Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 1 (2016), pp. 1151-1159

http://www.ripublication.com

Return Level Value Modeling of Rainfall and GCM for Extreme Rainfall Prediction in Indramayu Regency

Tia Fitria Saumi
Bogor Agricultural University Bogor, West Java, Indonesia.

Aji Hamim Wigena
Bogor Agricultural University Bogor, West Java, Indonesia.

Anik Djuraidah
Bogor Agricultural University Bogor, West Java, Indonesia.

Abstract

Quantification of rainfall extreme is important for planting plan purpose and common measure extreme event. One of method to predict rainfall extreme is return level modeling, because return level can describe extreme value that happens in certain period. Return level of rainfall at 15 stations in Indramayu are used as response and return level of general circulation model (GCM) as predictor. The return level can be obtained by parameters of general pareto distribution (GPD) and copula. GPD is used when data is single or univariate, and then the copula is used when data are multivariate and have spatial dependence. Spatial dependency can be detected through f-madogram plot of rainfall and GCM. The best model based on the smallest RMSEP produced by the return level modelling from rainfall of 15 stations in response and GCM in predictor.

Keywords: statistical downscaling, generalized pareto distribution, copula, return level value modeling, partial least square regression.

I. Introduction

The research about extreme rainfall prediction has developed for many years. Prang (2005) identified the extreme rainfall in Dramaga station by using block maxima, and generalized extreme value distribution (GEVD) approximates this sample maxima distribution. Irfan (2011) identified the extreme rainfall in the same station using peak over threshold (POT), the extreme value fits generalized pareto distribution (GPD) asymptotically. Block maxima approach and POT approach were used for univariate extreme rainfall.

The study of multivariate extreme rainfall has also well developed in climate change researches. Sari (2013) applied copula to identify the extreme rainfalls in 15 stations in Indramayu regency. The extreme rainfall from two or more stations needs the spatial dependences assumption. Sari (2013) evaluated the spatial dependences among extreme rainfall in 15 stations by using F-madogram.

Apart from using the rainfall data, the information of global climate has become an important tool in extreme rainfall prediction. The information of atmosphere circulation is gained from global circulation model (GCM) which has the global scale data with high multicolinearity. The method that is used to gain local scale information from GCM output is statistical downscaling (SD). The modeling of extreme rainfall prediction in this research was built using the return level value of rainfall and GCM output. SD approach that is used to overcome the large data and high multicolinearity was partial least square regression (PLSR).

II. Literature Review

2.1 Univariate Extreme Value

There are two approaches for understanding univariate extreme rainfall distribution, i.e. block maxima and peak over threshold (POT). Generalized extreme value distribution (GEVD) approximates the sample maxima and the data exceed above threshold follow generalized pareto distribution (GPD) asymptotically. GEV distribution function follows,

$$H_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left\{-\left[1+\ \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\}, \text{jika } \xi \neq 0 \\ \exp\left[-\exp\left(\frac{\mu-x}{\sigma}\right)\right], \text{jika } \xi = 0 \end{cases} \tag{1}$$

where, $H_{\mu,\xi,\sigma}$ *(x)* is GEV distribution, μ is location parameter, σ is scale parameter and ξ is shape parameter. The shape parameters $\xi \to 0$, $\xi > 0$ and $\xi < 0$ correspond respectively to Gumbel, Fr´echet and Weibull families. GPD distribution function follows,

$$G_{\xi,\sigma}(x) = \begin{cases} 1-\left(1+\frac{\xi x}{\sigma}\right)^{\frac{-1}{\xi}}\ jika\ \xi\ \neq 0 \\ 1-exp\left(-\frac{x}{\sigma}\right)\ jika\ \xi = 0 \end{cases} \tag{2}$$

where, $G_{\xi,\sigma}(x)$ is GPD distribution, σ is scale parameter and ξ is shape parameter.
Extreme value prediction can be obtained by calculating the return level value of GEV or GPD distribution. Return level value is the level expected to be exceeded in one out of *k* periods of length *p* (Gilli and Kellezi). The return level of GEV distribution follows,

$$\hat{R}_p^k = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}_i}{\hat{\xi}_i}\left[\left(-\ln\left(1-\frac{1}{k}\right)\right)^{-\hat{\xi}_i} - 1\right], \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}_i \log\left(-\log\left(1-\frac{1}{k}\right)\right), \hat{\xi} = 0 \end{cases} \tag{3}$$

and return level of GPD distribution follows,

$$\hat{R}_p^k = u + \frac{\hat{\sigma}_i}{\hat{\xi}_i}((k\delta u)^{\xi}\text{-}1) \tag{4}$$

where u is threshold, *k* is period, δu estimated by n/N, *n* sum of exceeded value and N is sum of data.

