most commonly used methods for measuring dependencies in time series data is the Auto Correlation Function (ACF). The pattern shape of the ACF plot became the basis of the researchers to choose a model suitable for time series data, but the concept of...
ACF cannot be applied to extreme values. Davis and Mikosch [9] introduce extremogram as a method that can be used to measure and see extreme value dependencies in time series data. Extremogram is a method that can measure dependencies between events that are above a certain threshold. Threshold used is the concept of quantile, where the quantile selection is upper quantile based on the purpose of the study. The Extremogram can only see the dependencies of a single time series variable. The method used to look at extreme dependencies for two time series variables is cross-extremogram which can be analogous to the concept of Cross-correlation Function (CCF). In addition to the quantity size of dependencies, both extremogram and cross-extremogram can also be presented in the form of graphs so that time trend patterns for extreme events can be seen. Several other studies using the concept of extremogram are Davis et al. [10] who developed extremograms in the form of multiple variables, and Nadezda [11] who applied the concept of extremogram to see extreme spike patterns in the case of electricity prices.

Indonesia is a tropical country with rain probability to occur over the year. Extreme rainfalls not only occur in the rainy season but can occur in the dry season. This is the reason why researchers want to study patterns and time dependencies on extreme rainfall. This study aims to see and measure the extreme rainfall dependencies in time series that occur in Indonesia, especially in areas of extreme precipitation zone in West Java using extremogram. It will also analyze the pattern of time dependencies of extreme precipitation time between stations in different time series data using cross-extremogram, and to see the dependency pattern based on the distance between stations.

2. Materials and Method

2.1. Measures of Serial Dependence

Auto Correlation Function (ACF) is a method for measuring time series data dependencies. ACF can see the correlation between lag, see patterns of trend or periodic data, and identify white noise. ACF is defined as the linear dependency measure of time series from time $t$ to time $t+h$, which can be written as:

$$\rho(h) = \frac{cov(X_t, X_{t+h})}{\sqrt{var(X_t)var(X_{t+h})}}$$

The ACF is estimated by the sample ACF as follows:

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^{T} (X_t - \bar{X}) (X_{t+h} - \bar{X})}{\sqrt{\sum_{t=1}^{T} (X_t - \bar{X})^2 \sum_{t=h+1}^{T} (X_{t+h} - \bar{X})^2}}$$

ACF is a measure of dependencies for a univariate time series. However, sometimes we have several time series and we would like to measure linear dependence between them. The cross-correlation function (CCF) is a well-known method used in this case. CCF is a measure of linear dependencies between two time series $X_t$ and $Y_{t+h}$, which can be written as:

$$\rho_{X,Y}(h) = \frac{cov(X_t, Y_{t+h})}{\sqrt{var(X_t)var(Y_{t+h})}}$$

The CCF is estimated by the sample CCF as follows:

$$\hat{\rho}_{X,Y}(h) = \frac{\sum_{t=h+1}^{T} (X_t - \bar{X}) (Y_{t+h} - \bar{Y})}{\sqrt{\sum_{t=1}^{T} (X_t - \bar{X})^2 \sum_{t=h+1}^{T} (Y_{t+h} - \bar{Y})^2}}$$
2.2. Extremogram

For a strictly stationary \( \mathbb{R}^d \)-valued time series \( (X_t) \), the extremogram is defined for two sets \( A \) and \( B \) bounded away from zero by:

\[
\rho_{A,B}(h) = \lim_{x \to \infty} P(\{x^{-1}X_0 \in A\} \cap \{x^{-1}X_h \in B\}) , h = 0, 1, 2, ...
\]

provided the limit exists. Since \( A \) and \( B \) are bounded away from zero, the events \( \{x^{-1}X_0 \in A\} \) and \( \{x^{-1}X_h \in B\} \) are becoming extreme in the sense the probabilities of these events are converging to zero with \( x \to \infty \), with \( h \) is the value of the time lag.

Estimation of the extremogram can be done by replacing the limit on \( x \) in (1) with \( a_m \), which is defined as \((1 - 1/m)\)-quantile of the stationary distribution of \(|X_t|\).

