International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET23103121

Rainfall Prediction in Statistical Downscaling Using Tweedie Compound Response and Lasso Penalty

Ma'rufah Hayati*, Reni Permatasari
Department of Statistic, The University of Nahdlatul Ulama Lampung, Sukadana, Indonesia

ARTICLEINFO

Article History:
Accepted: 10 June 2023
Published: 22 June 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
537-546

ABSTRACT

Statistical Downscaling (SD) is a technique in climatology to analyze the relationship between large-scale (global) data and small-scale (local) data using statistical modeling. The SD technique is used to overcome the inability of global scale data in the form of the General Circulation Model (GCM) as a low resolution predictor to predict local scale climatic conditions in the form of high resolution rainfall as a direct response. Rainfall consists of two components, namely continuous and discrete. The continuous component describes the intensity of rainfall while the discrete component describes the occurrence of rain. both components have an important role in predicting rainfall so it is necessary to choose the right distribution. One distribution that is able to handle both rain components is the mixed Tweedie distribution, namely the Gamma and Poisson distribution, hereinafter referred to as the Tweedie compound. GCM generally has multicollinearity problems in SD modeling. This can be handled using the Lasso penalty. This study aims to predict rainfall and rainfall events by taking into account the multicollinearity problem in the model for locations on different plains. Based on the research results, it was found that Cigugur Station from the highland gets the smallest RMSEP value and the biggest r-correlation. This model is not good enough to use for moderate plains rainfall data.

Keywords : Tweedie Compound, Poisson-Gamma, Statistical Downscaling, Lasso, Rainfall.

I. INTRODUCTION

Statistical Downscaling (SD) aims to link global scale variables (predictors) and local scale variables (responses) [1]. The SD technique helps overcome low-resolution GCMs that are unable to directly predict high-scale local climate conditions. the predictor variable used is the GCM outcome in the form of precipitation and the response variable used is rainfall. The role of GCM output is very important for predicting rainfall in SD modeling. Rainfall actually consists of two components, namely continuous and discrete. The continuous component describes the intensity of rainfall whose value is greater than 0, while the discrete component describes the occurrence of rain and no rain. The occurrence of rain indicates that there is recorded rainfall and there is no rain event, meaning the intensity is 0 because there is no recorded rainfall.[2][3]

Rainfall predictions generally only model one of the components. Several studies regarding the prediction of rainfall in SD were carried out by [4] using normal distribution with Fused Lasso penalty. [5] uses normal distribution with a generalized linear mixed model and Lasso penalty. [6] use the normal distribution and Lasso penalty. The three studies only modeled the continuous component. The most basic model for predicting rainfall for the continuous component is using the normal distribution in a linear model. This will cause a violation of the assumption because the rainfall data is generally skewed to the right which is identical to the Gamma distribution. Rain events can be modeled using Poisson distribution [7][8]. The general linear model is used as a solution. However, some rain events will be intensity 0 (no rain). This causes the Gamma distribution to be less precise for modeling. The Tweedie compound was proposed by Dunn in 2004 for rainfall modeling which represents the sum of continuous rainfall events. Because of its ability to accommodate both components of rainfall simultaneously. This is necessary because both components have important information for predicting future rainfall [3].

GCM outputs in SD often violate the assumption of multicollinearity. This needs to be addressed in order to obtain meaningful predictions. This study deals with multicollinearity problems using Lasso. Several studies regarding the prediction of rainfall using the Tweedie compound distribution have been carried out [9] modeling rainfall using the tweedio compound poisson gamma distribution with Lasso regularization compared to the tweedie GLM model and tweedie principle component analysis (PCA) and [10] carried out modeling using the Generalized linear mixed model of the tweedie compound poisson gamma response and the PCA reduction method. Based on this background, this study aims to compare rainfall predictions using Tweedie compound responses with Lasso ropes at three different locations, namely locations for high, medium and low altitudes.

