Bootstrap resampling to detect active zone for extreme rainfall in Indonesia

Heri Kuswanto, Sri Hidayati, Mutiah Salamah, and Brodjol Sutijo Ulama
Department of Statistics, Faculty of Mathematics and Natural Science
Instytut Teknologi Sepuluh Nopember, Indonesia
E-mail: heri.k@statistika.its.ac.id

Abstract. This research aims to develop a methodology to detect active zones related to the extreme rainfall event in one of regions in Indonesia. Active zone is defined as regions or zones on the atmospheric level which have significantly different pattern with other regions prior to the occurrence of extreme event happen in the study area. The detection will be useful for the forecasters to predict the extreme rainfall events, and hence, the risk of the disaster caused by the events can be minimized. In order to predict the active zone, this paper examines statistical procedure that is able to test the significant difference between weather phenomena at the atmospheric level onset with prior to the occurrence.

1. Introduction

Climate change happens within the last decade has lead to the increasing of extreme events intensity. Pall et al.[7] showed that the global gas emission anthropogenic greenhouse increases the risk of flood in England. Using the case of Northern Hemisphere data, Hansen et al.[5] showed that the human behavior towards the greenhouse gas has contributed to the increasing of heavy precipitation intensity. Dai[1] also found that the percentage of dry global area increased about 1.74% per decade spanning from 1950 to 2008 caused by the global warming. Min et al.[6] argues that extreme anomaly such as happen in Texas and Oklahoma in 2011 and Moscow 2010 were the consequence of global warming. Another research discussed the impact of climate change to the extreme weather is Rahmstorf and Coumou[11] among others.

Indonesia is also one of the most vulnerable countries towards the impact of climate change. Especially in East Java, extreme precipitation has led to severe flood and caused significant losses in term of infrastructure and human victim. To deal with this, actions to reduce disaster risk is necessarily to be carried out. One of the feasible mitigation strategies is to study the meteorological pattern of some weather variables prior to the occurrence of the extreme events.

This research proposes composite maps showing the atmospheric dynamic prior to the occurrence of extreme precipitation in Indramayu, East Java. Similar maps has been developed by Grotjahn and Faure [4] to predict several types of extreme events in Sacramento, California. In fact, the maps has a very good performance used as a guidance for the forecaster in the Meteorological Office. The composite maps developed in this research differ from Grotjahn and Faure [4] in the sense that the grid spaces is smaller, meaning that the proposed map has higher resolution. Grotjahn and Faure used 2.5 x 2.5 degree as the grid space, while the proposed maps
is developed with 0.25 x 0.25 degree. As a consequence of using more precise grid space, the predictability of the event will increase.

The map is composed from reforecast data treated with statistical resampling procedure namely bootstrap. The bootstrap will generate artificial data to calculate the climatological mean as the basis to identify regions or zones over Indonesia which show significant different pattern on one and two days prior to the heavy precipitation in Indramayu. Bootstrap is used due to the fact that the underlying distribution of weather variable is unknown or maybe not normal, and hence, bootstrap nonparametric approach can be a powerful method to solve the problem of lack information about distribution. As in illustration, the composite maps will be developed for Mean Sea Level Pressure (MSLP).

2. Research Methodology
In order to develop the composite maps, this research uses two kind of dataset as follow:

(i) Field observed Precipitation data
   This is daily rainfall data collected from one of the meteorological stations in Indramayu. The dataset span from 1979 to 2006.

(ii) Atmospheric data (Reforecast Era-Interim data)

   The steps of analysis are described below:

   • identify the date of extreme events from observation data
   • describe the MSLP on the Indonesian regions (specified by latitude and longitude)
   • for all grids, analyze the MSLP at 1 day and 2 day prior to the occurrence of extreme rainfall in Indramayu using bootstrap
   • apply bootstrap resampling to create artificial data for calculating the climatological mean
   • estimate the percentile bootstrap to create the confidence level with certain significant level
   • test the significant different between mean of bootstrap with climatological mean (both day-1 and day-2). The zone is said to be active if the mean of extreme events lies outside the interval.