2.2 Multivariate extreme values

Max-stable process and copula are tools in applied extreme spatial analysis. Max-stable processes provide a natural generalization of extremal dependence structures in continuous spaces (Padoan et. al. 2010). Let X_{ij} be *n* independent replications of a continuous stochastic and *d* dimension, $i= 1,2,....,n,\ j=1,2,..d.$ Assume that there are sequences of continuous functions $a_n>0$ and $b_n \in R$ such that,

$$\lim_{n\to\infty} \frac{max_{i=1}^{n} X_{ij} - b_n}{a_n} = Z(j) \tag{5}$$

If this limit exists, the limit process $Z(j)$ is a max-stable process (de Haan 1984).
Copula is the non parametric estimator that plays a significant role in measuring the dependences between random vector components. Let $X = (X_{1,......,}X_d)$ are random field, and *F* is distribution function of *X*. Function of multivariate variable C in $[0,1]^d$, i.e

$$G\ (x_1, \ldots\ldots x_d) = C\{F_1(x_1), \ldots\ldots F_d(x_d)\}. \tag{6}$$

Where, C is copula and $F_1, F_2, F_3, \ldots F_d$ is marginal distribution of $X_1, X_2, X_3, \ldots, X_d$. Estimators of copula parameters are obtained using *pseudo maximum likelihood Estimation* (PMLE) (Weiβ 2010). Extreme value prediction of copula can be calculated by return level equation of GEVD in equation (3).

2.3 F-madogram

Spatial dependences among extreme observations in several locations can be identified. Madogram evaluates the dependence among extreme observations located in two or more separated regions. Cooley et. al. (2006) proposed the F-madogram, which transforms the random variable by applying the cdf and it is finite for any distribution. If $Z(x)$ is stationer max-stable process and isotropic with $Frechet$ marginal distribution, then the F-madogram is:

$$v_F(h) = \frac{1}{2} E\left|F\big(Z(x+h)\big) - F\big(Z(x)\big)\right|. \tag{7}$$

where $\hat{v}_F(h)$ is F-madogram, h is distance between two locations, $Z(x+h)$ is observation at $x+h$ location and $Z(x)$ is observation at x location.

2.4 Partial Least Square Regression

Partial Least square Regression (PLSR) is the useful tool when the factors are many and highly collinear. PLS extract latent variables that accounting for as much of the manifest factor variation. This method results the extracted factors (*X*-scores and *Y*-score) based on the strong relationship between successive pairs of scores. The X-scores are used to predict the Y-scores, and then the predicted Y-scores are used to construct predictions for the responses. There is no limit and the measured factor or select it in PLSR. It will be useful tool when the prediction is the goal.

III. Methodology

3.1 Data

There are two types of data used, namely:

1. Monthly local rainfall data used is the data from climatology station Indramayu regency. Rainfall data from this station are in the period 1979 to 2008 (the data length of 30 years) so there are 360 observation data.
2. The GCM data used are monthly rainfall data from Climate Model Intercomparison Project (CMIP5) issued by KNMI, Netherlands in 1979 to 2008. The area of a square area measuring 8 x 8 grid around Indramayu district, so there are 64 variables

