Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2021, 2021:60
https://doi.org/10.28919/cmbn/5936
ISSN: 2052-2541

A NEW APPROACH TO STATISTICAL DOWNSCALING USING TWEEDIE COMPOUND POISSON GAMMA RESPONSE AND LASSO REGULARIZATION

MA'RUFAH HAYATI[1,2,*], AJI HAMIM WIGENA[1], ANIK DJURAIDAH[1], ANANG KURNIA[1]

[1]Department of Statistics, IPB University, Bogor 16680, Indonesia

[2]Department of Statistics, The University of Nahdlatul Ulama Lampung, Sukadana 34194, Indonesia

Abstract: Statistical Downscaling (SDS) is a technique in climatology to analyze the functional relationship of global scale GCM (Global Circulation Model) output data as predictor variables and local scale rainfall data as a response variable. Rainfall contains continuous and discrete components. The continuous component is related to the intensity of rainfall more than zero which can be assumed to be Gamma distribution while the discrete component is related to the occurrence of rain including zero which can be assumed to be Poisson distribution. A combination of both distributions is called Tweedie compound poisson gamma (TCPG). SDS modeling with TCPG response can be used to predict the occurrence of rain and also the rainfall intensity simultaneously. GCM output data generally contain multicollinearity problems which can be overcome by Lasso regularization. This study discusses SDS modeling which assumes TCPG distributed response and uses Lasso to predict some characteristics of rainfall such as the average number of daily rainfall events per month (λ), shape parameter (γ), the average intensity of daily rainfall per month ($\alpha\gamma$), probability of no rain event per month $\pi = \exp(-\lambda)$, the number of no rain per month $N\pi$. Based on the

*Corresponding author
E-mail address: aji_hw@apps.ipb.ac.id
Received May 3, 2021

smallest RMSEP and the high correlation of the actual and predicted data, the TCPG model with Lasso regulation is more reliable and needs to be considered for modeling rainfall than TCPG-generalized linear models and TCPG-principal component analysis.
Keywords: statistical downscaling; global circulation model; lasso; tweedie compound poissson gamma.

2010 AMS Subject Classification: 93A30.

1. INTRODUCTION

Statistical downscaling (SDS) is a technique to develop a relationship of global scale variables (covariate) and local scale variables (responses) [1]. The SDS technique is used to overcome the inability of low-resolution general circulation model (GCM) output to directly predict high-scale local climate conditions. Thus, the information of the GCM output is used to Predict rainfall.

Rainfall consists of two components, continuous and discrete components. The continuous component is used to measure the intensity of rainfall which is more than zero, while the discrete component includes the occurrence of rainfall with an intensity of zero because there is no recorded rainfall [2]. It is important to consider the estimation of the two components of rainfall for the sake of predicting the characteristics of rainfall.

Some proposed models for continuous components assumed the normal distribution in the linear model. This will lead to violations of assumptions because rainfall data are generally skewed to the right with gamma distribution. Generalized linear models are used as a solution. However, some rainfall events will be zero (no rain). These results in the distribution of gamma are less suitable for modeling the data. Proposing the distribution of Tweedie compound Poisson-gamma as a sum of continuous rainfall events can accommodate both components of rainfall simultaneously. This distribution is required because both components have important information for predicting rainfall in the future [3].

Precipitation data of GCM output are used as covariate variables in SDS modeling. The GCM output data has several problems, including multiple-dimensional explanatory variables, spatial correlation between grids, and multicollinearity between variables. This problem can be solved by several methods such as dimensional reduction, variable selection, and parameter shrinkage. The dimensional reduction can use the principal component analysis method (PCA) and the variable selection and parameter shrinkage often use the LASSO (least absolute shrinkage and selection operator) method.An example of the dimensional reduction method is the principal component analysis method (PCA). The lasso method has the advantage of selecting variables and estimating stable parameters in data analysis [10].

SDS research is generally carried out by modeling the two components of rainfall separately. SDS modeling by [10] uses gamma and Pareto distributions with lasso regularization. [12] conducted an SDS study using a rainfall response with a Gaussian distribution with fused lasso penalty. [11] used a general linear mixed model (GLMM) using a Gaussian response with lasso regularization. These models have not accommodated the two components of rainfall simultaneously in a single distribution. Therefore, this study will conduct SDS modeling that can handle several problems in rainfall modeling in one model, namely handling two rain components simultaneously using TCPG distribution, handling multicollinearity using Lasso regularization, and predicting several characteristics of rainfall in addition to rainfall intensity.