3. Bootstrap

Bootstrap is used to estimate the sample mean, variance and other measures. The method estimates the sampling distribution of statistical measure using a simple procedure i.e. resampling (Efron and Tibshirani [2]; Efron [3]). Bootstrap estimates the properties (statistical measures) of estimator by by measuring the property from the closest sampling distribution. One of the standard choices is through the empirical distribution of the analyzed data. In case where sequence of observations are assumed to be identical and independent, the empirical distribution can be developed by generating number of resampling from the observed data (with the number of sample same as the raw data), by which each of them are drawn randomly from the raw data. Bootstrap can also be used to test the hypothesis. In this case, bootstrap is usually used as alternative for statistical inference that need assumption about underlying distribution, or to calculate a very complex standard error.

   Below are the steps of bootstrap resampling:

   • given the sample with the size of n, do resampling with or without replacement and get R replication. The number of replication R has to be sufficient enough to estimate the sampling distribution.
   • calculate the mean estimator from each resampling
   • calculate the bootstrap mean by averaging mean values of all resampling.
Bootstrap has been applied to solve many problems in meteorology and climatology for instance in Lall and Sharma [8], Orlowsky et al.[10] as well as Mudelsee [11] who discusses bootstrap application for risk climate analysis.

4. Results and Discussion
The descriptive statistic of the daily rainfall in Indramayu for the annual period within 1979 to 2006 observed from Bangkir station is performed in order to study the characteristic of the rainfall to be used for further analysis.

![Average number of rainfall days](image1)

![Number of Extreme days](image2)

**Figure 1.** Number of rainfall based on each category.

**Figure 2.** Average number of rainfall every month.

Figure 1 shows that on the period of one year, the average days of light rainfall is 43 days, average or normal rainfall is 29 days and heavy rainfall in about 6 days, with 1 day is considered to be extreme. The information about daily rainfall on monthly period is depicted in Figure 2, and it shows that the rainy season in Indamayu happens on November to April, while the dry season happens on May to October.

4.1. Extreme Events Detection in Indramayu
Identification of extreme events can be seen from normal probability plot.

![Normal Probability Plot](image3)

**Figure 3.** Normal Probability Plot.

Figure 3 shows that the pattern of daily rainfall consists of tail distribution (extreme value), and hence the probability distribution function is away from normal distribution. This fact shows nonlinear pattern of the distribution which strongly indicates that the distribution is not normal. It is confirmed also by the p-value of the statistical test less than 0.01.

Another popular method to identify the extreme value statistically is by using Peak Over Threshold (POT). It uses a threshold value u as the cut off for determining extreme. This research uses Mean Residual Life Plot (MRLP) to find the threshold.
From the MRLP plot, the threshold is chosen by identifying the magnitude of the mean excess in which it shows a linear pattern. From Figure 4, the MRLP start to be stable at the point of 50, and hence, the threshold will be set as 50. This value is consistent with the extreme rainfall definition specified by the Agency for Meteorology, Climatology and Geophysics (BMKG) Indonesia. The extreme rainfall data are observations above the threshold as shown in Figure 5.

Another statistic reveals that extreme rainfall mostly happen in November, December, January, February and March. These months are exactly the period of rainy season in Indonesia. January is the month with the highest intensity of extreme rainfall with about 60 events. Further analysis will use only extreme rainfall detected within the rainy season due to the fact that heavy rainfall happening in dry season might be the result of convolution.

4.2. Analysis of Mean Sea Level Pressure over Indonesia Region

In this research, the MSLP data that will be analyzed covers the area of over Indonesia region. The region is sliced into grid space with the size of of 0.25 degree x 0.25 degree. The dataset are reforecast era interim data published by European Center for Medium Range Weather Forecasting Model (ECMWF) and available at http://ecmwf.int/.

4.3. Bootstrap to identify Active Zones for Mean Sea Level Pressure

We listed the date of extreme rainfall based on the threshold specified on the previous section. The idea of bootstrapping is to create artificial data corresponding to one and two days prior to the occurrence of the events. The analysis is done for each grid coordinate. Detail procedures of bootstrap resampling for the considered case are as follows:

- Dataset to be sampled span from 1979 to 2006
- The number of random sample drawn in each resampling process is 146 days (i.e. number of extreme rainfall events) both for day-1 and day-2
- The resampling process is done with replacement
- The number of replication is set to be 1000 times, and calculate the estimate mean of sample, and hence, there will be 1000 mean values
- For all estimated mean, sort the value and find the percentile with confidence level of 99%. Note that the active zones may consist of two condition i.e. very low level of pressure and very high level of pressure. Therefore, the test will be based on 1 % and 99% data.
- Test the mean of MSLP on day-1 and day-2 with mean of bootstrap sample. If the sample mean of MSLP is less than the 1 percent-th data, then the MSLP is significantly low at this zone. Moreover, if the mean of MSLP is above the the 99th percent of the bootstrap

**Figure 4.** Mean Residual Life Plot.  
**Figure 5.** Rainfall with the threshold.
data, then the MSLP is significantly high. The zone is said to be inactive zone (normal condition) if the mean of MSLP lies between 1

An illustration of bootstrap result is given on Figure 6. The figure performs histogram of mean bootstrap of MSLP in Indonesia on one coordinate, together with the vertical line showing the lower and upper limit as the basis for the significant test. From the figure, the lower bound is 92000 and the upper zone is 100500. If the mean sample of MSLP on day-1 lies outside the interval (let say lower than 92000), then the region is active zone. It simply means that on one day prior to the occurrence of extreme rainfall in Indramayu, this region will show MSLP that is significantly different pattern compared to other regions/zones. Basically we test for the following hypothesis:

\[
\begin{align*}
H_0 : & \quad \mu = \mu_0 \quad (\text{the MSLP on the corresponding coordinate is not active zone}) \\
H_1 : & \quad \mu \neq \mu_0 \quad (\text{the MSLP on the corresponding coordinate is active zone})
\end{align*}
\]

Figure 6. Illustration of bootstrap result in one grid point.  

Using bootstrap resampling method, the active zones can be plotted into a composite map as performed in Figure 7. The Figures revealed that most of the regions in Sumatra and few part of Kalimantan have plot with red color, which show that the MSLP on day-1 is significantly high on that region, while over Java and Bali, the color is black showing that the MSLP on that region is significantly low. We can see also that the Sea Pressure at the atmospheric level on most of the area above Java Sea does not show any significant different patterns with usual days. To deal with this, the forecaster can only focus on observing the active zones to predict the occurrence of extreme rainfall in Indramayu.

5. Conclusion

Indramayu is a region in Indonesia with low intensity of rainfall, however the variances of the rainfall intensity is very high, in a year, the average of light rainfall event is 43 days, normal rainfall is 29 days and heavy rainfall is 6 days. The threshold to characterize extreme rainfall event is 50 mm, which fits the definition of BMKG Indonesia. On day-1 and day-2 prior to the occurrence of the extreme rainfall in Indramayu, the MSLP on Sumatra and some Kalimantan regions will show a very high level of MSLP. Meanwhile, the MSLP in Bali and east part of Java will be very low.

References

[1] Dai A 2011 Drought under global warming: a review Climate Change 2(1) pp 45–65
[2] Efron B 1979 A Note on Bootstrapping the Sample Median Ann. Statist 7(1)
[3] Efron B and Tibshirani R J 1993 An Introduction to the Bootstrap Chapman and Hall, London
[4] Grotjahn R and Faure G 2008 Composite Predictor Maps of Extraordinary Weather Events in the Sacramento California Region Weather and Forecasting 23 pp 313–335
[5] Hansen J, Sato M and Ruedy R 2012 Perception of climate change Proceeding National Academic of Sciences US
[6] Min S, Zhang X and Hegerl G C 2011 Human contribution to more-intense precipitation extremes Nature 470 pp 378–381
[7] Pall P, Aina T, Stone D A, Stott P A, Nozawa T, Hilberts A G 2000 Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn Nature 470 pp 328–385
[8] Lall U and A Sharma 1996 A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series Water Resour. Res. 32(3) pp 679–693
[9] Mudelsee M 2011 The Bootstrap in Climate Risk Analysis Extremis pp 44–58
[10] Orlowsky B, Bothe O, Fraedrich K, Gerstengarbe F W, Zhu X 2010 Future climates from bias-bootstrapped weather analogues: an application to the Yangtze river basin Journal of Climate 23(13) pp 3509–3524
[11] Rahmstorf S and Coumou D 2011 Increase of extreme events in a warming world. Proc. Natl Acad. Sci. USA 108 pp 17905–17909