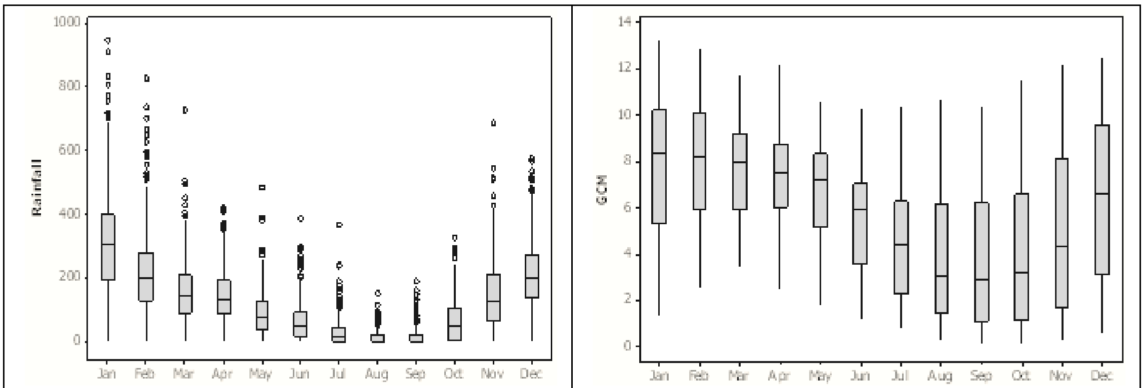
3.2 Method

1. Exploration Indramayu district rainfall data and GCM output.
2. Creating average rainfall and extreme rainfall in 15 stations in Indramayu.
3. Creating extreme value of GCM output or extreme GCM output.
4. Predicting the GPD parameter of rainfall average.
5. Calculating spatial dependence from extreme rainfall at 15 stations and extreme GCM using F-madogram.
6. Predicting the Copula prameter of extreme rainfall and Extreme GCM.
7. Determine return level value of GPD from rainfall average, return level value of copula from extreme rainfall and return level value of copula from extreme GCM.
8. Modeling the return level values based on PLSR.
   a. Return level value of extreme rainfall as response and return level value of extreme GCM as predictor.
   b. Return level value of rainfall average as response and return level value of extreme GCM as predictor.
9. Predicting the extreme value.
10. Calculating the evaluation of extreme value prediction using RMSEP and correlation.

IV. Results and Discussions

4.1 Data Exploration

Figure 1(a) below is the boxplot of rainfall in Indramayu district from 1979 to 2008, and Figure 1(b) is boxplot of GCM output. The two boxplots have different pattern, the rainfall formed monsoon and GCM output formed seasonal pattern. The monsoon pattern shows the largest variance of rainfall occurred in January, as for August and September were very low variance. Rainy season occurred in October until March, and dry season occurred between April and September. Outlier as extreme rainfall occurred in all months.



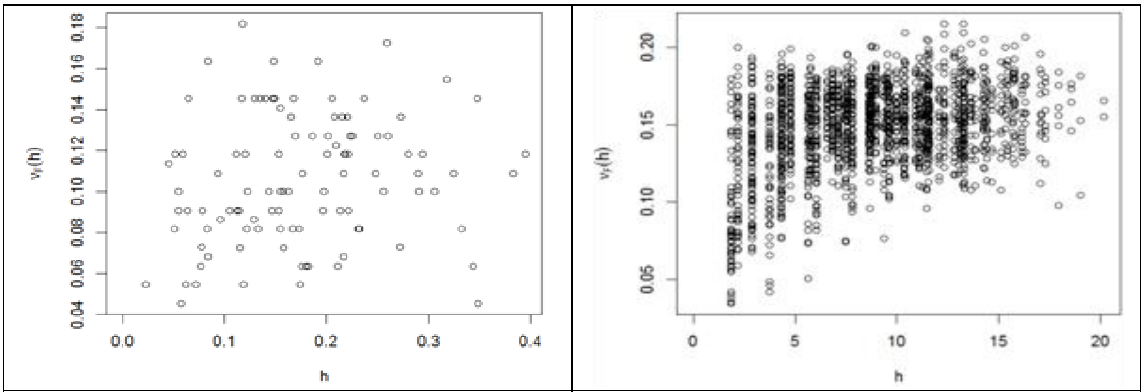
Month	month
(a)	**(b)**

Figure 1: Boxplot of Rainfall (a) and GCM output (b)

Exploration of GCM output uses the boxplot on Figure 1(b) shows that the highest average of GCM output occurred in January and February, whereas the lowest average occurred in September. The largest variance occurred in November, whereas the lowest variance occurred in April. There is no outlier occurred at GCM output.

Spatial Dependences

The spatial dependence is an assumption in multivariate extreme study. Spatial dependences of extreme rainfall at 15 stations in Indramayu and extreme GCM were identified using F-madogram plot. Figure 2(a) is F-madogrm plots of extreme rainfall at 15 stations. Its plots are positive values and increase monotonously, this shows that extreme rainfall have spatial dependences among the stations. Figure 2 (b) is F-modogram plot for extreme GCM output. F-madogram plots of extreme GCM output have positive values and increase monotonously clearly. Thus F-madogram plots said there are spatial dependences between extreme GCM and its grid.



(a)	(b)

Figure 2: F-madogram and extreme rainfall in 15 stations in Indramayu (a) and extreme GCM output (b)

Parameter Estimation

GPD parameter estimates of rainfall average are scale parameter (σ) and shape parameter (ξ). Copula parameter estimate of extreme rainfall at 15 stations are location parameter (μ), scale parameter (σ) and shape parameter (ξ). Location parameter describes the location of centralization data, scale parameter states the variance of data, and shape parameter describes the tail distribution characteristic. The GPD parameters estimates and copula parameter estimates in Table 1 bellow were analyzed in ten analysis period.