2. LITERATURE REVIEW

2.1. Statistical Downscaling and General Circulation Models

The SDS technique is used to overcome the inability of GCM (low resolution) to predict local-scale climate conditions (high resolution). The model usually used in SDS is as follows:

$$\boldsymbol{y}_{n\times 1} = f(\boldsymbol{X}_{n\times k})$$

where:

$\boldsymbol{y}_{t\times 1}$ is the rainfall data, $\boldsymbol{X}_{n\times k}$ is the precipitation of GCM output data, n is the number of observations, k is the number of covariate variables.

Some reasons that GCM output data cannot produce information for local scale directly are: (1) The description of spatial solutions about the structure of the earth surface, especially topography, is unclear; (2) Atmospheric hydrodynamics is nonlinear and there are nonlinear interactions between small scale grids; (3) Too many parameters that might not be right for small-scale processes. [4]. The illustration of Statistical Downscaling is shown in Figure 1.

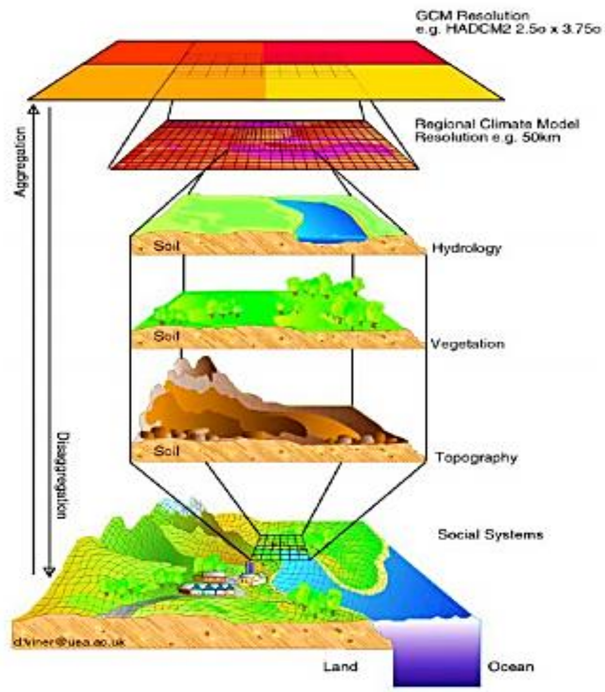

FIGURE 1. Illustration of Statistical Downscaling [15]

2.2. Tweedie Compound Poisson Gamma

The Tweedie model is a special member of the exponential dispersion model (EDM). The density function of EDM is defined by a two-parameter function, namely:

$$f_y\ (y|\theta,\phi) = a\ (y,\phi)\exp\left(\frac{1}{\phi}[y\theta - k(\theta)]\right) \tag{1}$$

where:

θ is canonical parameters in $\mathbb{R}$, $\phi > 0$ is dispersion Parameter $(0, +\infty)$, $k(\theta)$ is the cumulant function of the exponential dispersion model, $a\ (y,\phi)$ is the normalized quantity that has a size base that is independent of the parameter θ [5,6]

EDM has a mean characteristic, namely $\mu = k'(\theta) = dk(\theta)/d\theta$ and variance $Var(y) = \phi\ k''(\theta) = \phi Var(\mu)$, which can be calculated from the first and second derivatives of $k(\theta)$ w.r.t θ, because of the one-to-one mapping between θ and μ. $k''(\theta)$ can be denoted as a function of the mean μ, $k''(\theta) = Var(\mu)$. which is known as a variance function.

The Tweedie model specifies the power-law relationship between variance and mean $V(\mu) = \mu^p$. Tweedie is denoted by $T_{W_p}(\mu, \phi)$ wherein mean μ, dispersion parameter $\phi > 0$, and index or power parameter p. Based on the general form of EDM in equation (1). Tweedie's natural parameters and Tweedie's cumulative function are:

$$\theta_i = \theta(\mu_i) = \begin{cases} \frac{\mu_i^{1-p}}{1-p}, & p \neq 1 \\ log\mu_i, & p = 1 \end{cases} \quad \text{and} \quad b(\theta_i) = \begin{cases} \frac{\mu(\theta_i)^{2-p}}{2-p}, & p \neq 2 \\ log\mu_i, & p = 2 \end{cases}$$