II. METHODS AND MATERIAL

This sub-chapter will explain some of the supporting materials in the research including:

A. General Circulation Models (GCM) and Statistical Downscaling (SD)

The General Circulation Model is an important tool in the study of diversity and climate change [11]. This model describes the subsystems of the earth's climate, such as processes in the atmosphere, sea, land, and simulates climate conditions on a global scale. GCM simulates global climate variables on each grid with a size of $\pm 2.5^0$ or ± 300 km^2 in each layer of the atmosphere. However, GCM does not provide important information at higher resolution, such as temperature and precipitation on a local scale. GCM is still possible to use to obtain local scale information using the downscaling method. The downscaling method is the process of transforming data from a grid with large scale units into data on a grid with smaller scale units. One of the downscaling methods is statistical downscaling, in which data on a large scale grid in a certain period is used as a basis for determining data on a smaller scale grid. The equation for this method is:

$$y_{n\times 1} = f(X_{n\times k})$$

Where $y_{n\times 1}$ is rainfall, $X_{n\times k}$ is precipitation of GCM output data, n is the number of observations, k is the number of explanatory variables. Figure 1 is framework of statistical downscaling process [14]

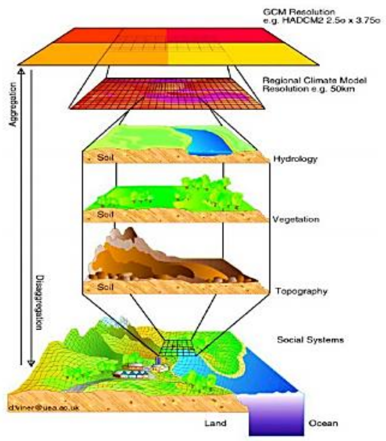

Figure 1 : Framework of Statistical Downscaling Process

B. Exponential Dispertion Model and Tweedie Compound

Tweedie is a special case of the Exponential Dispersion Model (EDM). The density function of the EDM is defined as a function of 2 parameters, namely:

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

θ is the canonical parameter in R , ϕ>0 is the dispersion parameter in (0, +∞) [12], k(θ) is a cumulative function of the exponential dispersion model, $a\ (y,\phi)$ is basis of normalized quantity that is independent of the parameter θ [13]. EDM has the property that the mean μ and variance Var(y) can be calculated from the first and second derivatives of k(θ) w.r.t θ. Due to the one-to-one mapping between θ and μ. k''(θ) can be denoted as a function of the mean μ, $k''(\theta) = c$ which is known as a function of the variance.

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

The normalized quantity $a\ (y,\phi)$ can be obtained as follows:

$$a\ (y,\phi) = \begin{cases} \frac{\mu_i^{1-p}}{1-p} & jika\ y = 0 \\ \frac{1}{y}\sum_{n=1}^{\infty} a_n(\ y,\phi,p) & jika\ y > 0 \end{cases}$$

$\sum_{n=1}^{\infty} a_n(\ y,\phi,p)$ is Wright's generalized Bessel function. Note that a (y,ϕ) in the tweedie model is also a function of p. The p parameter value is used as a determinant of the Tweedie distribution. Some of the general distributions that enter the tweedie family are known to have an analytic form including, $\rho = 0$ is normal distribution, $\rho = 1, \phi = 1$ is Poisson distribution $Tw_1(\mu, 1) = Poisson\ (\mu)$, $\rho = 2$is Gamma distribution $Tw_2(\mu,\phi) = Gamma\ (\mu,\phi)$, $\rho = 3$ is the inverse Gaussian distribution $Tw_3(\mu,\phi) = Invers - Gaussian\ (\mu,\phi)$, $1 < \rho < 2$ is a Tweedie compound which can model discrete and continuous components simultaneously, so the stretch can model both events and the amount of rainfall simultaneously, $\rho \geq 2$ can model positive data, the data is skewed to the right [14]. In the field of meteorology, Tweedie assumes Y as the total monthly rainfall, N_t is the total number of rain events per month and y_i is the precipitation from the i-th event [8] 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 $(y_i)_{i \geq} 1$, assumed to have an independent and identical Gamma distribution of the time of the rain events:

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

So $y_i\ \sim Gamma(\alpha,\gamma)$ is a probability density function with mean αγ and variance $\alpha\gamma^2$. If N=0 then Y=0, if

N>0 maka $Y = \sum_{i}^{N_t} y_i$ [12]. The probability density function for Y for N>0 is:

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

Rainfall contains zero value and a continuous positive value. Thus, Y is the total amount of monthly rainfall, which is represented by the Poisson's sum of the Gamma random variables. N is the number of rain events and y_i rainfall intensity for the ith rain event or the amount of rain per day. When Y is the total amount of monthly rainfall. So that Y has a Tweedie Compound distribution with the following parameter: λ is average number of rainfall events per month, γ is shape of the precipitation event, $\alpha\gamma$ is average amount of precipitation per event. The relationship between parameters $\{\lambda, \alpha, \gamma\}$ from Tweedie Compound and parameters $\{\mu, \phi, p\}$ from the tweedie model is as follows:

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

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

According to (Dunn & Smyth, 2005) the probability that no rain will occur is:

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

Equivalent to equation

$$P(Y, N = n|\lambda, \alpha, \gamma\} = d_0(y)e^{-\lambda} \, \|_{n=0} + \frac{y^{n\alpha-1}e^{-y/\beta}}{\beta^{n\alpha}\Gamma(n\alpha)} \frac{\lambda^n e^{-\lambda}}{n!} \, \|_{n>0} \quad (4)$$

where $d_0(y)$ Delta dirac function at zero. The joint distribution $P(Y, N = n|\lambda, \alpha, \gamma\}$ according to [13] has a close form expression by substituting equation (2) into equation (4) so that the joint density function is represented by $\{\mu, \phi, p\}$ as:

$$P(Y, N = n|\, \mu, \phi, p\} = \left[\exp\left(-\frac{\mu^{2-p}}{\phi(2-p)}\right]^{\|_{n=0}}$$

$$* \left[\begin{array}{c} \exp\left(n\left(-\frac{\log(\phi)}{p-1} + \frac{2-p}{p-1}\log\left(\frac{y}{p-1}\right) - \log(2-p)\right) - log\Gamma(n+1\right) \\ -\frac{1}{\phi}\left(\frac{\mu^{1-p}y}{p-1} + \frac{\mu^{2-p}}{2-p}\right) - log\Gamma\left(\frac{2-p}{p-1}n\right) - \log(y)) \end{array}\right]^{\|_{n>0}}$$

C. LASSO (Least Absolute Shringkage and Selection Variable)

The Least Absolute Shringkage and selection operator (LASSO) method was introduced by Thibsirani in 1996. This method is used for variable selection by shrinking the linear regression parameter coefficients of predictors that are highly correlated with error, to almost 0 or exactly zero, by adding a penalty called with L1 regulation. The L1 regulation is by giving constraints $\sum_{j=1}^{p}|\beta_k| \leq t$, $t \geq 0$ in the modeling objective function. This provides two advantages, namely variable selection and stable parameter estimation. Parameter estimation in linear modeling with L1 regulation has the following solution:

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

with $L(y, \beta_k)$ is response probability function., n is umber of observations, λ is tuning parameters (parameters controlling the LASSO coefficient shrinkage) with λ ≥0, β_k is parameter regression coefficient. [9]

D. Data Analysis Framework and Data Description

The application of the Tweedie Compound model with Lasso penalty is carried out in the field of climatology, especially rainfall with statistical downscaling techniques. This study uses R software assisted by the statmod and tweedie packages to determine index parameters, and dispersion. The Hdtweedie package is used for Tweedie Compound modeling with Lasso penalty, determination of regression parameters, and predictions. The model used in this study is as follows:

$$\log(\mu) = \beta_0 + \beta^T x \qquad (4)$$

By maximizing the penalty model equation (5) using a 2 layer loop algorithm which incorporates the blockwise majorization descent method into iteratively re-weighted least squares (IRLS-BMD) proposed by [12].