Estimator of the extremogram based on the observations \( X_1, X_2, ..., X_n \) is given by [12]:

\[
\hat{\rho}_{A,B}(h) = \frac{\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_{t+h} \in B \} \cap \{a_m^{-1}X_t \in A\}}} {\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in A\}}}
\]

\( \hat{\rho}_{A,B}(h) \) is asymptotically normal distribution that has been proven as follows [9]:

\[
\sqrt{n/m} (\hat{\rho}_{A,B}(h) - \rho_{A,B;m}(h)) \to N \left( 0, \sigma^2_{A,B}(h) \right)
\]

where: \( \rho_{A,B;m}(h) = P(\{a_m^{-1}X_h \in B \} | a_m^{-1}X_0 \in A) = \frac{P(a_m^{-1}X_0 \in A, a_m^{-1}X_h \in B)}{P(a_m^{-1}X_0 \in A)} \).

In simple terms it can be said that the extremogram is a method to estimate the probability of occurrence of extreme event in time \( (t + h) \) provided that there is an extreme event in period \( t \). Unfortunately extremogram is a method that can measure extreme dependencies within only one time series. Sometimes we want to see how far another time series may affect the series that is being analyzed. If the assumption of normal distribution is used then CCF is the right tool, whereas cross-extremogram is the right method to look at the extreme dependencies of bivariate time series. Cross-extremogram defined as:

\[
\rho_{A,B}(h) = \lim_{x \to \infty} P(\{x^{-1}Y_h \in B \} | x^{-1}X_0 \in A), h = 0, 1, 2, ...
\]

cross-extremogram estimator calculated as follows:

\[
\hat{\rho}_{A,B}(h) = \frac{\sum_{t=1}^{n-h} I_{\{a_m^{-1}Y_{t+h} \in B \} \cap \{a_m^{-1}X_t \in A\}}} {\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in A\}}}
\]

where \( a_m \) is \((1 - 1/m)\)-quantile that indicate extreme events in time series \( X_t \) and \( Y_t \).

2.3. Data

The data are monthly rainfall data (1981-2012) of 30 stations in West Java. Data from Meteorological, Climatological, and Geophysical Agency (BMKG). The region used is 9 zones from the extreme rainfall zone [5], which can be seen in Figure 1.
2.4. Methods

The steps of this study are itemized below:
1. Identifying extreme rainfall data with maximum rainfall plot.
2. Calculating the temporal dependencies of monthly rainfall using extremogram and analyzing extreme rainfall patterns over time through the extremogram plot.
3. Calculating the temporal dependencies of the monthly rainfall of each station to one another, using cross-extremogram.
4. Making a plot between the value of cross-extremogram and the distance between stations.
   a. The distance between stations is calculated by the difference of longitude and latitude coordinates which are converted into kilometers.
   b. The cross-extremogram used is a lag-0, indicating the correlation of extreme rainfall between stations at the same time.
   c. Cross-extremogram estimators for lag-0 can be written as:

\[
\hat{\beta}_{AB}(0) = \frac{\sum_{t=1}^{n} I[a_{AB}^{-1}Y_t \in B, a_{AB}^{-1}X_t \in A]}{\sum_{t=1}^{n} I[a_{AB}^{-1}X_t \in A]}
\]

3. Result and Discussion

3.1. Identification of Extreme Rainfall

The maximum monthly rainfall values from 1981-2012 shown in Figure 2 shows that the maximum precipitation for each year varies greatly. Although the values used are the maximum value, it turns out that the majority of rainfall in a year still have outliers at the top of the diagram grid lines on boxes with longer lines. This indicates that the monthly rainfall in
each year has the opportunity to far exceed the average value, in other words the chance of occurrence of extreme rainfall events in almost every year.

![Figure 2 Box plot of Maximum Monthly Rainfall in 1981-2014](image)

### 3.2. Extreme Dependencies with Extremogram

Several zones were selected to view time series dependencies, the zones are zone 3, 4, 5, 14, 15, 16, 17, 18, 19. The multiple zones chosen to see the diversity of extreme rainfall dependencies by successive time. Zones 3, 4, and 5 represent zones with extreme high intensity. Zones 15, 16 and 17 represent moderate intensity rainfall, while zones 14, 18, and 19 represent low-intensity rainfall. Furthermore from each rain station per zone is calculated the value of extremogram and presented in extremogram plot.