Tweedie EDM with $\{\mu, \phi, p\}$ and $p \in (1,2)$ equivalent to describing Tweedie Compound Poisson Gamma (TCPG) which is parameterized with $\{\lambda, \alpha, \gamma\}$. Tweedie assumes that the arrival of an event has Poisson distribution and the intensity of each event has Gamma distribution. In the field of meteorology, Tweedie assumes Y as total monthly rainfall, N is the total number of rainfall events per month and y_i is precipitation in the i-event [7] mathematically written as:

$$P(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}, \forall n \in N_t$$

$$N = \sum_{t \geq 1} 1_{[t,\infty)}(t)$$

The amount of rainfall is represented as the total amount of rain from each rain event, say $(y_i)_{i\geq 1}$ is assumed to have a gamma distribution that is independent and identically to the time of rain:

$$Y = \begin{cases} \sum_{i=1}^{N} y_i & N = 1,2,3, \ldots \\ 0 & N = 0, \end{cases}$$

$y_i \sim Gamma(\alpha, \gamma)$ is the probability density function with mean $\alpha\gamma$ and variance $\alpha\gamma^2$. If $N = 0$ then $Y = 0$ and if $N > 0$ then $Y = \sum_i^N y_i$ [6]. Probability density function Y for $N > 0$ is:

$$f(y) = \begin{cases} \frac{\gamma^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\gamma y}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Y has the Poisson-Gamma distribution with the following parameters:

λ is the average number of rain events per month, γ is the shape of precipitation events, $\alpha\gamma$ is the average amount of precipitation per incident. Relationship between parameters $\{\lambda, \alpha, \gamma\}$ from TCPG and parameter $\{\mu, \phi, p\}$ from Tweedie model that is:

(2) $\begin{cases} \mu = \lambda\alpha\gamma \\ p = \frac{\alpha+2}{\alpha+1} \\ \phi = \frac{\lambda^{1-p}(\alpha\gamma)^{2-p}}{2-p} \end{cases}$ parameterization with $\begin{cases} \lambda = \frac{\mu^{2-p}}{\phi(2-p)} \\ \alpha = \frac{2-p}{p-1} \\ \gamma = \phi(p-1)\mu^{p-1} \end{cases}$

According to [8] the probability of no rain is:

(3) $$\pi = \Pr(Y=0) = e^{-\lambda} = \exp\left(\frac{\mu^{2-p}}{\phi(2-p)}\right)$$

The distribution of probability for data $Y > 0$ is:

(4) $$f(y|\mu,p,\phi) = a(y,\phi)\exp\left(\frac{1}{\phi}\left(\frac{y\mu^{1-p}}{(1-p)} - \frac{\mu^{2-p}}{2-p}\right)\right)$$

2.3. LASSO *(Least Absolute Shrinkage and Selection Operator)*

The least absolute shrinkage and selection operator (LASSO) was introduced by Thibsirani in 1996. The basic idea of the Lasso method is to add a penalty called L_1 regularization with the constraint $\sum_{j=1}^{p}|\beta_k| \le t$, $t \ge 0$ on the objective function. Penalty L_1 is used for variable selection by shrinking the linear regression parameter coefficients of highly correlated covariate variables so that they are close to zero or exactly zero. Suppose there is an input vector $\boldsymbol{X}^T = (x_1, x_2, \dots, x_k)$ used to predict the response of y using a linear regression model Y in the form:

$$\boldsymbol{Y} = \beta_0 + \sum_{j=1}^{k} x_j\beta_j + \boldsymbol{\varepsilon}$$

The least-squares method is used to estimate $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$ by minimizing the number of squares of the error [9]. The distribution of responses is not always normal. Thus, the Generalized linear model is used as a solution that uses the link function between the linear predictors and the mean response. Parameter estimation of the generalized linear model (GLM) with L_1 regularization has the following solutions:

(5) $$\underset{\beta_j}{\arg min}\left\{-\frac{\log\ L(\boldsymbol{y},\beta_j)}{n} + \lambda\sum_{j=1}^{p}|\beta_j|\right\}$$

where:

$L(\boldsymbol{y}, \beta_j)$ is the response of likelihood function, n is The number of observations, λ is the tuning parameters Lasso that is the the coefficient controller shrinkage parameters with $\lambda \geq 0$, β_j is the regression coefficient.