$$\left(\hat{\beta}_0, \hat{\beta}\right) = argmin_{(\hat{\beta}_0, \hat{\beta})} l(\beta_0, \beta) + \lambda_1 \left\|\beta_j\right\| \qquad (5)$$

This study uses rainfall data as a predictor variable and the outcome of the General Circulation Model (GCM) in the form of monthly precipitation as a response variable with a time period from January 1981 to December 2009 of 348 months. Rainfall data is located between -7.78^0 to -6.28^0 South Latitude and 108.40^0 to 107.87^0 East Longitude in West Java province. The data unit used is mm/day. This was obtained from the Center for Meteorology, Climatology and Geophysics. GCM data as an explanatory variable was obtained from The National Centers for Environmental Prediction (NCEP) in the form of a Climate Forcast System Reanalysis (CSFR) model which can be downloaded on the website https:|//rda.ncar.edu/ (Saha et al. 2010) . The locations used are 3 rain stations representing each plain. This is considered to see the ability of the model used in modeling and predicting each plain which has different characteristics.

TABLE I. INFORMATION OF RESEARCH VARIABLE

Variable	Information
$Y_{n\ x\ 1}$	3 rainfall data are used to represent 3 stations for each land, namely: 1. Highlands: Cigugur 2. Moderate plain: Lekong 3. Lowland: Pusakanegara
$X_{n\ xp}$	GCM (General Circulation Model) n=348 observation, p=40 Predictors

The research steps are as follows:

1) Exploration of data through plots and histograms to see the characteristics of rainfall.
2) Determining the parameters of the p index and phi parameter (ϕ)
3) Tweedie Compound Modeling with Lasso Penalty
4) Rainfall prediction
5) Evaluate the model by looking at the RMSEP and Rsquare values of each station

Determination of parameters α, γ, the number of daily rainfall events λ, Average daily rainfall intensity per month αγ, probability of no monthly rainfall (π), prediction of monthly rainfall μ

III.RESULTS AND DISCUSSION

The first step of this research is data exploration to see the characteristics of rainfall data at each rain station can be seen in Figure 2. The histogram in Figure 2 shows that the rainfall data at each rain station has a data frequency dominated by zero and positive values . Thus, the rainfall data appears to be in accordance with the distribution characteristics of the Tweedie compound, namely the data is positive and contains exact zero. The box plots of the three rain stations have a monsoon pattern where the lowest rainfall is between June and September.

The Tweedie compound distribution index parameter is in the range of values $1 < \rho < 2$. Thus, the next step is to determine the index parameter value for each rain station, as well as the ϕ parameter which will also be obtained simultaneously when searching for the index parameter value which can be obtained with the R package tweedie software, the tweedie.profile() function. The index parameter is selected from several candidate index values that have the smallest profile likelihood value accompanied by the dispersion parameter output.

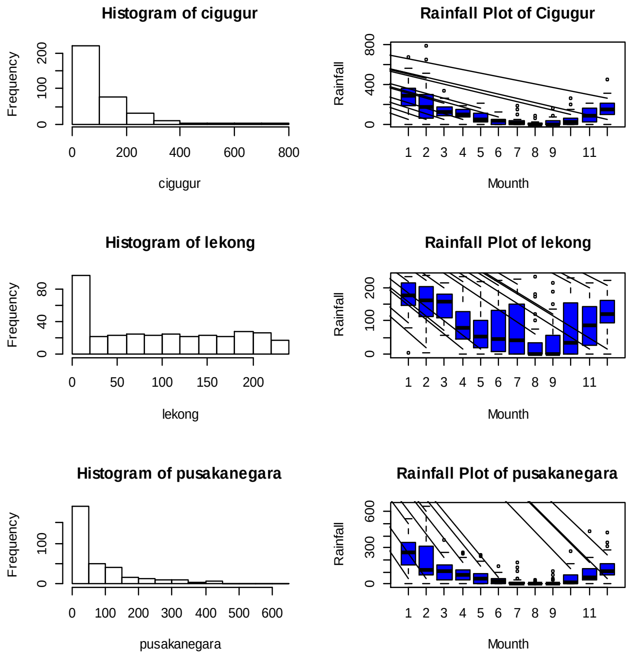

Figure 2: histogram and rainfall plot of each rain station

The profile likelihood plot for determining the best index parameter for each rain station can be seen in Figure 3. The candidate index parameter values entered into the tweedie.profile function are between 1.2 to 2.

