

Quantile Regression with Elastic-net in Statistical Downscaling to Predict Extreme Rainfall

Tri Budi Novia Cahyani¹, Aji Hamim Wigena, Anik Djuraidah

*Department of Statistics,
Bogor Agricultural University, Indonesia.*

Abstract

Rainfall prediction is necessary since extreme rainfall has a big impact to the environment. A method commonly used to predict rainfall is statistical downscaling. This technique develops a functional relation between local scale rainfall data and global scale General Circulation Model (GCM) output data. The multicollinearity problem in GCM output data can be overcome by the elastic-net regularization which is the combination of ridge and lasso regularizations. The elastic-net is better than only lasso or ridge regularization especially if among predictors are highly correlated. Quantile regression can be used to predict the extreme rainfall. This paper discusses quantile regression in statistical downscaling with elastic-net regularization to predict extreme rainfall. The results show that the extreme rainfall which occur in January and December 2013 was predicted properly at $Q(0.75)$ and $Q(0.90)$ respectively, and elastic-net penalized quantile regression model was consistent.

AMS subject classification:

Keywords: Elastic-net Regularization, General Circulation Model, Quantile Regression, Statistical Downscaling.

1. Introduction

Rainfall has an important role in many sectors, especially in agriculture. The higher rainfall can cause flood that will cause give negative impacts for plants. Rainfall prediction becomes necessity to reduce those negative impacts that may occur. Statistical downscaling using quantile regression can be used to predict such the extreme rainfall.

¹Corresponding Author: tribudi.nc@gmail.com

Statistical downscaling develops a functional relation of General Circulation Models (GCM) output and rainfall data. GCM output are global scale data with low spatial resolution while rainfall data are local scale data with high resolution [11]. The function is in the form of multiple regression, $\mathbf{y} = f(\mathbf{X})$ where \mathbf{y} is a vector of response variable such as rainfall data and \mathbf{X} is a matrix of explanatory variables such as GCM output data.

The problem arises because the GCM output data are not only large dimension but also high correlation or multicollinearity. Principal component analysis is usually used to overcome both problems. Another way to overcome multicollinearity is to apply a regularization, such as ridge, lasso, or elastic-net regularization.

Hoerl and Kennard [3] introduced ridge regularization. However, many variables are still included in the model since this regularization does not set any regression coefficient to 0 and does not give an easy interpretation of the model. Tibshirani [10] proposed another regularization called Least Absolute Shrinkage and Selection Operator (lasso). Lasso shrinks some regression coefficients towards zero and some coefficients may be exactly to zero. Although the lasso produces model with fewer explanatory variables and the interpretation becomes easier, ridge regularization is still better in prediction performance for $n > p$ situations [10]. Considering these limitations, Zou and Hastie [13] proposed the elastic-net regularization.

Quantile regression can be used to measure the effect of the explanatory variables at the various quantile values (top, centre, or bottom) of the data distribution. Djuraidah & Wigena [2] built quantile regression model to explore rainfall in Indramayu and concluded that the quantile regression can be used to detect both extreme dry at 0.05th quantile ($Q(0.05)$) and extreme wet at 0.95th quantile ($Q(0.95)$). Mondiana [5] used quantile regression with principal component analysis to overcome multicollinearity. Sari [8] also developed quantile regression to predict extreme rainfall using functional principal component analysis. Soleh [9] developed modelled rainfall using the linear model based on Gamma and generalized Pareto distribution with lasso regularization. Santri [7] applied lasso regularization in quantile regression.

The purpose of this study is to develop a statistical downscaling model to predict extreme rainfall with elastic-net regularized quantile regression.

2. Preliminary Notes

2.1. Quantile Regression

Quantile regression was first introduced by Koenker and Bassett [4]. Quantile regression can estimate the response in various quantile of the data distribution. It is very useful when the purpose of the modelling only focuses on a particular part of the data distribution, such as modelling extreme value on the top quantile of the data distribution.