The estimate of β_j cannot be done deductively by calculus, but it uses optimization methods. Common numerical optimization methods that can be used to get optimum solutions is the iteratively re-weighted least square (IRWLS) method based on Newton Raphson approximation. By using the IRWLS, the likelihood function will get an estimate that reaches a maximum value if the Newton Raphson iteration converges [10].

3. METHODOLOGY

3.1. Data

This study uses monthly rainfall and precipitation data of GCM output in a period of January 1981 to December 2009 (348) months. Rainfall data (mm) at Cigugur station in West Java province is used as the response. The rainfall data were from Meteorology, Climatology, and Geophysical Agency. GCM output data as explanatory variables were from the national centers for environmental prediction (NCEP) in the form of a climate forecast system reanalysis (CSFR) model (website HTTP: //rda.ncar.edu).

Rainfall data in this study is divided into two parts, namely training data from January 1981 to December 2008 and testing data from January 2009 to December 2009 which can be seen in Table 1. The research scheme is the comparison of the TCPG-Lasso, TCPG-GLM, and TCPG-PCA on the data.

TABLE 1. Description of research data

Variable	Location of the rainfall observation station
$\boldsymbol{Y}_{\boldsymbol{n} \times \mathbf{1}}$	Rainfall from Cigugur station
$\boldsymbol{X}_{\boldsymbol{n} \times \boldsymbol{k}}$	Precipitation as The GCM (general circulation model) output data

3.2. TCPG-Lasso for Statistical Downscaling Model

There are three distributions in the Exponential Dispersion Model (EDM) that are found to be suitable for rainfall modeling, namely normal, gamma, and TCPG distributions. So, this study compared 3 different methods namely TCPG-Lasso, TCPG in terms of GLM (TCPG-GLM), TCPG with multicollinearity handling using principal component analysis (TCPG-PCA). This research assumes that the response of Y is TCPG distributed so that the model used in this study is as follows:

$$l(og(\mu)) = \beta_0 + \sum_{j=1}^{k} x_j \beta_j \tag{6}$$

Estimation of the coefficient parameters TCPG-Lasso model is done by minimizing the equation of penalty model (5) using a two-layer loop algorithm that incorporates the blockwise majorization descent method into an iteratively re-weighted least square (IRLS-BMD) proposed by [6]. TCPG-GLM and TCPG-PCA have the same parameter estimation method, namely the likelihood estimation method. The estimations of the parameter in this Research were assisted by software R version 3.6.0 with Tweedie package.

3.3. Best model selection criteria

This study uses Root Means Square Error Prediction (RMSEP) and the correlation between actual data and prediction data to determine the best parameter estimation and reduction method.

3.3.1. Root Means Square Error Prediction (RMSEP)

RMSEP is a method to measure the difference between the predicted value and the actual value defined as follows [1]:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n}} \tag{7}$$

3.3.2. Correlation of Y_i and $\hat{Y}_i$ (r)

Correlation is a method for measuring the closeness of a linear relationship between two random variables. Correlation used in this study is a correlation to measure the closeness of the relationship between actual data and prediction data which is defined as follows:

$$r = \frac{cov(Y_i, \hat{Y}_i)}{\sigma_{Y_i} \sigma_{\hat{Y}_i}} \tag{8}$$

where

Y_i is Rainfall data, $\hat{Y}_i$ is prediction data, n is the number of rainfall data. $\sigma_{Y_i} and\ \sigma_{\hat{Y}_i}$ are variance for rainfall and prediction data.

4. RESULTS

The distribution pattern of density, histogram, and boxplot of rainfall data are shown in Figure 2. The density plot in Figure 2 (a) shows that data pattern is right-skewed or positive continuous and exact zero. Thus, the rainfall data appear to be following the distribution characteristics of the TCPG. Other evidence that rainfall follows the characteristics of the TCPG distribution can be seen in Figure 2 (b) in the presence of a frequency of zero values and continuous positive values in the data. The box plot for the Cigugur station in Figure 2 (c) has a monsoon pattern where the lowest rainfall is between June and September.