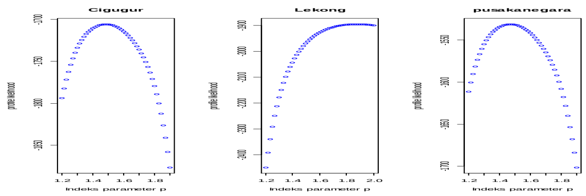

Figure 3: Likelihood parameter index plot for Cigugur, Lekong and Pusakanegara stations

The parameters obtained besides the index parameter include index value, dispersion, furthermore the minimum α and lambda values from the penalty lasso can be obtained by modeling the Tweedie compound with the penalty Lasso. The estimated results can be seen in Table 2.

TABLE 2. PARAMETER ESTIMATE FOR p, DISPERSION ϕ, α , AND LASSO PENALTY λ_1

Value parameter estimate	Cigugur	Lekong	Pusakanegara
profile likelihood of p	1.48	1.86	1.47
CI 95 %	(1.43,1.54)	(1.79 , 1.95)	(1.42, 1.53)
Phi. Value	15.49	2.36	19.75
Alpha	1.06	0.15	1.12
Lamda.min (λ_1)	3.606624e-05	4.024771e-06	4.638868e-05

The predicted index parameter p from the profile likelihood plot is between the values $1 < \rho < 2$ for the three rain stations. So, it is true that the rainfall data for each rain station has a Tweedie compound distribution. Therefore, further Tweedie compound modeling can be carried out with the Lasso Penalty assisted by the R package Hdtweedie software.

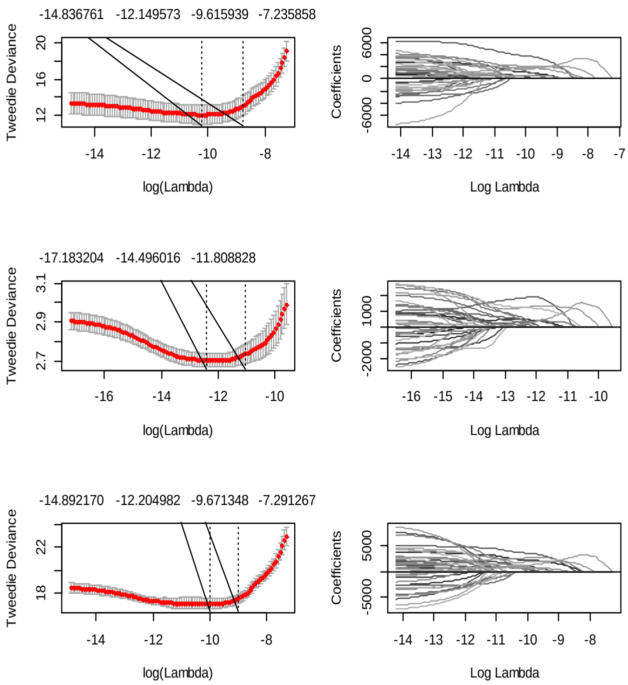

Figure 4: The selected Lambda plot with the smallest crossed validate error value and plot of coefficients based on the lambda values λ_1

The analysis process using the HDtweedie function requires the response y, the variable x in the form of a matrix, and the parameter index values obtained previously. The results of the analysis obtained a non-zero regression coefficient and several penalty values λ_1. The plot of the analysis results can be seen in Figure 4. The value of λ_1 with the minimum and best evaluation is used for modeling based on n-fold cross validation with the cv.Hdtweedie function. The evaluation results can be seen in Figure 4 in the form of a curve (red dotted line) and an upper and lower standard deviation curve for all rows λ_1 (error bar). The vertical black line indicates the selected lambda value.