Table 1: GPD parameter estimates of average rainfall and copula parameter estimates of extreme rainfall

Data Analyze	GPD parameter estimates		Copula parameter estimates				
	ξ	σ	μ		σ		ξ
			Min	Max	Min	Max	
1 Jan 1979-31 Des 1998	-0.04	98.27	313.61	349.25	71.99	104.14	-0.02
1 Jan 1979-31 Des 1999	-0.04	95.95	317.21	340.31	79.7	96.51	-0.02
1 Jan 1979-31 Des 2000	-0.03	95.73	319.57	343.99	55.3	115.38	0.00
1 Jan 1979-31 Des 2001	-0.04	95.26	314.67	341.22	65.92	117.56	-0.02
1 Jan 1979-31 Des 2002	-0.04	98.73	316.77	340.65	69.56	101.56	0.00
1 Jan 1979-31 Des 2003	-0.04	97.04	316.86	330.97	65.07	116.97	-0.03
1 Jan 1979-31 Des 2004	-0.02	97.67	322.68	341.75	55.89	146.32	-0.03
1 Jan 1979-31 Des 2005	-0.02	96.69	311.71	335.91	71.04	126.23	-0.01
1 Jan 1979-31 Des 2006	-0.02	96.29	321.81	329.99	77.12	106.34	-0.04
1 Jan 1979-31 Des 2007	-0.03	98.07	324.08	337.6	72.29	134.64	-0.03

The shape parameter estimate of two data was similar, the values between-0.04 until-0.01. It means the average rainfall and extreme rainfall had same characteristic of tail distribution. The negative values mean their tail was finite. The scale parameter estimates of average rainfall were 95.00-98.00, its variance tend to homogenous. The scale parameter estimates of extreme rainfall at 15 stations about 55.00-145.00. Thus values showed that the variance of extreme rainfall in 15 stations were not homogenous. Location parameter estimate of copula showed the centralization of extreme rainfall between 310-350 mm.
Extreme GCM outputs have three parameter estimates, location parameter, scale parameter and shape parameter estimates, it is given in Table 2. The minimum value of location parameter estimate was 8.00 and the maximum value was 14.00. The scale parameter estimates about 1.88-2.20. The values showed that the extreme GCM outputs had similar data centralization and homogenous variances. The shape parameter estimates showed that extreme GCM output have finite tail distribution.

Table 2: Copula parameter estimates of extreme GCM outputs

Data Analyze	Copula parameter estimates				
	μ		σ		ξ
	Min	Max	Min	Max	
1 Jan 1979-31 Des 1998	8.047	14.982	1.877	2.202	-0.254
1 Jan 1979-31 Des 1999	8.063	14.987	1.885	2.203	-0.263
1 Jan 1979-31 Des 2000	8.075	15.001	1.888	2.205	-0.262
1 Jan 1979-31 Des 2001	8.077	14.996	1.881	2.209	-0.257
1 Jan 1979-31 Des 2002	8.076	15.003	1.878	2.212	-0.246
1 Jan 1979-31 Des 2003	8.085	15.001	1.878	2.213	-0.228
1 Jan 1979-31 Des 2004	8.078	15.007	1.881	2.218	-0.221
1 Jan 1979-31 Des 2005	8.077	15.022	1.888	2.217	-0.214
1 Jan 1979-31 Des 2006	8.078	15.044	1.887	2.224	-0.212
1 Jan 1979-31 Des 2007	8.081	15.029	1.882	2.218	-0.208

Return Level Value

After the parameter estimation value was obtained, thus return level value can be calculated. Return level value describes the extreme event in particular period, it can say that return level value is an extreme value prediction. The best return level value based on smallest RMSEP. Table 1 is return level value of rainfall average, rainfall and extreme rainfall.
Table 3 shows that the return level values of rainfall average and extreme fainfal.l were underestimates. The best prediction was return level values of rainfall generally, because it had the smallest RMSEP than others. The return level values of extreme rainfall were overestimates, but it was useful to predict the most extreme rainfall. Return level value of extreme rainfall gave the closest prediction to the most extreme rainfall in January 1997. Table 2 is return level value from extreme GCM. Return level values of extreme GCM output from 64 grids were presented in minimum value,

maximum value and average. The table shows that return level value extreme GCM were constant.