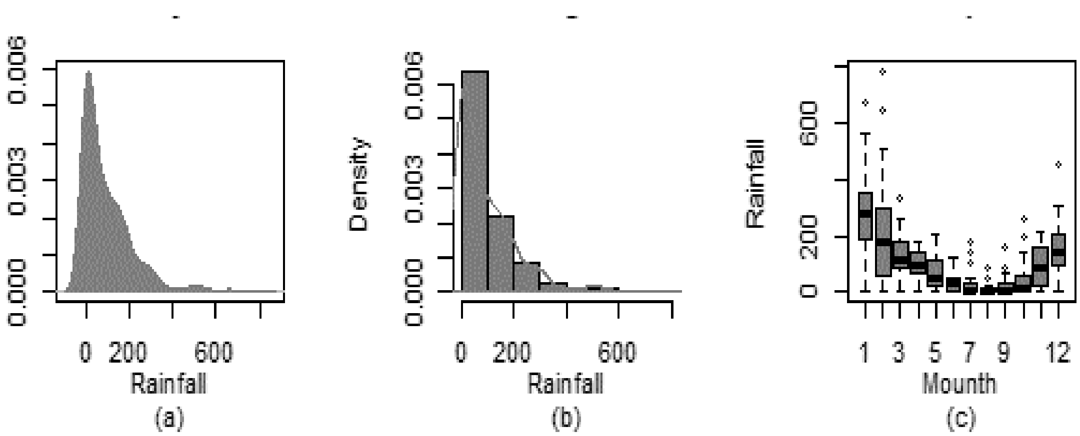

FIGURE 2. Density plot (a), Histogram (b), and Box plot (c) of monthly rainfall data in Cigugur station 1981-2009.

The index parameters of TCPG distribution are in the value range $1 < p < 2$. The index parameters and the dispersion parameters are estimated simultaneously, the results of which are described in Table 2. The selected index parameters which have the smallest likelihood profile can be seen in Figure 3. The index parameters are estimated first to determine the distribution of the data.

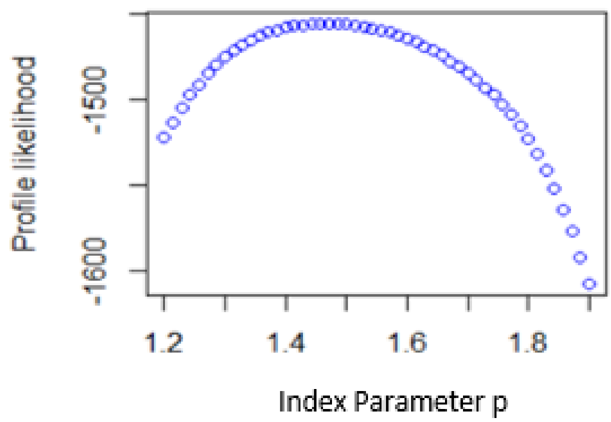

FIGURE 3. The plot of Profile Likelihood of Index Parameters for rainfall data at Cigugur station.

Figure 3 shows that the index parameter with the highest likelihood profile is around the values 1.4 to 1.5. Some of the parameters are estimated before modeling the TCPG-Lasso distribution that can be seen in Table 2, namely the estimated index parameter is 1.47, the dispersion (ϕ) is 16.31, α is 1.12 and the confidence interval for the index parameter (p) is 1.41 to 1.52. Confidence intervals for index parameters can be used to construct several TCPG models with different index parameters according to the value at the selected confidence interval so that the best model can be selected.

TABLE 2. Estimated parameters p, dispersion ϕ, α

Value parameter estimate	Cigugur
profile likelihood of $\boldsymbol{p}$	1.47
CI 95 % of	(1.41,1.52)
$\boldsymbol{\phi}$	16.31
$\boldsymbol{\alpha}$	1.12

From Table 2, it can be seen that the value of the p index parameter calculated from the profile likelihood for the data is between the values $1 < p < 2$. So, it is true that the rainfall data for the rain station of Cigugur has a TCPG distribution. The GCM output in the form of precipitation is used as a covariate variable which is in the form of a grid. The grid domain used is 5x8 which is equivalent to

40 covariates with a grid resolution of $2.5^0 \times 2.5^0$. Grid size is selected by looking for the correlation between existing grids in the selected domain with rainfall at the selected rain station. The selected grid is contiguous and correlates with the correlation value ≥ 0.3 [14]. The grids are spatially oriented and correlated so that there will be a violation of the multicollinearity assumption in the modeling. This multicollinearity problem can be handled by entering a penalty L_1, called penalty lasso. Therefore, the TCPG with lasso regulation modeling can be done for the data. Lasso involved in the model is a regression analysis method that can select variables and regulations to improve the accuracy of predictions and a statistical model that can be interpreted more easily.