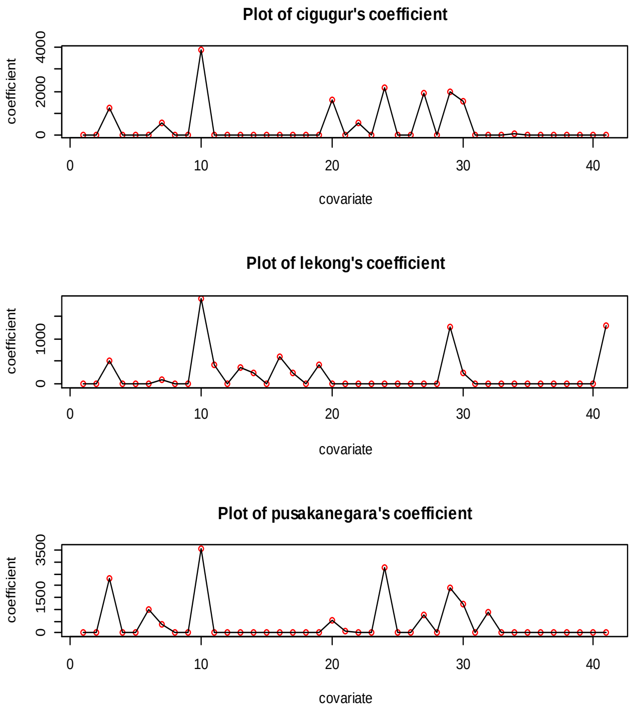

Figure 5: Regression coefficients for Cigugur, Lekong and Pusakanegara rain stations

The regression coefficient in Figure 4 was obtained based on lambda λ_1 with a minimum mean cross-validate error and was then used for variable selection. The selected explanatory variable will have a regression coefficient of zero, while the unselected explanatory variable will have a value of more than zero. This can be seen in Figure 5. Based on Figure 5, the explanatory variables that were not selected for each rain station are

TABLE 3. EXPLANATORY VARIABLES THAT ARE NOT SELECTED FOR EACH RAIN STATION

Rain station	Predictor variable	total
Cigugur	X2,X6,X9,X19,X21,X23,X26, X28	8
Lekong	X2,X6,X9,X10,X12,X13,X15,X16, X18,X28,X40	11
Pusakanegara	X2,X5,X6,X9,X19,X20,X23,X26,X28,X30	9

The three rain stations obtained the explanatory variables that were not much different in number and the same explanatory variables were selected for the three rain stations. This shows that the Tweedie compound model with Lasso penalty has almost the same ability for each rain station. Although the selection of rain stations has different characteristics for the highlands, medium and low. Predictive models can be created based on the cv.Hdtweedie() function. The minimal lambda is used in the prediction model using the predict() function. This function requires the cv.Hdtweedie() function, a new x matrix for the desired observations, and a minimum lambda value. Furthermore, the predicted and actual data are visualized through plots to compare the two rainfall data which can be seen in Figures 6 to 8 for the three rain stations. Predictions are made for 2009 from January to December.

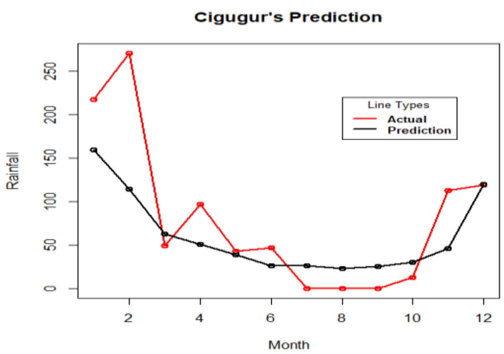

Figure 6: Cigugur station rainfall prediction

Plot of actual and predicted rainfall for the Cigugur station can be seen on Figure 6. The plot shows that the actual and predicted rainfall is not much different because the pattern of data in the plot is not much different. This shows that the model is quite good at predicting. Thus, the model can be considered for predicting rainfall over a period of several years, especially for data from upland rain stations.