Table 3: Return level value of average rainfall, rainfall and extreme rainfall

Analysis Period	Return Level Value		Actual Value	Time of Realization
	Rainfall average	Extreme Rainfall		
1 Jan 1979-1 Des 1994	275.821	540.49	424	Jan 95
1 Jan 1979-1 Des 1995	282.341	532.79	414	Jan 96
1 Jan 1979-1 Des 1996	283.293	553.69	582	Jan 97
1 Jan 1979-1 Des 1997	285.55	578	346	Nov 98
1 Jan 1979-1 Des 2000	282.06	555.49	326	Dec 01
1 Jan 1979-1 Des 2001	281.19	560.37	455	Jan 02
1 Jan 1979-1 Des 2002	286.1	545.85	247	Jan 03
1 Jan 1979-1 Des 2004	284.13	594.15	235	Feb 05
1 Jan 1979-1 Des 2005	281.69	575.81	409	Feb 06
1 Jan 1979-1 Des 2007	282.16	591.62	439	Mar 08
RMSEP	144.4	203.8		

Table 4: Return level value of Extreme GCM output

Analysis Period	Prediction Value		Average of Prediction value
	min	max	
1 Jan 1979-1 Dec 1998	11.461	18.988	15.224
1 Jan 1979-1 Dec 1999	11.459	18.956	15.208
1 Jan 1979-1 Dec 2000	11.482	18.979	15.230
1 Jan 1979-1 Dec 2001	11.486	18.999	15.243
1 Jan 1979-1 Dec 2002	11.523	19.064	15.293
1 Jan 1979-1 Dec 2003	11.603	19.148	15.375
1 Jan 1979-1 Dec 2004	11.626	19.190	15.408
1 Jan 1979-1 Dec 2005	11.667	19.238	15.453
1 Jan 1979-1 Dec 2006	11.674	19.283	15.478
1 Jan 1979-1 Dec 2007	11.681	19.273	15.477

Return Level Modeling

Prediction model of extreme rainfall was built from the return level of rainfall and GCM. Responses used in modeling were the return level value of average rainfall and extreme rainfall. Predictor used was return level value of extreme GCM output. Prediction model formed from the combination of return level values of rainfall and GCM output. Prediction model was evaluated by RMSEP and correlation. Criterion of the best model was based on the highest correlation and the smallest RMSEP.

Table 8: Correlation value and RMSEP and RMSE

Return level value		Correlation	RMSEP
Y	X		
Extreme rainfall	Extreme GCM	0.911	154.108
Rainfall average	Extreme GCM	0.999	123.648

The high correlation was resulted by return level value model of extreme rainfall and extreme GCM and return level value model of average rainfall and extreme GCM, they were 0.911 and 0.999. The best model was return level value model of average rainfall and extreme GCM because it has the smaller RMSEP was 123.648 than RMSEP of another model was 154.108.

Conclusion

Return level values of rainfall average gave the smaller RMSEP than return value of extreme rainfall. It means that GPD approach is better than copula approach. The return level modeling gave two models, return level of rainfall average and extreme GCM, and return level of extreme rainfall and extreme GCM. Both of models gave the hight correlation, 0.911 and 0.999. The best model of them was the second model because the it had the smaller RMSEP (123.648) than the first one (154.108).

References

[1] Buch-Larsen T, Nielsen JP, Guillen M, Bolance C. 2005. Kernel density estimation for heavy-tailed distribution using the Champernowne transformation. Statistics. 39(6):503-518. doi:10.1080/02331880500439782.

[2] Cooley D, Naveau P, Poncet P. 2006. Variograms for spatial max-stable random fields, *Lecture Notes in Statistics*. 187: 373-390, Springer, New York.

[3] Irfan M. 2011. General Pareto Distribution to Predict Extreme Rainfall (Case study: Rainfall period 2001-2010 in Darmaga Station) [skripsi]. Bogor: Bogor Agricultural University.

[4] Këllezi E M. 2006. An application of extreme value theory for measuring financial risk. *Computational Economics*. 27: 207-228, Springer, New York.

[5] Prang JD. 2006. General Extreme Value in Rainfall Phenomenon [thesis]. Bogor: Postgraduate Program, Bogor Agriculture University.

[6] Sari FM. 2012. Extreme Rainfall Forecasting in Spatial (Case Study: Monthly rainfall in Indramayu Regency) [thesis]. Bogor: Postgraduate Program. Bogor Agriculture University.

[7] Wigena AH. 2006. Statistical Downscaling modeling with Projection Pursuit Regression for Monthly Extreme Rainfall Forecasting in Indramayu. Bogor: Postgraduate Program, Bogor Agriculture University.