The explanatory variable chosen will have a regression coefficient other than zero, while the explanatory variable that is not selected will be zero. The plot of the selected explanatory variables is shown in Figure 4.

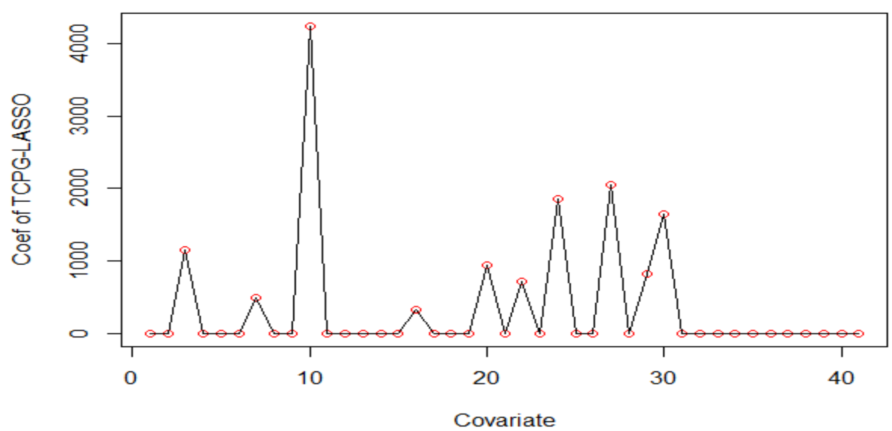

FIGURE. 4. The regression coefficients were selected for the Cigugur stations

The TCPG-Lasso model has nine covariate variables selected which are indicated by a regression coefficient value other than zero. Furthermore, the prediction and actual rainfall is visualized through the plot to see the comparison of the two rainfall data. The plots of the prediction data from the three models and the actual data can be seen in Figure 5.

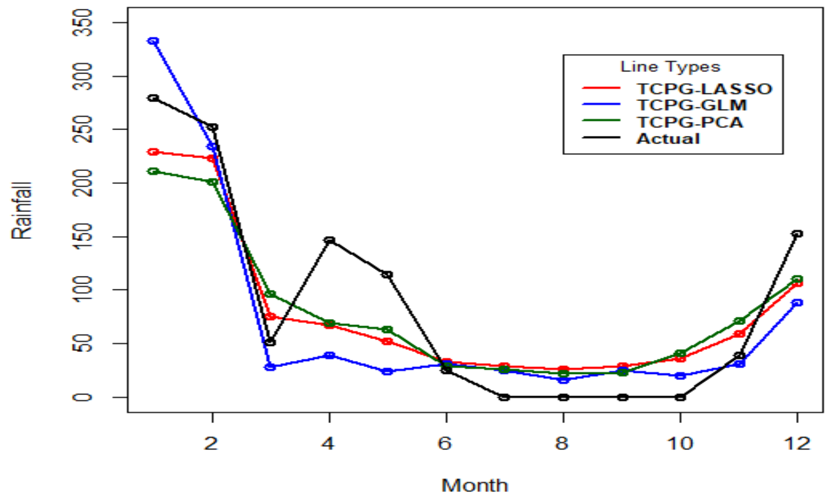

FIGURE 5. Rainfall Prediction for Cigugur Station in 2009

The plot of actual and predicted rainfall data for the Cigugur rain station from three methods was carried out in 2009 from January to December. The plot shows that the overall actual and predicted data patterns are similar to the actual data only slightly different in April and May. The TCPG-Lasso method is good at capturing high rainfall intensity. This shows that the model is good for the prediction of rainfall, both high and low rainfall. The measure of the goodness of the model is seen through the smallest Root Mean Square Error Prediction value and the highest correlation value between actual and predicted data of the three models. the goodness of fit model can be seen in Table 3.

TABLE 3. RMSEP and Correlation for Actual and Prediction rainfall

Method	RMSEP	r
TCPG-Lasso	23.33	0.93
TCPG-GLM	49.67	0.87
TCPG-PCA	44.87	0.93

Table 3 shows that the smallest RMSEP was obtained by the TCPG-Lasso model and the highest correlation value is obtained by the TCPG-Lasso and TCPG-PCA. Based on Table 2. The TCPG-Lasso model is more reliable for rainfall modeling than other models because it has the smallest RMSEP and the highest correlation.