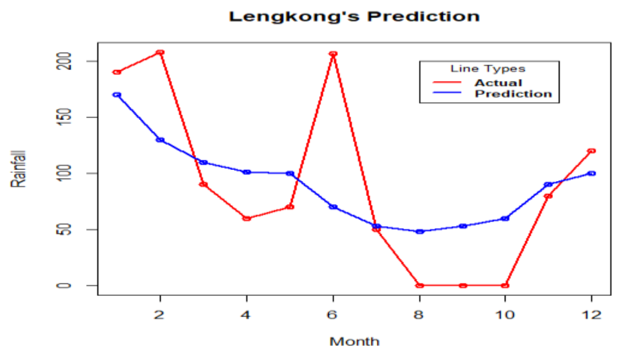

Figure 7: Lekong station rainfall prediction

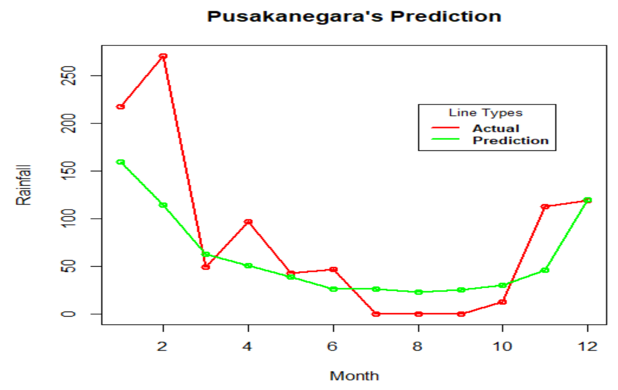

Figure 8: Pusakanegara station rainfall prediction

The rainfall plot for the Lekong rain station can be seen in Figure 7. It can be seen that the actual and predicted rainfall data differ quite a lot, especially in February and June, but the overall pattern is not much different. Thus, the model can still be considered to model rainfall data from temperate plains. The rainfall plot for the Pusakanegara rain station can be seen in Figure 8. It can be seen that the actual and predicted rainfall data differ in January, February, November and December, but the overall pattern is not much different or even similar. Thus, the model is good enough to model rainfall data from the lowlands.

The goodness of the model is seen through the Root Mean Square Error Prediction and r-correlation between actual data and predictiom values shown in Table 4. The smallest RMSEP was obtained by the Cigugur rain station and the highest was Lekong. The largest r-correlation is obtained by Cigugur and the smallest is obtained by Lekong. When looking at Figure 1 again, the rainfall plots for the Cigugur and Pusakanegara rain stations have a monsoon pattern. the two rain stations come from the highlands and lowlands and have a p index parameter value of around 1.4-1.5 . Meanwhile, Lekong does not perfectly form the pattern and the index parameter value of 1.8 is close to 2, which tends to approach the gamma distribution. Thus, the Tweedie compound model with the penalty lasso is very good for rainfall data that has a Monsoon rain pattern with an index parameter of 1.4-1.5 and is not good for rainfall data originating from the lowlands.

TABLE 4

FONT SIZES FOR PAPERS

Stasiun	RMSEP	r-corelation
Cigugur	23.33	0.93
Lekong	60.07	0.83
Pusaka negara	28.04	0.87

Prediction of some parameters of rainfall characteristic Tweedie compound model was chosen only for Cigugur station because it has the largest RMSEP value and correlation. The rainfall characteristics from Table 5 can be interpreted that the average monthly daily rainfall (λ) in January is twice, the shape parameter γ in January is 98, the average daily rainfall per month ($\alpha\gamma$) in January is 110.63, there is no chance rainfall events per month (π) for January is 0.12, and the number of events without rain per month ($N\pi$) for January is one.