The τ th quantile of a random variable Y is defined as

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\}$$

where $F_Y(y) = P(Y < y)$ is the distribution function of Y and $\tau \in [0, 1]$.

In the regression model $\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}$ the coefficient estimators is defined as a solution of optimization problem:

$$\min_{\beta_0, \boldsymbol{\beta}} \left[\sum_{i \in \{i: y_i \geq \beta_0 + \mathbf{x}'_i \boldsymbol{\beta}\}} \tau |y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta}| + \sum_{i \in \{i: y_i < \beta_0 + \mathbf{x}'_i \boldsymbol{\beta}\}} (1 - \tau) |y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta}| \right].$$

The problem can also be written as

$$\min_{\beta_0, \boldsymbol{\beta}} \left[\sum_{i=1}^n \rho_\tau(y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta}) \right]$$

where

$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u > 0 \\ (\tau - 1)u & \text{if } u \leq 0 \end{cases}$$

is the loss function for quantile regression.

2.2. Elastic-net Regularization

Zou and Hastie [13] proposed a regularization technique which combine ridge and lasso regularization. The technique is called elastic-net. The elastic-net coefficient estimates are obtained by minimizing the regression loss function using elastic-net penalty:

$$\sum_{j=1}^p [\alpha |\beta_j| + (1 - \alpha) \beta_j^2] \leq k$$

When $\alpha = 0$, then it becomes a ridge penalty $\left(\sum_{j=1}^p \beta_j^2 \right)$ and when $\alpha = 1$, it becomes

lasso penalty $\left(\sum_{j=1}^p |\beta_j| \right)$. Elastic-net does both the coefficient shrinkage in a group of

high correlated variables as in ridge and variable selection as in lasso. The coefficient estimator in elastic-net regularized quantile regression is defined as:

$$\mathbf{b} = \arg \min_{\beta_0, \boldsymbol{\beta}} \left[\sum_{i=1}^n \rho_\tau(y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta}) + \lambda \sum_{j=1}^p [\alpha |\beta_j| + (1 - \alpha) \beta_j^2] \right]$$

3. Data and Methods

This study uses monthly rainfall data of Indramayu ZOM 79 in 1981 to 2013 as response variable and GCM precipitation data from Climate Model Intercomparison Project (CIMP5) obtained from <http://climexp.knmi.nl> at $1.25^\circ - 18.75^\circ$ S and $101.25^\circ - 118.75^\circ$ E as

explanatory variables. The area consist of 8×8 grid, so it contains 64 variables. Data is divided into training data (1981-2012) and testing data (2013).

GCM variables are shifted based on optimum time lag with the highest cross correlation function (CCF) value with rainfall variable. The purpose is to get explanatory variables which highly correlated with response variable [6]. SD modelling using elastic-net penalized quantile regression is performed on training data by first selecting the optimum tuning parameters using cross validation (CV). The optimum parameters selected are parameters with minimum CV error. Quantile regression models are constructed using the regularization parameters at $Q(0.75)$, $Q(0.90)$, and $Q(0.95)$, where extreme values are mostly occurred. The models are then used for prediction using testing data (2013) and calculating root mean square error of prediction (RMSEP) and correlation between actual observations and the predicted values to analyse the goodness of the model. Analysis are performed by R software using “hqreg” package [12].

4. Results and Discussion

In period 1981-2013, the maximum value of rainfall was 498 mm which occurred in January 2006. The pattern of rainfall in Indramayu is showed by boxplot in Figure 1. The pattern of monthly rainfall in Indramayu resembles the letter “U”. The rainy season occurs at the beginning and end of the year, and the dry season occurs at the middle of the year. Extreme values of rainfall mostly occur in January and December, the lowest values are around July to September.