TCPG distribution has several parameters that can be estimated to explain some characteristics of rainfall such as the average number of daily rainfall events per month (λ), shape parameter (γ), the average intensity of daily rainfall per month ($\alpha\gamma$), probability of no rain event per month $\pi = \exp(-\lambda)$, many events of no rain per month $N\pi$. That will be described in Table 4 and Figure 76

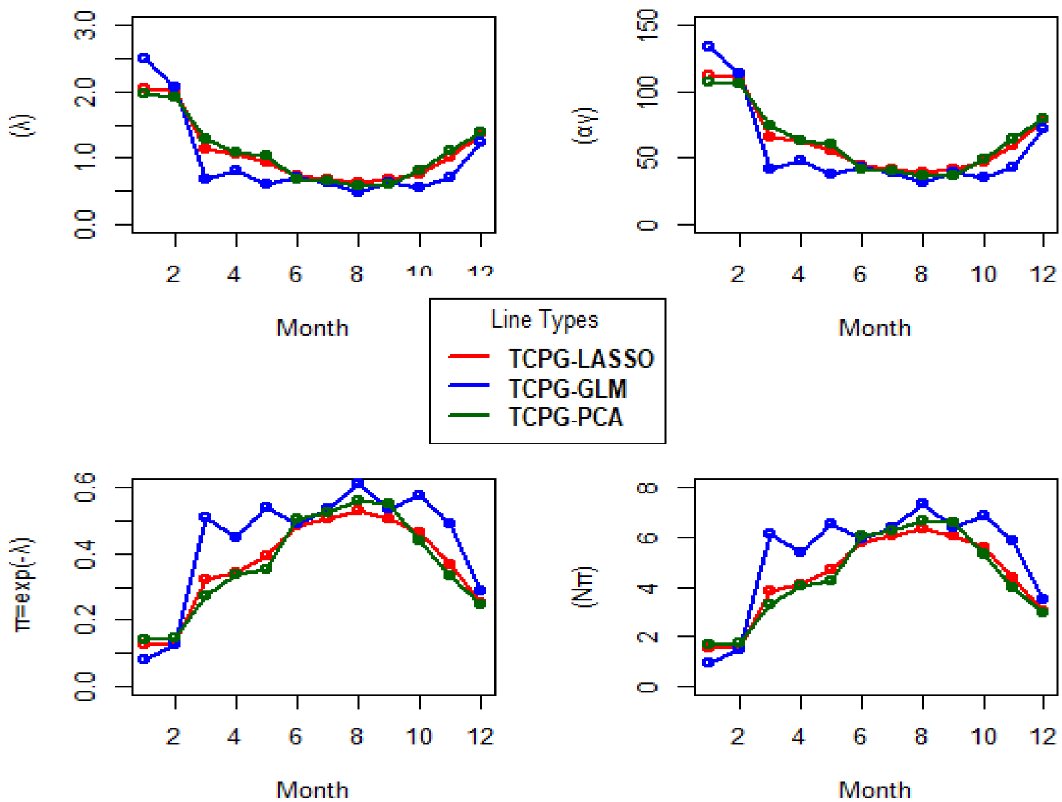

FIGURE 6. The plot of parameter estimates for (a) $\lambda, (b)\ \alpha\gamma,\ (c)\pi = \exp(-\lambda),\ (d)\ N\pi$ for rainfall in Cigugur station

From Figure 6 (a) to 6 (d), it can be interpreted that the TCPG-Lasso method has a parameter estimate plot between the estimated parameter plots of the TCPG-GLM and TCPG-PCA methods. This shows that the TCPG-Lasso method is better than other models. The four plots have a pattern following the letter U called the monsoonal rainfall pattern following the characteristics of the rainfall pattern in West Java Province which has one peak of the rainy season. The dry season occurs in June, July, and August, which is indicated by the chance that there will be no rain and many non-rainy events in Figure 6 (c) increase in June, July, and August, while wet months occur in December, January, and February. The other six months are the transitional season.

Table 4 is the estimates of some TCPG parameters that describe the characteristics of rainfall which can be interpreted that the average number of daily rainfall events per month (λ) for January is twice, shape parameters γ for January is 98, the average daily rainfall per month $(\alpha\gamma)$ for January is 110.63, probability of no rain event per month (π) for January is 0.13, and many events of no rain per month $(N\pi)$ for January is twice.