Figures 3 until 7 are an extremogram plot for each station in their respective zones. In general, the form of extremogram plot shows periodic trend except for zone 14 and zone 18 which are not visible to be a periodic trend. Almost of extremogram plot show periodic trend with a period 12. As the data was monthly rainfall, this means that there is a high probability given an extreme event at time $t$ to observe another one in the next 12, (24, 36, etc.) months. Beside lag-12, lag-11 and lag-13 also show significant probabilities of extreme rainfall. This is consistent with the rainfall pattern in Indonesia which tends to be high in November, December and January. It can be concluded that the occurrence of extreme rainfall will have a great chance of occurring indeed in the rainy months.
Figure 3 Extremogram Plot in Zone 3 and Zone 4

Figure 4 Extremogram Plot in Zone 5 and Zone 14
Figure 5 Extremogram Plot in Zone 15

Figure 6 Extremogram Plot in Zone 16 and Zone 17
3.3. Extreme Dependency with Cross-extremogram

Lw Gede station has the best periodic rainfall pattern compared to other rain stations, it can be seen through the extremogram plot of figure 3. Furthermore, to see how far extreme rainfall at station one can affect other stations, then we use cross-extremogram with reference point is Lw Gede rain station which has the best periodic rainfall pattern. In general, cross-extremogram values for Lw Gede stations against other rain stations have a periodic pattern with peaks of extreme values lag-11, 12 and 13. This indicates that the extreme rainfall have a high probability to occur every year, and the possibility of occurrence in November, December, and January. The cross-extremogram plot can be seen in Figure 12.

The cross-extremogram can be used to see how far the rainfall of one station and the other affect each other based on distance. From Figure 8 it can be seen that greater distance from one station to another station, the time dependency value for extreme rainfall is also smaller. This indicates that there is a spatial dependency for extreme rainfall, in line with Rinaldi et al. [5] which has proved the existence of spatial dependencies on extreme rainfall in West Java.
Figure 8 Cross-extremogram Plot with Lw Gede Station as Reference Point
4. Conclusion

Extreme rainfall in West Java generally has a periodic pattern and tends to recur each year. Extreme rainfall has a great chance of occurring in the rainy season especially in November, December and January. This is a reference for stakeholders to plan for disaster anticipation especially caused by extreme rainfall, such as floods and landslides especially in disaster prone areas. Besides the existence of temporal dependencies it turns out that extreme rainfall also has spatial dependencies, the closer the extreme rainfall area the stronger the effect on the surrounding area as well as vice versa.

References
[1] Sadik K 1999 Pemodelan nilai ekstrim terampat untuk proses lingkungan (studi kasus pada curah hujan harian) [tesis] Bogor: Program Pascasarjana Bogor Agriculture University
[2] Prang J D 2006 Sebaran nilai ekstrim terampat dalam fenomena curah hujan [tesis] Bogor: Program Pascasarjana Bogor Agriculture University
[3] Irfan M 2011 Sebaran pareto terampat untuk menentukan curah hujan ekstrim (studi kasus: curah hujan periode 2001-2010 pada stasiun Darmaga) [skripsi] Bogor: Program Sarjana Bogor Agriculture University
[4] Sari F M 2013 Prediksi curah hujan ekstrim secara spasial (studi kasus: curah hujan bulanan di kabupaten Indramayu) [tesis] Bogor: Program Pascasarjana Bogor Agriculture University
[5] Rinaldi A, Djuraidah A, Wigena A H, Mangku I W, and Gunawan D 2017 Spatial extreme models with copula to determine extreme rainfall zone Applied Mathematical Sciences 11(27) 1327 – 1336

[6] Davison A C, Padoan S A, and Ribatet M 2012 Statistical modeling of spatial extremes Statistical Science 27 161-186

[7] Davison A C and Smith R L 1990 Models for exceedances over high thresholds Journal of the Royal Statistical Society Series B (Methodological) 393–442

[8] Smith R L 1989 Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone Statistical Science 367–377

[9] Davis R A and Mikosch T 2009 The extremogram: a correlogram for extreme events Bernoulli 15 977-1009

[10] Davis R A, Mikosch T, and Cribben I 2011 Estimating extremal dependence in univariate and multivariate time series via the extremogram arXiv:1107.5592v1 [stat.ME]

[11] Nadezda F 2016 Estimation of extreme value dependence: application to Australian spot electricity prices [tesis] Department of Mathematical and Statistical Sciences University of Alberta

[12] Davis R A, Mikosch T, and Cribben I 2012 Towards estimating extremal serial dependence via the bootstrapped extremogram Journal of Econometrics 170(1) 142–152