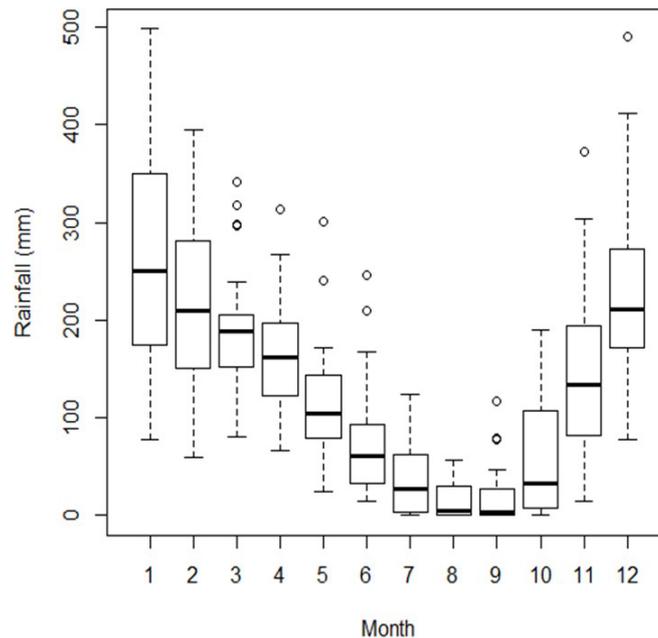


Figure 1: The pattern of rainfall in Indramayu (1981–2013).

In elastic-net penalized quantile regression, there are two parameters used. The parameters are $(\lambda \times \alpha)$ for lasso penalty and $(\lambda \times [1 - \alpha])$ for ridge penalty, $\lambda \in [0, 1]$. The values of α and λ used in model are determined by CV. Optimum parameters and RMSE of the elastic-net penalized quantile regression are presented in Table 1. It shows that the values of lasso parameters (0.94×10^{-3} , 5.40×10^{-3} , 2.43×10^{-3}) are greater than the values of ridge parameters (4.04×10^{-4} , 6.00×10^{-4} , 2.70×10^{-4}) for each quantiles.

Table 1: Elastic-net parameters (α and λ), CVE, and RMSE.

Quantile	α	λ	Lasso	Ridge	CVE	RMSE
$Q(0.75)$	0.7	1.35×10^{-3}	0.94×10^{-3}	4.04×10^{-4}	22.57	17.77
$Q(0.90)$	0.9	6.00×10^{-3}	5.40×10^{-3}	6.00×10^{-4}	13.22	21.44
$Q(0.95)$	0.9	2.70×10^{-3}	2.43×10^{-3}	2.70×10^{-4}	7.87	27.44

The models are then used for testing data to predict extreme rainfalls. RMSEP and correlation obtained from elastic-net penalized quantile regression models are showed in Table 2. The lower the RMSEP, the higher the correlation of the prediction. Figure 2 shows that elastic-net penalized quantile regression models can estimate the quantile of the actual data very well. The model provides prediction value of rainfall with similar pattern with the actual, which resembles the letter “U”. Some actual points falls right in prediction lines.

Table 2: RMSEP and correlation for each quantile model.

Quantile	RMSEP	Correlation
$Q(0.75)$	14.22	9.90×10^{-1}
$Q(0.90)$	14.14	9.93×10^{-1}
$Q(0.95)$	20.07	9.89×10^{-1}

The plot of actual rainfall and predicted quantile rainfall in 2013 is presented in Figure 3. Elastic-net penalized quantile regression model can predict extreme values in January and December very well. Rainfall in January (323.38 mm) is predicted well at $Q(0.75)$ (335.74 mm), and in December (345.50 mm) is predicted well at $Q(0.90)$ (368.65 mm).