TABLE 4. Parameter estimates of $\lambda, \gamma, \alpha\gamma, \pi = \exp(-\lambda), N\pi$ for rainfall in Cigugur

Month	Actual	Prediction $(\hat{y})$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}\hat{\gamma}$	$\hat{\pi}$	$N\hat{\pi}$
1	280	236	2.02	98.67	110.63	0.13	1.58
2	252	227	2.01	98.04	109.93	0.13	1.60
3	51	61	1.15	59.94	67.20	0.31	3.76
4	147	64	1.08	56.41	63.25	0.33	4.06
5	114	43	0.95	50.57	56.70	0.48	4.60
6	25	28	0.75	40.68	45.61	0.47	5.66
7	0	24	0.7	38.68	43.26	0.49	5.91
8	0	21	0.66	36.41	40.83	0.51	6.18
9	0	24	0.70	38.68	43.37	0.49	5.90
10	0	29	0.78	42.21	47.33	0.45	5.49
11	39	53	1.01	53.01	59.44	0.36	4.37
12	153	111	1.35	68.81	77.15	0.25	3.10

5. CONCLUSION

Based on the above discussion, it can be concluded that:

1. The TCPG distribution is a flexible and accurate model for rainfall modeling because it can handle two components of rainfall, namely the discrete component which describes no rain with zero rainfall and the continuous component which describes the occurrence of rain with a rainfall intensity of more than zero.
2. The TCPG model with Lasso regularization is used to handle multicollinearity in the model. The addition of Lasso can reduce the RMSEP value compared to the TCPG-PCA and TCPG-GLM methods. Therefore, the TCPG-Lasso Model is highly recommended for rainfall modeling in order to obtain accurate prediction results.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] Sutikno, Statistical downscaling luaran GCM dan pemanfaatannya untuk peramalan produksi padi, Disertation, IPB University, 2008.

[2] M.M. Hasan, P.K. Dunn, A simple Poisson–gamma model for modelling rainfall occurrence and amount simultaneously, Agric. Forest Meteorol. 150 (2010), 1319–1330.

[3] P.K. Dunn,Tweedie: Tweedie exponential family models; R Package. Version 2.2.1; R Core Team: Vienna, Austria, 2014.

[4] E. Zorita, H. von Storch, The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Climate, 12 (1999), 2474-2489.

[5] W. Qian, Y. Yang, H. Zou, Tweedie's compound poisson model with grouped elastic net, J. Comput. Graph. Stat. 25 (2016), 606–625.

[6] Y. Yang, R. Luo, R Khorsidi, Y Liu, Inferring tweedie compound poisson mixed models with adversarial variational methods, Presented at NIPS 2017 Workshop on Advances in Approximate Bayesian Inference, 2017.

[7] P.K. Dunn, Occurrence and quantity of precipitation can be modeled simultaneously. Int. J. Climatol. 24 (2004), 1231–1239.

[8] P.K. Dunn, G.K. Smyth, Series evaluation of Tweedie exponential dispersion model densities, Stat. Comput. 15 (2005), 267–280.

[9] T. Hastie, R. Tibshirani, J. Friedman. The elements of statistical learning: data mining, inference, and prediction. (2009), https://web.stanford.edu/~hastie/ElemStatLearn/

[10] A.M. Soleh, Pemodelan linier sebaran gamma dan pareto terampat dengan regulasi L_1 pada statistical downscaling untuk pendugaan curah hujan, Disertation, IPB University, 2015.

[11] A. Muslim, M. Hayati, B. Sartono, K.A. Notodiputro, A combined modeling of generalized linear mixed model and LASSO techniques for analizing monthly rainfall data, IOP Conf. Ser.: Earth Environ. Sci. 187 (2018), 012044.

[12] F. Novkaniza, M. Hayati, B. Sartono, K. Anwar Notodiputro, Fused lasso for modeling monthly raifall in Indramayu sub distric West Java Indonesia, IOP Conf. Ser.: Earth Environ. Sci. 187 (2018), 012046.

[13] N.C. Dzupire, P. Ngare, L. Odongo, A Poisson-Gamma model for zero inflated rainfall data, J. Probab. Stat. 2018 (2018), 1012647.

[14] N. Fadhli, A.H. Wigena, A. Djuraidah, Determination of general circulation model grid resolution to improve accuracy of rainfall prediction in West Java, Int. J. Sci. Res. Publ. 9 (2019), p91106.

[15] Y.S. Trzaska, E. Schnarr. A review of downscaling methods for climate change projection, (2014). https://www.climatelinks.org/resources/review-downscaling-methods-climate-change-projections