TABLE 5. PRE PREDICTION OF OTHER RAINFALL CHARACTERISTICS FOR $\lambda, \gamma, \alpha\gamma, \pi = \text{EXP}(-\lambda)$, $N\pi$

Month	Actual	Prediction (μ)	λ	γ	$\alpha\gamma$	π	$N\pi$
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

IV.CONCLUSION

This sub-chapter answers the research objectives that have been presented previously. The results of the data analysis above can be concluded that:

1) The Tweedie compound model with a Lasso penalty is quite good and needs to be considered for rainfall modeling, especially for rainfall data from the highlands and lowlands.

2) This model is very well used for rainfall data with monsoon patterns with index parameters between 1.4-1.5

This model is not only able to predict rainfall intensity but also able to predict other parameters such as monthly average daily rainfall events (λ), shape parameter γ months, average daily rainfall per month (αγ), probability of no rain events per month (π) and the number of events without rain per month (Nπ).

V. REFERENCES

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

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

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

[4]. 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. 2018.

[5]. 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

[6]. M. Hayati and A. Muslim. Generalized Linear Mixed Model and Lasso Regularization for Statistical Downscaling. ENTHUSIASTIC: International Journal of Statistics and Data Science. vol. 1, No. 1, April 2021, pp. 36-52. 2021.

[7]. WH Bonat, C.C. Kokonendji. Flexible Tweedie regression models for continuous data. J Stat Comput Simul. 87(11):2138–2152. doi:10.1080/00949655.2017.1318876, 2017.

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

[9]. M Hayati, A. H. Wigena, A Djuraidah, A. Kurnia. A New Approach To Statistical Downscaling Using Tweedie Compound Poisson Gamma Response And Lasso Regularization. Commun. Math. Biol. Neurosci, 2021:60. ISSN: 2052-254, 2021.

[10]. M Hayati, A. H. Wigena, A Djuraidah, A. Kurnia. The Spatio-Temporal Model For The Tweedie Compound Poisson Gamma Response In Statistical Downscaling. J. Math. Comput. Sci. 2022, 12:35, ISSN: 1927-5307, 2022,.

[11]. E. Zorita, H. von Storch, The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Climate, 12 , 2474-2489, 1999.
[12]. W. Qian, Y. Yang, H. Zou, Tweedie's compound poisson model with grouped elastic net, J. Comput. Graph. Stat. 25 , 606–625, 2016.
[13]. Y. Yang, R. Luo, R Khorsidi, Y Liu, Inferring tweedie compound poisson mixed models with adversarial variational methods, Presented at NIPS Workshop on Advances in Approximate Bayesian Inference, 2017.
[14]. Ma R. Jϕrgensen B. Nested Generalized Linier Mixed Models: an ortodhox Best Linear Unbiased Predictor Approach, J.R. Statist.Soc.B. 69(4) , 625-641, 2006.
[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
[16]. P.K. Dunn,Tweedie: Tweedie exponential family models; R Package. Version 2.2.1; R Core Team: Vienna, Austria, 2014.
[17]. P.K. Dunn, Occurrence and quantity of precipitation can be modeled simultaneously. Int. J. Climatol. 24, 1231–1239, 2004.
[18]. P.K. Dunn, G.K. Smyth, Series evaluation of Tweedie exponential dispersion model densities, Stat. Comput. 15, 267–280, 2005.
[19]. T. Hastie, R. Tibshirani, J. Friedman. The elements of statistical learning: data mining, inference, and prediction. 2009. https://web.stanford.edu/~hastie/ElemStatLearn

Cite this article as :

Marufah Hayati, Reni Permatasari, "Rainfall Prediction in Statistical Downscaling Using Tweedie Compound Response and Lasso Penalty", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 3, pp. 537-546, May-June 2023. Available at doi : https://doi.org/10.32628/IJSRSET23103121
Journal URL : https://ijsrset.com/IJSRSET23103121