Another thing should be considered to build a good SD model is the relationship between the response and explanatory variables should not change in time [1]. Figure 4 presents RMSEP and correlation between the actual and predicted rainfall in 2010, 2011, 2012, and 2013. It shows that on average, the RMSEP at $Q(0.75)$ are lower than RMSEP at $Q(0.90)$, and RMSEP at $Q(0.90)$ are always lower than RMSEP at $Q(0.95)$ in all years. The correlation values are around 0.99 in all years. Those results show that

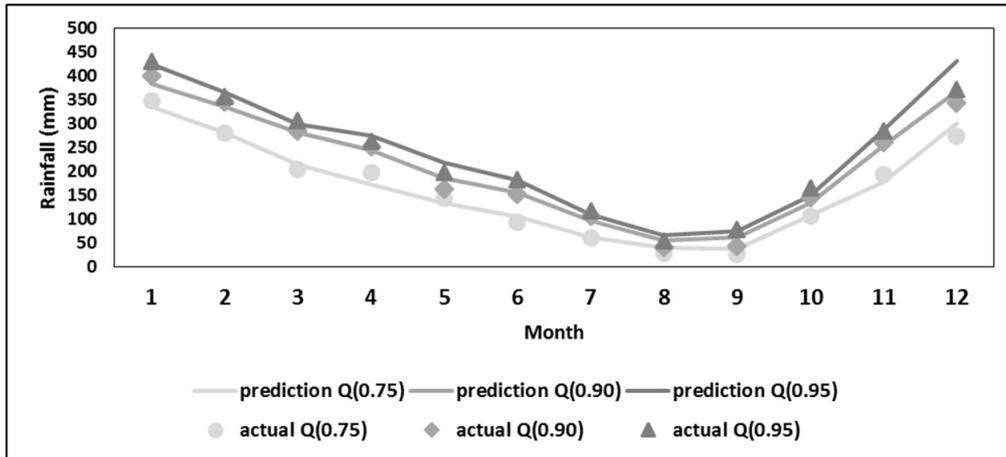


Figure 2: Actual and Predicted Quantile Rainfall.

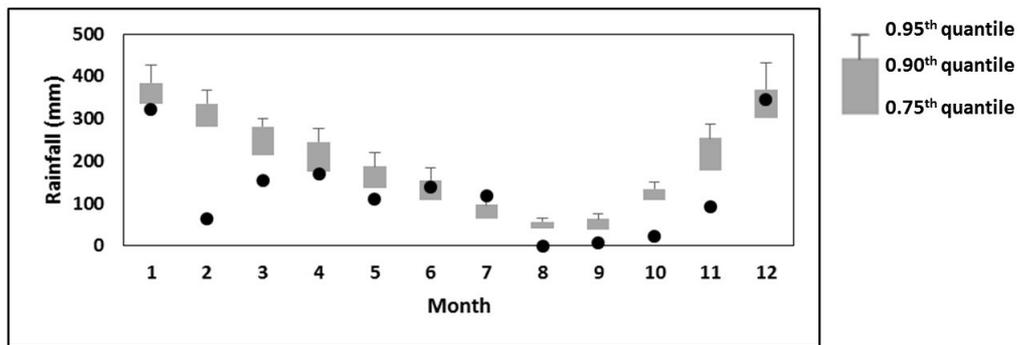
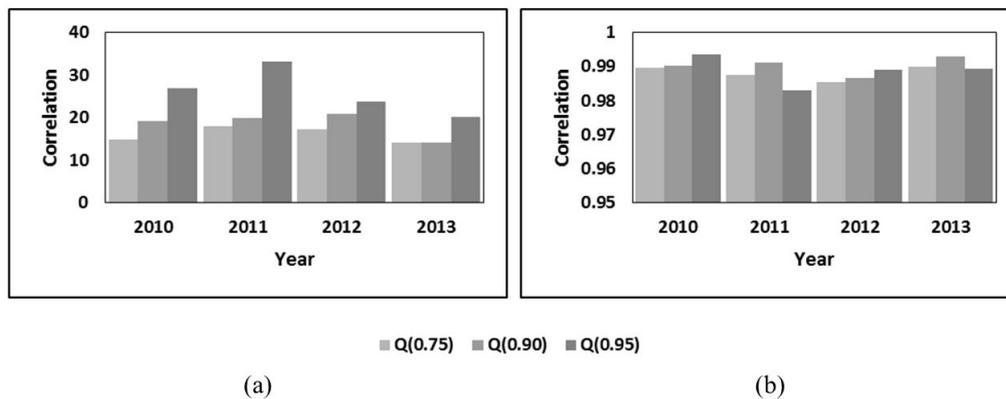


Figure 3: Actual rainfall (●) and predicted quantile rainfall in 2013.

the RMSEP and correlation from the model are consistent at various different time, so the elastic-net penalized quantile regression model is consistent.



(a)

(b)

Figure 4: RMSEP (a) and correlation (b) in 2010–2013.

5. Conclusion

Elastic-net regularized quantile regression models could predict extreme rainfall properly as in January and December at $Q(0.75)$ and $Q(0.90)$ respectively. The patterns of predicted and actual rainfall were similar and consistent at different times.

References

- [1] Busuioc, A., Chen, D., and Hellstrom, C., 2001, "Performance of Statistical Downscaling Models in GCM Validation and Regional Climate Change Estimates: Application for Swedish Precipitation," *International Journal of Climatology*, 21, pp. 557–578.
- [2] Djuraidah, A. and Wigena, A.H., 2011, "Regresi Kuantil untuk Eksplorasi Pola Curah Hujan di Kabupaten Indramayu," *Jurnal Ilmu Dasar*, 12(1), pp. 50–56.
- [3] Hoerl, A.E. and Kennard, R.W., 1970, "Ridge Regression, Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12(1), pp. 55–67.
- [4] Koenker, R. and Bassett, G. Jr., 1978, "Regression Quantiles," *Econometrica*, 46(1), pp. 33–50.
- [5] Mondiana, Y.Q., 2012, "Pemodelan Statistical Downscaling dengan Regresi Kuantil Untuk Pendugaan Curah Hujan Ekstrim," Thesis, Bogor Agricultural University, Indonesia.
- [6] Sahriman, S., Wigena, A.H., and Djuraidah, A., 2015, "Statistical Downscaling dengan Pergeseran Waktu Berdasarkan Korelasi Silang," *Jurnal Meteorologi dan Geofisika*, 16(1), pp. 19–24. Koenker R., Bassett G. Jr., 1978, "Regression Quantiles," *Econometrica*, 46(1), pp. 33–50.
- [7] Santri, D., 2016, "Statistical Downscaling Modeling with Quantile Regression Using Lasso to Estimate Extreme Rainfall," *Proceedings of the 7th SEAMS UGM International Conference on Mathematics and Its Applications*, <http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.4940862>.
- [8] Sari, W.J., 2015, "Pemodelan Statistical Downscaling dengan Regresi Kuantil Komponen Utama Fungsional untuk Prediksi Curah Hujan Ekstrim," Thesis, Bogor Agricultural University, Indonesia.
- [9] Soleh, A.M., Wigena, A.H., Djuraidah, A., and Saefudin, A., 2015, "Statistical Downscaling to Predict Monthly Rainfall Using Linear Regression with L1 Regularization (LASSO)," *Applied Mathematical Sciences*, 9(108), pp. 5361–5369.
- [10] Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B*, 58(1), pp. 267–288.
- [11] Wigena, A.H., Djuaridah, A., and Rizki, A., 2015, "Semiparametric modeling in statistical downscaling to predict rainfall," *Applied Mathematical Sciences*, 9(88), pp. 4371–4382.

- [12] Yi, C. and Huang, J., 2015, “Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss and Quantile Regression,” <http://arxiv.org/abs/1509.02957>.
- [13] Zou, H. and Hastie, T., 2005, “Regularization and Variable Selection via the Elastic Net,” *Journal of the Royal Statistical Society Series B*, 67(2), pp. 301–320.